Econometric Theory, **38**, 2022, 1140–1174.
doi:10.1017/S0266466621000530

AVERAGE DENSITY ESTIMATORS: EFFICIENCY AND BOOTSTRAP CONSISTENCY

Matias D. Cattaneo
Princeton University

Michael Jansson
University of California at Berkeley

This paper highlights a tension between semiparametric efficiency and bootstrap consistency in the context of a canonical semiparametric estimation problem, namely the problem of estimating the average density. It is shown that although simple plug-in estimators suffer from bias problems preventing them from achieving semiparametric efficiency under minimal smoothness conditions, the nonparametric bootstrap automatically corrects for this bias and that, as a result, these seemingly inferior estimators achieve bootstrap consistency under minimal smoothness conditions. In contrast, several "debiased" estimators that achieve semiparametric efficiency under minimal smoothness conditions do not achieve bootstrap consistency under those same conditions.

1. INTRODUCTION

Peter Phillips is a towering figure in econometrics. Among other things, his pathbreaking work on nonstationary time series (e.g., Phillips (1987) and Phillips and Perron (1988) in the case of unit-root autoregression and Phillips and Durlauf (1986) and Phillips and Hansen (1990) in the case of cointegration) has forcefully demonstrated that estimators can be useful without having limiting distributions that are "simple."In this paper, we show that a similar phenomenon occurs in a seemingly very different setting, namely a canonical semiparametric estimation problem in a model with independent and identically distributed (*i.i.d.*) data.

The specific semiparametric estimation problem we consider is the problem of estimating the average density of a continuously distributed random vector

For comments and suggestions, we are grateful to the Co-Editor, two referees, and seminar participants at the Cowles Foundation conference celebrating Peter Phillips's 40 years at Yale, the 2018 LAMES conference, the 2019 CIREQ Montreal Econometrics Conference, ITAM, University of North Carolina, Princeton University, UC Santa Barbara, and Oxford University. Cattaneo gratefully acknowledges financial support from the National Science Foundation through grants SES-1459931 and SES-1947805, and Jansson gratefully acknowledges financial support from the National Science Foundation through grants SES-1459967 and SES-1947662, and the research support of CREATES (funded by the Danish National Research Foundation under grant no. DNRF78). Address correspondence to Michael Jansson, Department of Economics and CREATES, University of California at Berkeley, Berkeley, CA, USA; e-mail: mjansson@berkeley.edu.

(of which we have a random sample of observations). In that setting, a well-known apparent shortcoming of simple "plug-in" estimators is that they have biases that are avoidable and potentially nonnegligible. In particular, the biases in question prevent the plug-in estimators from achieving semiparametric efficiency under minimal smoothness conditions. In recognition of this, several methods of "debiasing" have been proposed and have been found to be successful insofar as they give rise to estimators that do achieve semiparametric efficiency under minimal smoothness conditions. (The particular examples given in this paper were obtained by applying and combining ideas from Hall and Marron (1987), Bickel and Ritov (1988), and Powell, Stock, and Stoker (1989).)

Recognizing that construction of an estimator is often a means to the end of conducting inference, a natural question is whether existing average density estimators permit valid inference to be conducted under minimal smoothness conditions. In this paper, we answer a specific version of the latter question by investigating whether average density estimators achieve bootstrap consistency under minimal smoothness conditions. Looking at estimators through the lens of the bootstrap is of interest for several reasons, most notably because one can answer questions motivated by inference considerations without having to make additional (and potentially arbitrary) assumptions about the behavior of standard errors (i.e., estimators of nuisance parameters). In other words, because bootstrap consistency (or lack thereof) can be interpreted as a property of an estimator, it has the potential to shed new light on the relative merits of competing estimators. In this paper, we show that average density estimation provides an example where this potential is realized.

To be specific, whereas several distinct approaches to debiasing achieve semiparametric efficiency under minimal smoothness conditions, we find that many of the estimators produced by these approaches fail to achieve bootstrap consistency under minimal smoothness conditions. In contrast, in spite of failing to achieve semiparametric efficiency under minimal smoothness conditions, simple plug-in estimators achieve bootstrap consistency under minimal smoothness conditions. In other words, we find that plug-in estimators enjoy certain nontrivial advantages over some of their debiased counterparts.

The paper proceeds as follows. Section 2 presents the setup and introduces the formal questions we set out to answer. Studying the most prominent average density estimators, Sections 3 and 4 are concerned with efficiency and bootstrap consistency, respectively. Alternative bootstrap procedures are discussed in Section 5, whereas alternative estimators are analyzed in Section 6. Finally, Section 7 offers concluding remarks, and the Appendix collects proofs of our main results.

2. SETUP

Suppose $X_1,\ldots,X_n$ are *i.i.d.* copies of a continuously distributed random vector $X \in \mathbb{R}^d$ with an unknown density f_0. Assuming f_0 is square integrable, a widely

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

studied estimand in this setting is

$$\theta_0 = \mathbb{E}[f_0(X)],$$

the average density. Influential work on estimating θ_0 includes Hall and Marron (1987), Bickel and Ritov (1988), and Ritov and Bickel (1990); see also Giné and Nickl (2008a) and the references therein. In econometrics, estimators of θ_0 are often viewed as prototypical examples of two-step semiparametric estimators (in the terminology of Newey and McFadden (1994)) and therefore provide a natural starting point when attempting to shed light on the properties of two-step semiparametric estimators.

In what follows, we shall explore the extent to which certain prominent estimators of θ_0 enjoy one (or both) of two desirable properties. The first of these properties is a very conventional one, namely (semiparametric) efficiency. It is well known (e.g., Pfanzagl, 1982, Exam. 9.5.2; Ritov and Bickel, 1990) that if f_0 is bounded, then the efficient influence function L_0 is well defined and given by

$$L_0(x) = 2\{f_0(x) - \theta_0\}.$$

Accordingly, an estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ of θ_0 is said to be efficient if it satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} L_0(X_i) + o_{\mathbb{P}}(1). \tag{1}$$

Our analysis will proceed under the following condition on the density.

Condition D For some $s > d/4$ with $2s \notin \mathbb{N}$, f_0 is bounded and belongs to the Besov space $B^s_{2\infty}(\mathbb{R}^d)$.

As alluded to earlier, the assumption that f_0 is bounded serves the purpose of ensuring that

$$\sigma_0^2 = \mathbb{V}[L_0(X)],$$

the semiparametric variance bound implied by (1), is well defined and finite. As pointed out by Bickel and Ritov (1988) and Ritov and Bickel (1990), however, some (additional) assumptions are required on the part of f_0 for semiparametric efficiency to be achievable. For our purposes, it is convenient and turns out to be sufficient to assume that f_0 is smooth in the sense that it belongs to $B^s_{2\infty}(\mathbb{R}^d)$, as that assumption will enable us to employ results from Giné and Nickl (2008b) when showing asymptotic negligibility of certain remainder terms. In particular, and as further discussed below, the magnitude "smoothing" bias of the kernel-based estimators under consideration in this paper turns out to depend on f_0 through the smoothness of the function f_0^Δ given by

$$f_0^\Delta(x) = \int_{\mathbb{R}^d} f_0(u) f_0(x+u) du.$$

Condition D is convenient, because it follows from Giné and Nickl (2008b, Lem. 12) that f_0^{Δ} belongs to the Hölder space $\mathbf{C}^{2s}(\mathbb{R}^d)$ whenever f_0 is bounded and belongs to $B^s_{2\infty}(\mathbb{R}^d)$ with $2s \notin \mathbb{N}$. The second property of interest is (nonparametric) bootstrap consistency. In the setting of this paper, the most attractive definition of that property is the following. Letting $X^*_{1,n}, \ldots, X^*_{n,n}$ denote a random sample from the empirical distribution of $X_1, \ldots, X_n$ and letting $\hat{\theta}^*_n = \hat{\theta}_n(X^*_{1,n}, \ldots, X^*_{n,n})$ denote the natural bootstrap analog of $\hat{\theta}_n$, the bootstrap is said to be consistent if

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \le t] - \mathbb{P}^*_n[\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) \le t] \right| = o_{\mathbb{P}}(1), \tag{2}$$

where $\mathbb{P}^*_n$ denotes a probability computed under the bootstrap distribution conditional on the data.

To motivate interest in (2), recall that the (nominal) level $1-\alpha$ bootstrap confidence interval for θ_0 based on the "percentile method" (in the terminology of van der Vaart (1998)) is given by

$$\mathsf{CI}^{\mathsf{P}}_{n,1-\alpha} = \left[\hat{\theta}_n - q^*_{n,1-\alpha/2}\, ,\, \hat{\theta}_n - q^*_{n,\alpha/2}\right], \quad q^*_{n,a} = \inf\{q \in \mathbb{R} : \mathbb{P}^*_n[\hat{\theta}^*_n - \hat{\theta}_n \le q] \ge a\}.$$

This interval is said to be consistent if

$$\lim_{n\to\infty} \mathbb{P}[\theta_0 \in \mathsf{CI}^{\mathsf{P}}_{n,1-\alpha}] = 1-\alpha \tag{3}$$

and to be efficient if its end points satisfy

$$\sqrt{n}(\hat{\theta}_n - q^*_{n,a} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{1\le i\le n} L_0(X_i) - \Phi^{-1}(a)\sigma_0 + o_{\mathbb{P}}(1), \quad a \in \{\alpha/2, 1-\alpha/2\}, \tag{4}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. In addition to being "heuristically necessary,"the bootstrap consistency property (2) turns out to be sufficient for (3) and (4) in the cases of interest in this paper. In turn, the property (4) implies (by the duality between hypothesis tests and confidence intervals) that efficient two-sided tests of simple hypotheses about θ_0 can be based on $\mathsf{CI}^{\mathsf{P}}_{n,1-\alpha}$ whenever the interval is efficient. In other words, the property (2) has strong and obvious implications for inference, and although those implications may seem more important than bootstrap consistency per se, much of our subsequent discussion of the bootstrap focuses on (2) for specificity because that property seems more "fundamental" than (3) and (4) in the sense that it is not directly associated with a particular inference method.

At any rate, because the properties of $\hat{\theta}^*_n$ and $\mathsf{CI}^{\mathsf{P}}_{n,1-\alpha}$ are governed solely by (the density f_0 and) the functional form of $\hat{\theta}_n$, the properties (2)–(4) can all be interpreted as properties of the estimator $\hat{\theta}_n$, and one of the main purposes of this paper is to explore the relationship between those properties and the more familiar (efficiency) property (1).

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

The (nominal) level $1-\alpha$ bootstrap confidence interval for θ_0 based on "Efron's percentile method" (in the terminology of van der Vaart (1998)) is given by

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

Suppose (2) holds. Then, $\mathsf{CI}^{\mathsf{E}}_{n,1-\alpha}$ is consistent if also (1) holds. On the other hand, and in contrast to $\mathsf{CI}^{\mathsf{P}}_{n,1-\alpha}$, it turns out that in the cases of interest, in this paper, the interval $\mathsf{CI}^{\mathsf{E}}_{n,1-\alpha}$ is inconsistent when (1) fails. Partly, for this reason, we focus on intervals based on the percentile method.

Suppose (1) holds. Letting $\hat{\sigma}^2_n$ denote an estimator of σ^2_0, a natural (nominal) level $1-\alpha$ confidence interval motivated by the distributional approximation $\sqrt{n}(\hat{\theta}_n - \theta_0) \dot{\sim} \mathcal{N}(0,\hat{\sigma}^2_n)$ is the "Normal" interval given by

$$\mathsf{CI}^{\mathsf{N}}_{n,1-\alpha} = \left[\hat{\theta}_n - \Phi^{-1}(1-\alpha/2)\hat{\sigma}_n/\sqrt{n}\,,\ \hat{\theta}_n - \Phi^{-1}(\alpha/2)\hat{\sigma}_n/\sqrt{n}\right].$$

This interval is consistent if $\hat{\sigma}^2_n$ is consistent. The bootstrap consistency property (2) is neither necessary nor sufficient for the "bootstrap variance consistency" property

$$\hat{\sigma}^{2,*}_n = n\mathbb{V}[\hat{\theta}^*_n | X_1,\ldots,X_n] \to_{\mathbb{P}} \sigma^2_0. \tag{5}$$

Following Bickel and Freedman (1981), one way of ensuring that bootstrap variance consistency is implied by bootstrap consistency is to employ the Mallows metric d_2 when defining bootstrap consistency. The examples studied herein have the feature that (5) can hold even if (2) (and therefore also convergence in the Mallows metric) fails. Partly, for this reason, it seems more attractive (to us at least) to define bootstrap consistency as in (2), hereby treating bootstrap consistency and bootstrap variance consistency as distinct (i.e., nonnested) properties.

3. AVERAGE DENSITY ESTIMATORS: EFFICIENCY

Our discussion of efficiency (or otherwise) of average density estimators $\hat{\theta}_n$ will be based on the natural decomposition of the estimation error $\hat{\theta}_n - \theta_0$ into its bias and "noise" components $\mathbb{E}[\hat{\theta}_n] - \theta_0$ and $\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]$. If these components satisfy

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n] - \theta_0) = o(1) \tag{6}$$

and

$$\sqrt{n}(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]) = \frac{1}{\sqrt{n}} \sum_{1\le i\le n} L_0(X_i) + o_{\mathbb{P}}(1), \tag{7}$$

respectively, then (1) holds. Moreover, if (7) holds, then the easy-to-interpret bias condition (6) is necessary and sufficient for (1). The latter observation is particularly useful for our purposes, as it turns out that the estimators of interest satisfy (7) under very mild conditions.

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

The simplest average density estimator is arguably the kernel-based "plug-in" estimator

$$\hat{\theta}_n^{\mathsf{AD}} = \frac{1}{n} \sum_{1 \le i \le n} \hat{f}_n(X_i),$$

where, for some kernel K and some bandwidth h_n,$\hat{f}_n$ denotes the kernel density estimator

$$\hat{f}_n(x) = \frac{1}{n} \sum_{1 \le j \le n} K_n(x - X_j), \qquad K_n(x) = \frac{1}{h_n^d} K\left(\frac{x}{h_n}\right).$$

When developing results for $\hat{\theta}_n^{\mathsf{AD}}$ and other estimators, we impose the following standard condition on the kernel, in which $\|\cdot\|_1$ denotes the ℓ_1-norm and u^l is shorthand for $u_1^{l_1} \cdots u_d^{l_d}$ when $u = (u_1, \ldots, u_d)' \in \mathbb{R}^d$ and $l = (l_1, \ldots, l_d)' \in \mathbb{Z}_+^d$.

Condition K For some $P > d/2, K$ is even and bounded with

$$\int_{\mathbb{R}^d} |K(u)| \, (1 + \|u\|_1^P) du < \infty$$

and

$$\int_{\mathbb{R}^d} u^l K(u) du = \begin{cases} 1 & \text{if } l = 0, \\ 0 & \text{if } l \in \mathbb{Z}_+^d \text{ and } 0 < \|l\|_1 < P. \end{cases}$$

The constant P in Condition K is the order of the kernel. Condition K therefore implies that K is a higher-order kernel when $d \ge 4$. As usual, we employ higher-order kernels in order to ensure that the magnitude of the smoothing bias of $\hat{f}_n$ is sufficiently small.

Under Conditions D and K, the density estimator $\hat{f}_n$ is consistent (pointwise) provided the bandwidth satisfies the following condition.

Condition B$^-$ As $n \to \infty, h_n \to 0$ and $nh_n^d \to \infty$.

More importantly, Condition B$^-$ implies that the average density estimator $\hat{\theta}_n^{\mathsf{AD}}$ satisfies (7) under Conditions D and K.[1] As a consequence, under Conditions D, K, and B$^-$, the estimator $\hat{\theta}_n^{\mathsf{AD}}$ is efficient if and only if it satisfies the bias condition (6).

Using the representation $\theta_0 = f_0^{\Delta}(0)$, the bias of $\hat{\theta}_n^{\mathsf{AD}}$ can be shown to admit the approximation

$$\mathbb{E}[\hat{\theta}_n^{\mathsf{AD}}] - \theta_0 \approx \frac{K(0)}{nh_n^d} + \int_{\mathbb{R}^d} K(t)[f_0^{\Delta}(h_n t) - f_0^{\Delta}(0)] dt, \tag{8}$$

where the approximation error is of order n^{-1}, the first term is a "leave-in" bias term (in the terminology of Cattaneo, Crump, and Jansson (2013)), and the second

[1] Conversely, Condition B$^-$ is minimal in the sense that the methods of Cattaneo, Crump, and Jansson (2014b) can be used to show that (7) can fail if Condition B$^-$ is violated.

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

term is a smoothing bias term. As previously mentioned, the function f_0^{Δ} belongs to the Hölder space $\mathbf{C}^{2s}(\mathbb{R}^d)$ under Condition D. Using this fact, it follows from a routine calculation (e.g., Tsybakov, 2009, Prop. 1.2) that if Conditions D and K are satisfied and if $h_n \to 0,$ then

$$\int_{\mathbb{R}^d} K(t)[f_0^{\Delta}(h_n t) - f_0^{\Delta}(0)]dt = O(h_n^{2S}), \qquad S = \min(P/2, s).$$

As a consequence, under Conditions D and K, the estimator $\hat{\theta}_n^{\mathsf{AD}}$ is efficient provided Condition B$^-$ is strengthened to the following condition.

Condition B$^+$ As $n \to \infty, nh_n^{4S} \to 0$ and $nh_n^{2d} \to \infty$.

Existence of a bandwidth sequence satisfying Condition B$^+$ requires that the parameter s governing the smoothness of f_0 satisfies $s > d/2,$ a stronger condition than the (minimal) condition $s > d/4$ included in Condition D.

This shortcoming of $\hat{\theta}_n^{\mathsf{AD}}$ is attributable to its leave-in bias, as it is the presence of the leave-in bias that requires a strengthening of the lower bound on the bandwidth from $nh_n^d \to \infty$ to $nh_n^{2d} \to \infty$. Of course, the leave-in bias of $\hat{\theta}_n^{\mathsf{AD}}$ is easily avoidable. One option is to employ a kernel satisfying $K(0) = 0$. Recognizing that all standard kernels have $K(0) \neq 0,$ a more natural option is to use the "bias-corrected" version of $\hat{\theta}_n^{\mathsf{AD}}$ given by

$$\hat{\theta}_n^{\mathsf{AD\text{-}BC}} = \hat{\theta}_n^{\mathsf{AD}} - \frac{K(0)}{nh_n^d}.$$

By construction, the bias of this estimator satisfies

$$\mathbb{E}[\hat{\theta}_n^{\mathsf{AD\text{-}BC}}] - \theta_0 \approx \int_{\mathbb{R}^d} K(t)[f_0^{\Delta}(h_n t) - f_0^{\Delta}(0)]dt = O(h_n^{2S}),$$

so under Conditions D and K, the bias condition (6) is satisfied by $\hat{\theta}_n^{\mathsf{AD\text{-}BC}}$ provided $nh_n^{4S} \to 0,$ implying in turn that $\hat{\theta}_n^{\mathsf{AD\text{-}BC}}$ is asymptotically efficient under Conditions D and K provided the bandwidth satisfies the following condition, which requires no additional smoothness (as measured by the value of s) relative to Condition D.

Condition B As $n \to \infty, nh_n^{4S} \to 0$ and $nh_n^{d} \to \infty$.

The leave-in bias of $\hat{\theta}_n^{\mathsf{AD}}$ is proportional to $1/(nh_n^d)$. Equipped with only that knowledge, the method of generalized jackknifing constructs a debiased version of $\hat{\theta}_n^{\mathsf{AD}}$ as a weighted sum of two (or more) versions of $\hat{\theta}_n^{\mathsf{AD}}$ implemented using different values of the bandwidth, where the weights are judiciously chosen to remove the leave-in bias. To give the simplest example, let $\hat{\theta}_n^{\mathsf{AD}}(h)$ denote the version of $\hat{\theta}_n^{\mathsf{AD}}$ associated with the bandwidth h. Then, for any $c \neq 1,$ the "generalized jackknife" version of $\hat{\theta}_n^{\mathsf{AD}}$ obtained by combining $\hat{\theta}_n^{\mathsf{AD}} = \hat{\theta}_n^{\mathsf{AD}}(h_n)$ and

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

$\hat{\theta}_n^{\texttt{AD}}(ch_n)$ is given by

$$\hat{\theta}_n^{\texttt{AD-GJ}} = \frac{1}{1-c^d}\hat{\theta}_n^{\texttt{AD}} - \frac{c^d}{1-c^d}\hat{\theta}_n^{\texttt{AD}}(ch_n).$$

Like $\hat{\theta}_n^{\texttt{AD}}$, the estimator $\hat{\theta}_n^{\texttt{AD-GJ}}$ satisfies (7) under Conditions D, K, and B$^-$. Moreover, because

$$\mathbb{E}[\hat{\theta}_n^{\texttt{AD}}(ch_n)] - \theta_0 \approx \frac{1}{c^d}\frac{K(0)}{nh_n^d} + \int_{\mathbb{R}^d} K(t)[f_0^{\Delta}(ch_n t) - f_0^{\Delta}(0)]dt,$$

the bias condition (6) is satisfied by $\hat{\theta}_n^{\texttt{AD-GJ}}$ under Condition B.

Finally, as its name suggests, the leave-in bias can also be avoided by employing "leave-out" estimators of f_0. A generic average density estimator based on leave-out density estimators is of the form

$$\hat{\theta}_n^{\texttt{AD-LO}} = \frac{1}{n}\sum_{1\le i\le n} \hat{f}_{i,n}^{\texttt{LO}}(X_i),$$

where $\hat{f}_{i,n}^{\texttt{LO}}$ is a kernel density estimator constructed using observations belonging to a set that does not include X_i. Relative to $\hat{\theta}_n^{\texttt{AD-BC}}$ and $\hat{\theta}_n^{\texttt{AD-GJ}}$, an attractive feature of $\hat{\theta}_n^{\texttt{AD-LO}}$ is that it can be constructed without knowledge of the functional form of the leave-in bias. For concreteness, we shall develop results for $\hat{\theta}_n^{\texttt{AD-LO}}$ only in the (leading) special case where the sample $X_1,\ldots,X_n$ is partitioned into $B_n \in \{2,\ldots,n\}$ disjoint blocks of (approximately) equal size and $\hat{f}_{i,n}^{\texttt{LO}}$ is constructed using observations from all blocks except the one to which the ith observation belongs. To be specific, we assume that $\hat{f}_{i,n}^{\texttt{LO}}$ is of the form

$$\hat{f}_{i,n}^{\texttt{LO}}(x) = \sum_{1\le j\le n} w_{ij,n}K_n(x-X_j), \qquad w_{ij,n} = \frac{\mathbb{1}(\lceil iB_n/n\rceil \ne \lceil jB_n/n\rceil)}{\sum_{1\le k\le n}\mathbb{1}(\lceil iB_n/n\rceil \ne \lceil kB_n/n\rceil)}.$$

When $B_n = n$, $\hat{f}_{i,n}^{\texttt{LO}}$ is the ith "leave-one-out" estimator of f_0, and the estimator $\hat{\theta}_n^{\texttt{AD-LO}}$ reduces to the estimator introduced in Hall and Marron (1987) and further studied by Giné and Nickl (2008a) (among many others). At the opposite extreme, when B_n is kept fixed, the estimator $\hat{\theta}_n^{\texttt{AD-LO}}$ is a "cross-fit" estimator (using a B_n-fold nonrandom partition of $\{1,\ldots,n\}$) in the terminology of Newey and Robins (2018).

Regardless of the choice of B_n, under Conditions D, K, and B$^-$, the estimator $\hat{\theta}_n^{\texttt{AD-LO}}$ is similar to $\hat{\theta}_n^{\texttt{AD-BC}}$ and $\hat{\theta}_n^{\texttt{AD-GJ}}$ insofar as it satisfies (7) and has

$$\mathbb{E}[\hat{\theta}_n^{\texttt{AD-LO}}] - \theta_0 \approx \int_{\mathbb{R}^d} K(t)[f_0^{\Delta}(h_n t) - f_0^{\Delta}(0)]dt = O(h_n^{2S}),$$

implying in particular that $\hat{\theta}_n^{\texttt{AD-LO}}$ is asymptotically efficient under Conditions D, K, and B.

The following result collects and summarizes the main findings of this section.

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

THEOREM 1. *Suppose Conditions D, K, and B are satisfied. Then,* $\hat{\theta}_n^{\texttt{AD-BC}}$, $\hat{\theta}_n^{\texttt{AD-GJ}}$, *and* $\hat{\theta}_n^{\texttt{AD-LO}}$ *satisfy (1). If Condition B is strengthened to Condition* B^+, *then* $\hat{\theta}_n^{\texttt{AD}}$ *satisfies (1).*

Remark. Because $\hat{\theta}_n^{\texttt{AD}}$ is a linear functional of $\hat{f}_n$, the generalized jackknife estimator $\hat{\theta}_n^{\texttt{AD-GJ}}$ can be interpreted as a version of the plug-in estimator $\hat{\theta}_n^{\texttt{AD}}$ based on a modified kernel: Defining

$$K^{\texttt{GJ}}(x) = \frac{1}{1-c^d}\left[K(x) - K\left(\frac{x}{c}\right)\right],$$

we have

$$\hat{\theta}_n^{\texttt{AD-GJ}} = \frac{1}{n}\sum_{1\le i\le n}\hat{f}_n^{\texttt{GJ}}(X_i),$$

where

$$\hat{f}_n^{\texttt{AD-GJ}}(x) = \frac{1}{n}\sum_{1\le j\le n}K_n^{\texttt{GJ}}(x-X_j), \qquad K_n^{\texttt{GJ}}(x) = \frac{1}{h_n^d}K^{\texttt{GJ}}\left(\frac{x}{h_n}\right).$$

The modified kernel satisfies $K^{\texttt{GJ}}(0) = 0$, so this interpretation provides an explanation of the fact that $\hat{\theta}_n^{\texttt{AD-GJ}}$ satisfies (6) under Condition B. A similar interpretation is not available for generalized jackknife versions of estimators that are nonlinear functionals of $\hat{f}_n$; examples of such estimators are given by $\hat{\theta}_n^{\texttt{ISD-GJ}}$ and $\hat{\theta}_n^{\texttt{LR-GJ}}$ studied in Section 6.

4. AVERAGE DENSITY ESTIMATORS: BOOTSTRAP CONSISTENCY

Letting $X_{1,n}^*,\dots,X_{n,n}^*$ denote a random sample from the empirical distribution of $X_1,\dots,X_n$, the natural bootstrap analogs of the estimators studied in the previous section are given by

$$\hat{\theta}_n^{\texttt{AD},*} = \frac{1}{n}\sum_{1\le i\le n}\hat{f}_n^*(X_{i,n}^*), \qquad \hat{f}_n^*(x) = \frac{1}{n}\sum_{1\le j\le n}K_n(x-X_{j,n}^*),$$

$$\hat{\theta}_n^{\texttt{AD-BC},*} = \hat{\theta}_n^{\texttt{AD},*} - \frac{K(0)}{nh_n^d},$$

$$\hat{\theta}_n^{\texttt{AD-GJ},*} = \frac{1}{1-c^d}\hat{\theta}_n^{\texttt{AD},*} - \frac{c^d}{1-c^d}\hat{\theta}_n^{\texttt{AD},*}(ch_n),$$

and

$$\hat{\theta}_n^{\texttt{AD-LO},*} = \frac{1}{n}\sum_{1\le i\le n}\hat{f}_{i,n}^{\texttt{LO},*}(X_{i,n}^*), \qquad \hat{f}_{i,n}^{\texttt{LO},*}(x) = \sum_{1\le j\le n}w_{ij,n}K_n(x-X_{j,n}^*),$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

respectively, where $\hat{\theta}_n^{\mathtt{AD},*}(ch_n)$ denotes the version of $\hat{\theta}_n^{\mathtt{AD},*}$ associated with the bandwidth ch_n. The main goal of this section is to explore the extent to which these estimators enjoy the bootstrap consistency property (2) under Conditions D, K, and B.

If $\hat{\theta}_n$ is efficient in the sense that it satisfies (1), then $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{N}(0, \sigma_0^2)$, implying in particular that the bootstrap consistency property (2) admits the following characterization:

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \rightsquigarrow_{\mathbb{P}} \mathcal{N}(0, \sigma_0^2), \tag{9}$$

where $\rightsquigarrow_{\mathbb{P}}$ denotes conditional weak convergence in probability.

Similarly to the analysis of the previous section, it seems natural to base verification of (9) on a decomposition of the bootstrap estimation error $\hat{\theta}_n^* - \hat{\theta}_n$ into its bias and noise components $\mathbb{E}_n^*[\hat{\theta}_n^*] - \hat{\theta}_n$ and $\hat{\theta}_n^* - \mathbb{E}_n^*[\hat{\theta}_n^*]$, where $\mathbb{E}_n^*[\cdot] = \mathbb{E}[\cdot|X_1, \ldots, X_n]$. The resulting sufficient condition for (9) is given by the pair

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^*] - \hat{\theta}_n) = o_{\mathbb{P}}(1) \tag{10}$$

and

$$\sqrt{n}(\hat{\theta}_n^* - \mathbb{E}_n^*[\hat{\theta}_n^*]) \rightsquigarrow_{\mathbb{P}} \mathcal{N}(0, \sigma_0^2), \tag{11}$$

where (10) is the natural bootstrap analog of (6), (11) is a bootstrap version of the main distributional implication of (7), and where (10) is necessary and sufficient for (9) when (11) holds.

In perfect analogy with (7), it turns out that (11) holds under very mild bandwidth conditions. Indeed, under Conditions D and K, the estimators $\hat{\theta}_n^{\mathtt{AD},*}, \hat{\theta}_n^{\mathtt{AD-BC},*}, \hat{\theta}_n^{\mathtt{AD-GJ},*}$, and $\hat{\theta}_n^{\mathtt{AD-LO},*}$ all satisfy (11) whenever Condition B^- holds.[2] As a consequence, the question once again becomes whether the estimators have biases that are sufficiently small. Under Conditions D, K, and B^-, the bootstrap bias of $\hat{\theta}_n^{\mathtt{AD},*}$ satisfies

$$\mathbb{E}_n^*[\hat{\theta}_n^{\mathtt{AD},*}] - \hat{\theta}_n^{\mathtt{AD}} = \frac{K(0)}{nh_n^d} - \frac{1}{n}\hat{\theta}_n^{\mathtt{AD}} = \frac{K(0)}{nh_n^d} + O_{\mathbb{P}}(n^{-1}). \tag{12}$$

Therefore, the bias condition (10) is satisfied by $\hat{\theta}_n^{\mathtt{AD},*}$ provided $nh_n^{2d} \to \infty$. In other words, $\hat{\theta}_n^{\mathtt{AD},*}$ satisfies (2) (and therefore also (3) and (4)) under Conditions D, K, and B^+.

More surprisingly, perhaps, although the estimator $\hat{\theta}_n^{\mathtt{AD-BC}}$ is efficient under Conditions D, K, and B, stronger conditions are required for its bootstrap analog $\hat{\theta}_n^{\mathtt{AD-BC},*}$ to satisfy (2). This is so because

$$\mathbb{E}_n^*[\hat{\theta}_n^{\mathtt{AD-BC},*}] - \hat{\theta}_n^{\mathtt{AD-BC}} = \mathbb{E}_n^*[\hat{\theta}_n^{\mathtt{AD},*}] - \hat{\theta}_n^{\mathtt{AD}} = \frac{K(0)}{nh_n^d} + O_{\mathbb{P}}(n^{-1}) \tag{13}$$

[2]Conversely, Condition B^- is minimal in the sense that the methods of Cattaneo, Crump, and Jansson (2014a) can be used to show that (11) can fail if Condition B^- is violated.

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

under Conditions D, K, and B. A similar remark applies to $\hat{\theta}_n^{\texttt{AD-LO}}$, as its bootstrap analog satisfies

$$\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{AD-LO},*}]-\hat{\theta}_n^{\texttt{AD-LO}}=\hat{\theta}_n^{\texttt{AD}}-\hat{\theta}_n^{\texttt{AD-LO}}=\frac{K(0)}{nh_n^d}+o_{\mathbb{P}}(n^{-1/2})$$

under Conditions D, K, and B.

On the other hand, because

$$\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{AD},*}(ch_n)]-\hat{\theta}_n^{\texttt{AD}}(ch_n)=\frac{1}{c^d}\frac{K(0)}{nh_n^d}-\frac{1}{n}\hat{\theta}_n^{\texttt{AD}}(ch_n)=\frac{1}{c^d}\frac{K(0)}{nh_n^d}+O_{\mathbb{P}}(n^{-1}),$$

the bootstrap analog of $\hat{\theta}_n^{\texttt{AD-GJ}}$ satisfies

$$\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{AD-GJ},*}]-\hat{\theta}_n^{\texttt{AD-GJ}}=O_{\mathbb{P}}(n^{-1}),$$

so this estimator satisfies (2) under Conditions D, K, and B.

It turns out that $\hat{\theta}_n^{\texttt{AD},*}$ satisfies (2)–(4) under conditions that are weaker than the conditions under which $\hat{\theta}_n^{\texttt{AD}}$ is efficient. In generic notation, suppose the estimators $\hat{\theta}_n$ and $\hat{\theta}_n^*$ satisfy (7) and (11), respectively. Then, (2) is still sufficient for (3) and (4) to hold. Moreover, as also observed by Cattaneo and Jansson (2018), the bootstrap consistency condition (2) itself is satisfied under the following generalization of the bias conditions (6) and (10):

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^*]-\hat{\theta}_n)=\sqrt{n}(\mathbb{E}[\hat{\theta}_n]-\theta_0)+o_{\mathbb{P}}(1). \tag{14}$$

Now, as discussed above, the estimators $\hat{\theta}_n^{\texttt{AD}}$ and $\hat{\theta}_n^{\texttt{AD},*}$ satisfy (7) and (11), respectively, under Conditions D, K, and B. Under the same conditions, it follows from (8) and (12) that (14) is satisfied.

The following result collects and summarizes the main findings of this section.

THEOREM 2. *Suppose Conditions D, K, and B are satisfied. Then* $\hat{\theta}_n^{\texttt{AD},*}$ *and* $\hat{\theta}_n^{\texttt{AD-GJ},*}$ *satisfy (2). If Condition B is strengthened to Condition* B^+, *then* $\hat{\theta}_n^{\texttt{AD-BC},*}$ *and* $\hat{\theta}_n^{\texttt{AD-LO},*}$ *satisfy (2).*

Comparing Theorems 1 and 2, we see that efficiency is neither necessary nor sufficient for bootstrap consistency. In fact, the results indicate that there can be a tension between efficiency and bootstrap consistency in semiparametric settings. What seems most noteworthy to us is that whereas "debiased" estimators such as $\hat{\theta}_n^{\texttt{AD-BC}}$ and $\hat{\theta}_n^{\texttt{AD-LO}}$ may appear to be superior to the simple plug-in estimator $\hat{\theta}_n^{\texttt{AD}}$ insofar as they achieve efficiency under weaker (indeed, minimal) conditions, the ranking gets reversed when the estimators are looked at through the lens of the bootstrap. As pointed out by Chen, Linton, and Van Keilegom (2003) and Cheng and Huang (2010), bootstrap-based inference is particularly attractive in semiparametric settings. The results above demonstrate by example that efficiency-based rankings of estimators can be quite misleading in cases where

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

construction of an estimator is simply a means to the end of conducting bootstrap-based inference.

For the estimators under consideration in Theorems 1 and 2 (and elsewhere in this paper), perhaps the easiest way to explain the tension between efficiency and bootstrap consistency is the following. Under mild conditions (namely, under Conditions D, K, and B$^-$), the estimators are efficient if they satisfy (6) and the nonparametric bootstrap is consistent if (14) is satisfied. Both (6) and (14) are bias conditions, but clearly neither condition implies the other in general. As exemplified by $\hat{\theta}_n^{\mathsf{AD},*}, \hat{\theta}_n^{\mathsf{AD\text{-}BC},*}$, and $\hat{\theta}_n^{\mathsf{AD\text{-}LO},*}$, the bootstrap bias $\mathbb{E}_n^*[\hat{\theta}_n^*] - \hat{\theta}_n$ tends to be nonnegligible (i.e., it is not necessarily $o_{\mathbb{P}}(1)$) when only mild smoothness conditions are imposed. Because the conditions (6) and (14) are mutually exclusive whenever the bootstrap bias is nonnegligible, two distinct conclusions can be drawn. First, bootstrap consistency typically fails for efficient estimators. That observation is arguably the main finding of this paper, and we have deliberately chosen to document the finding by means of the simplest possible example(s). Second, nonparametric bootstrap consistency can hold for inefficient estimators as long as the source of the inefficiency is bias. This is precisely what happens for $\hat{\theta}_n^{\mathsf{AD}}$, and, in fact, it turns out that the finding that bootstrap consistency holds for plug-in estimators even if they are not efficient generalizes well beyond the setting of this paper (for details, see Cattaneo and Jansson, 2018).

As conjectured by the Co-Editor, the estimators $\hat{\theta}_n^{\mathsf{AD}}$, $\hat{\theta}_n^{\mathsf{AD\text{-}BC}}, \hat{\theta}_n^{\mathsf{AD\text{-}GJ}}$, and $\hat{\theta}_n^{\mathsf{AD\text{-}LO}}$ can all be shown to satisfy the bootstrap variance consistency property (5) under Conditions D, K, and B. The estimators $\hat{\theta}_n^{\mathsf{AD\text{-}BC}}, \hat{\theta}_n^{\mathsf{AD\text{-}GJ}}$, and $\hat{\theta}_n^{\mathsf{AD\text{-}LO}}$ therefore enjoy the property that the intervals $\mathsf{CI}_{n,1-\alpha}^{\mathsf{N}}$ based on the bootstrap variance estimator are consistent (indeed, efficient) under Conditions D, K, and B.

An important source of the bootstrap consistency result for $\hat{\theta}_n^{\mathsf{AD}}$ is the ability of the bootstrap to automatically perform a bias correction when approximating the distribution of $\hat{\theta}_n^{\mathsf{AD}} - \theta_0$. The same mechanism can be exploited for estimation purposes: Setting $\alpha = 1$, the interval $\mathsf{CI}_{n,1-\alpha}^{\mathsf{P}}$ becomes a singleton and can therefore be interpreted as a bootstrap-based estimator of θ_0. As a by-product of our results about $\hat{\theta}_n^{\mathsf{AD}}$, it can be shown that the resulting estimator

$$2\hat{\theta}_n^{\mathsf{AD}} - \inf\{q \in \mathbb{R} : \mathbb{P}_n^*[\hat{\theta}_n^{\mathsf{AD},*} \leq q] \geq 1/2\}$$

is efficient under Conditions D, K, and B.

The bootstrap analog of $\hat{\theta}_n^{\mathsf{AD}}$ employs a density estimator $\hat{f}_n^*$ that uses the same bandwidth h_n as is used when constructing $\hat{f}_n$. Doing so is important for the purposes of obtaining the bootstrap consistency result for $\hat{\theta}_n^{\mathsf{AD}}$. Indeed, if $\hat{f}_n^*$ were defined using a possibly different bandwidth h_n^* (say), then the bootstrap consistency result under Condition B can fail unless $h_n^*/h_n \to_{\mathbb{P}} 1$. On the other hand, the flavor of the bootstrap results about $\hat{\theta}_n^{\mathsf{AD\text{-}BC}}, \hat{\theta}_n^{\mathsf{AD\text{-}LO}}$, and $\hat{\theta}_n^{\mathsf{AD\text{-}GJ}}$ does not change if a different bandwidth is used when defining their bootstrap analogs.

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

Remark. In important special cases (such as when $\hat{\theta}_n$ equals $\hat{\theta}_n^{\mathtt{AD},*}, \hat{\theta}_n^{\mathtt{AD\text{-}BC},\ *}$, or the leave-one-out version of $\hat{\theta}_n^{\mathtt{AD\text{-}LO},*}$), the fact that $\mathbb{E}_n^*[\hat{\theta}_n^*] - \hat{\theta}_n$ tends to be nonnegligible can be interpreted as a manifestation of the following generic fact about U-statistics: If $X_{1,n}^*, \ldots, X_{n,n}^*$ denotes a random sample from the empirical distribution of $X_1, \ldots, X_n$ and if $\kappa : \mathbb{R}^r \to \mathbb{R}$ is permutation symmetric in its arguments, then

$$\mathbb{E}_n^*[\mathcal{U}_{\kappa,n}^*] = \mathcal{V}_{\kappa,n},$$

where

$$\mathcal{U}_{\kappa,n}^* = \binom{n}{r}^{-1} \sum_{1 \le i_1, \ldots, i_r \le n, i_1 < \cdots < i_r} \kappa\left(X_{i_1,n}^*, \ldots, X_{i_r,n}^*\right)$$

is the rth-order U-statistic (with kernel κ) constructed from $X_{1,n}^*, \ldots, X_{n,n}^*$ and where

$$\mathcal{V}_{\kappa,n} = n^{-r} \sum_{1 \le i_1, \ldots, i_r \le n} \kappa\left(X_{i_1}, \ldots, X_{i_r}\right)$$

is the rth-order V-statistic (with kernel κ) constructed from $X_1, \ldots, X_n$. In other words, under the nonparametric bootstrap distribution, the expected value of a U-statistic is given by the corresponding V-statistic. Whenever $r \ge 2$, the statistic $\mathcal{V}_n$ contains "diagonal" terms (i.e., terms of the form $\kappa\left(X_{i_1}, \ldots, X_{i_r}\right)$ with overlapping subscripts $i_1, \ldots, i_r$) not present in the U-statistic

$$\mathcal{U}_{\kappa,n} = \binom{n}{r}^{-1} \sum_{1 \le i_1, \ldots, i_r \le n, i_1 < \cdots < i_r} \kappa\left(X_{i_1}, \ldots, X_{i_r}\right).$$

It is the presence of such diagonal terms that gives rise to a potentially nonnegligible bias in the bootstrap distribution of estimators that involve U-statistics of order 2 (or greater). Indeed, it is precisely this phenomenon, that is, the source of the celebrated counterexample (to bootstrap consistency) reported in (Bickel and Freedman, 1981, pp. 1209–1210).

5. ALTERNATIVE BOOTSTRAP PROCEDURES

In light of Theorem 2, it is of interest to construct bootstrap-based approximations to the distributions of $\hat{\theta}_n^{\mathtt{AD\text{-}BC}}$ and $\hat{\theta}_n^{\mathtt{AD\text{-}LO}}$ that are consistent under Conditions D, K, and B. In generic notation, suppose $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ is the estimator whose distribution we seek to approximate. One option is to find an estimator $\tilde{\theta}_n = \tilde{\theta}_n(X_1, \ldots, X_n)$ (say) whose natural bootstrap analog $\tilde{\theta}_n^* = \tilde{\theta}_n(X_{1,n}^*, \ldots, X_{n,n}^*)$ satisfies

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \le t] - \mathbb{P}_n^*[\sqrt{n}(\tilde{\theta}_n^* - \hat{\theta}_n) \le t] \right| = o_{\mathbb{P}}(1). \tag{15}$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

As we shall see, both $\hat{\theta}_n^{\texttt{AD-BC}}$ and $\hat{\theta}_n^{\texttt{AD-LO}}$ lend themselves well to a construction of this type. Nevertheless, in some circumstances, it may be equally (if not more) attractive to achieve consistency by finding a bootstrap probability measure $\mathbb{P}_n^\star$ (say) governing the distribution of $X_{1,n}^\star,\ldots,X_{n,n}^\star$ such that $\hat{\theta}_n^\star = \hat{\theta}_n(X_{1,n}^\star,\ldots,X_{n,n}^\star)$ satisfies

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}[\sqrt{n}(\hat{\theta}_n-\theta_0)\le t]-\mathbb{P}_n^\star[\sqrt{n}(\hat{\theta}_n^\star-\hat{\theta}_n)\le t]\right| = o_{\mathbb{P}}(1). \quad \textbf{(16)}$$

A construction of this type turns out to be useful in the case of the cross-fit version of $\hat{\theta}_n^{\texttt{AD-LO}}$.

First, consider the problem of approximating the distribution of $\hat{\theta}_n^{\texttt{AD-BC}}$. It follows from (13) that a bias-corrected version of $\hat{\theta}_n^{\texttt{AD-BC},*}$ is given by

$$\tilde{\theta}_n^{\texttt{AD-BC},*} = \hat{\theta}_n^{\texttt{AD-BC},*} - \frac{K(0)}{nh_n^d}.$$

Rather than showing (15) by analyzing $\tilde{\theta}_n^{\texttt{AD-BC},*}$ directly, we find it more insightful to obtain the consistency result by means of an argument which highlights and exploits the relationship between $\tilde{\theta}_n^{\texttt{AD-BC},*}$ and $\hat{\theta}_n^{\texttt{AD},*}$. Heuristically, $\tilde{\theta}_n^{\texttt{AD-BC},*}$ "should" satisfy (15) under Conditions D, K, and B, because the percentile interval associated with $\tilde{\theta}_n^{\texttt{AD-BC},*}$ is identical to the percentile interval associated with $\hat{\theta}_n^{\texttt{AD},*}$.[3] These heuristics can be made rigorous with the help of the equality

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\left|\mathbb{P}[\sqrt{n}(\hat{\theta}_n^{\texttt{AD-BC}}-\theta_0)\le t]-\mathbb{P}_n^*[\sqrt{n}(\tilde{\theta}_n^{\texttt{AD-BC},*}-\hat{\theta}_n^{\texttt{AD-BC}})\le t]\right|\\ &=\sup_{t\in\mathbb{R}}\left|\mathbb{P}[\sqrt{n}(\hat{\theta}_n^{\texttt{AD}}-\theta_0)\le t]-\mathbb{P}_n^*[\sqrt{n}(\hat{\theta}_n^{\texttt{AD},*}-\hat{\theta}_n^{\texttt{AD}})\le t]\right|,\end{aligned}$$

which implies, in particular, that $\tilde{\theta}_n^{\texttt{AD-BC},*}$ satisfies (15) if and only if $\hat{\theta}_n^{\texttt{AD},*}$ satisfies (2). As a consequence, the fact $\tilde{\theta}_n^{\texttt{AD-BC},*}$ satisfies (15) under Conditions D, K, and B is simply a restatement of the bootstrap consistency result for $\hat{\theta}_n^{\texttt{AD},*}$. Turning next to $\hat{\theta}_n^{\texttt{AD-LO}}$, our preferred modification of this estimator is motivated by the observation that

$$\mathbb{P}[\tilde{f}_{i,n}^{\texttt{LO}}(X_i) = \hat{f}_{i,n}^{\texttt{LO}}(X_i)] = 1,$$

where

$$\tilde{f}_{i,n}^{\texttt{LO}}(x) = \sum_{1\le j\le n} w_{ij,n}\tilde{K}_n(x-X_j), \qquad \tilde{K}_n(x) = \mathbb{1}(x\ne 0)K_n(x).$$

[3] In generic notation, the percentile interval associated with an estimator $\tilde{\theta}_n^*$ is given by

$$\widetilde{\mathsf{CI}}_{n,1-\alpha} = \left[\hat{\theta}_n - \tilde{q}_{n,1-\alpha/2}^*\,,\ \hat{\theta}_n - \tilde{q}_{n,\alpha/2}^*\right], \qquad \tilde{q}_{n,a}^* = \inf\{q\in\mathbb{R} : \mathbb{P}_n^*[(\tilde{\theta}_n^*-\hat{\theta}_n)\le q]\ge a\}.$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

An immediate implication of this observation is that

$$\mathbb{P}[\tilde{\theta}_n^{\texttt{AD-LO}} = \hat{\theta}_n^{\texttt{AD-LO}}] = 1, \qquad \tilde{\theta}_n^{\texttt{AD-LO}} = \frac{1}{n}\sum_{1\le i\le n} \tilde{f}_{i,n}^{\texttt{LO}}(X_i).$$

Nevertheless, unlike $\hat{\theta}_n^{\texttt{AD-LO}}$ itself, the modification $\tilde{\theta}_n^{\texttt{AD-LO}}$ has a natural bootstrap analog

$$\tilde{\theta}_n^{\texttt{AD-LO},*} = \frac{1}{n}\sum_{1\le i\le n} \tilde{f}_{i,n}^{\texttt{LO},*}(X_{i,n}^*), \qquad \tilde{f}_{i,n}^{\texttt{LO},*}(x) = \sum_{1\le j\le n} w_{ij,n}\tilde{K}_n(x - X_{j,n}^*),$$

whose bias is small: Under Conditions D, K, and B,

$$\mathbb{E}_n^*[\tilde{\theta}_n^{\texttt{AD-LO},*}] = \frac{1}{n}\sum_{1\le i\le n} \tilde{f}_n^{\texttt{LO}}(X_i) = \tilde{\theta}_n^{\texttt{AD-LO}} + o_{\mathbb{P}}(n^{-1/2}), \quad \tilde{f}_n^{\texttt{LO}}(x) = \frac{1}{n}\sum_{1\le j\le n} \tilde{K}_n(x - X_j).$$

In fact, it can be shown that (15) is satisfied by $\tilde{\theta}_n^{\texttt{AD-LO},*}$ under Conditions D, K, and B.

For cross-fit estimators, an arguably more attractive option is to construct a bootstrap-based distributional approximation which employs a bootstrap probability measure that is itself of cross-fit (i.e., split sample) type. To illustrate the idea, we consider the simplest special case. When $B_n = 2$, the estimator $\hat{\theta}_n^{\texttt{AD-LO}}$ reduces to

$$\hat{\theta}_n^{\texttt{AD-CF}} = \frac{1}{n}\sum_{1\le i\le n} \hat{f}_{i,n}^{\texttt{CF}}(X_i),$$

where

$$\hat{f}_{i,n}^{\texttt{CF}}(x) = \begin{cases} \frac{1}{n-\lfloor n/2\rfloor}\sum_{\lfloor n/2\rfloor+1\le j\le n} K_n(x - X_j), & i \in \{1,\ldots,\lfloor n/2\rfloor\}, \\ \frac{1}{\lfloor n/2\rfloor}\sum_{1\le j\le \lfloor n/2\rfloor} K_n(x - X_j), & i \in \{\lfloor n/2\rfloor + 1,\ldots,n\}. \end{cases}$$

The $B_n = 2$ version of the "cross-fit bootstrap" is defined as follows. Conditional on $X_1,\ldots,X_n$, let $X_{1,n}^\star,\ldots,X_{n,n}^\star$ be mutually independent with $X_{1,n}^\star,\ldots,X_{\lfloor n/2\rfloor,n}^\star$ being a random sample from the empirical distribution of $X_1,\ldots,X_{\lfloor n/2\rfloor}$ and $X_{\lfloor n/2\rfloor+1,n}^\star,\ldots,X_{n,n}^\star$ being a random sample from the empirical distribution of $X_{\lfloor n/2\rfloor+1},\ldots,X_n$. Then,

$$\hat{\theta}_n^{\texttt{AD-CF},\star} = \frac{1}{n}\sum_{1\le i\le n} \hat{f}_{i,n}^{\texttt{CF},\star}(X_{i,n}^\star)$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

is the corresponding cross-fit bootstrap version of $\hat{\theta}_n^{\texttt{AD-CF}}$, where

$$\hat{f}_{i,n}^{\texttt{CF},\star}(x) = \begin{cases} \frac{1}{n-\lfloor n/2\rfloor}\sum_{\lfloor n/2\rfloor+1\le j\le n} K_n(x-X_{j,n}^{\star}), & i\in\{1,\ldots,\lfloor n/2\rfloor\}, \\ \frac{1}{\lfloor n/2\rfloor}\sum_{1\le j\le \lfloor n/2\rfloor} K_n(x-X_{j,n}^{\star}), & i\in\{\lfloor n/2\rfloor+1,\ldots,n\}. \end{cases}$$

The bootstrap distribution of $\hat{\theta}_n^{\texttt{AD-CF},\star}$ is correctly centered in the sense that $\mathbb{E}_n^{\star}[\hat{\theta}_n^{\texttt{AD-CF},\star}] = \hat{\theta}_n^{\texttt{AD-CF}}$, where $\mathbb{E}_n^{\star}[\cdot]$ denotes the expected value computed under the cross-fit bootstrap distribution. In fact, the bootstrap distribution satisfies (16) under Conditions D, K, and B.

As pointed out by a referee, yet another way of achieving consistency on the part of a bootstrap-based distributional approximation is to center the distribution of $\hat{\theta}_n^*$ at an estimator $\tilde{\theta}_n$ satisfying

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}[\sqrt{n}(\hat{\theta}_n-\theta_0)\le t]-\mathbb{P}_n^*[\sqrt{n}(\hat{\theta}_n^*-\tilde{\theta}_n)\le t]\right| = o_{\mathbb{P}}(1). \tag{17}$$

Because the estimators under consideration here all satisfy (7) and (11), the following analog of (14) is sufficient for (17):

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^*]-\tilde{\theta}_n) = \sqrt{n}(\mathbb{E}[\hat{\theta}_n]-\theta_0)+o_{\mathbb{P}}(1).$$

As already mentioned in connection with (14), the displayed condition is satisfied by $\tilde{\theta}_n = \hat{\theta}_n$ in the case of $\hat{\theta}_n^{\texttt{AD}}$. For the other estimators (i.e., for $\hat{\theta}_n^{\texttt{AD-BC}}$, $\hat{\theta}_n^{\texttt{AD-GJ}}$, and $\hat{\theta}_n^{\texttt{AD-LO}}$), because they satisfy the bias condition (6), the displayed condition is satisfied by $\tilde{\theta}_n = \mathbb{E}_n^*[\hat{\theta}_n^*]$.

6. ALTERNATIVE ESTIMATORS

This section considers two alternative classes of estimators. The first class is motivated by the integrated squared density representation

$$\theta_0 = \int_{\mathbb{R}^d} f_0(x)^2\,dx,$$

an interesting feature of which is that it involves a nonlinear functional of f_0. The second class is motivated by the representation

$$\theta_0 = 2\mathbb{E}[f_0(X)] - \int_{\mathbb{R}^d} f_0(x)^2\,dx,$$

an interesting feature of which is that it is "locally robust"/"Neyman orthogonal" (in the terminology of Chernozhukov et al. (2020)).

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

6.1. Integrated Squared Density Estimators

A kernel-based plug-in integrated squared density estimator is

$$\hat{\theta}_n^{\mathtt{ISD}} = \int_{\mathbb{R}^d} \hat{f}_n(x)^2\, dx.$$

Like $\hat{\theta}_n^{\mathsf{AD}}$, this estimator has a (potentially) nonnegligible bias: Under Conditions D, K, and B,

$$\mathbb{E}[\hat{\theta}_n^{\mathtt{ISD}}] - \theta_0 = \frac{\int_{\mathbb{R}^d} K(u)^2 du}{n h_n^d} + o(n^{-1/2}),$$

where the first term is a "nonlinearity" bias term (in the terminology of Cattaneo et al. (2013)) attributable to the fact that $\hat{\theta}_n^{\mathtt{ISD}}$ is a nonlinear functional of $\hat{f}_n$.

The nonlinearity bias of $\hat{\theta}_n^{\mathtt{ISD}}$ is easily avoidable, a simple bias-corrected version of $\hat{\theta}_n^{\mathtt{ISD}}$ being

$$\hat{\theta}_n^{\mathtt{ISD\text{-}BC}} = \hat{\theta}_n^{\mathtt{ISD}} - \frac{\int_{\mathbb{R}^d} K(u)^2 du}{n h_n^d}.$$

Similarly, because the nonlinearity bias of $\hat{\theta}_n^{\mathtt{ISD}}$ is proportional to $1/(nh_n^d)$, the following generalized jackknife version of $\hat{\theta}_n^{\mathtt{ISD}}$ is an efficient estimator of θ_0:

$$\hat{\theta}_n^{\mathtt{ISD\text{-}GJ}} = \frac{1}{1-c^d}\hat{\theta}_n^{\mathtt{ISD}} - \frac{c^d}{1-c^d}\hat{\theta}_n^{\mathtt{ISD}}(ch_n),$$

where $c \neq 1$ is a user-chosen constant and where $\hat{\theta}_n^{\mathtt{ISD}}(ch_n)$ denotes the version of $\hat{\theta}_n^{\mathtt{ISD}}$ associated with the bandwidth ch_n.

On the other hand, because the source of the nonlinearity bias of $\hat{\theta}_n^{\mathtt{ISD}}$ is different from the source of the leave-in bias of $\hat{\theta}_n^{\mathsf{AD}}$, there is no particular reason to expect leave-out estimators of the form

$$\hat{\theta}_n^{\mathtt{ISD\text{-}LO}} = \frac{1}{n}\sum_{1\le i\le n}\int_{\mathbb{R}^d} \hat{f}_{i,n}^{\mathtt{LO}}(x)^2 dx$$

to have favorable bias properties. Indeed, under Conditions D, K, and B and assuming B_n is proportional to n, we have:[4]

$$\mathbb{E}[\hat{\theta}_n^{\mathtt{ISD\text{-}LO}}] - \theta_0 = \frac{1}{1-B_n^{-1}}\frac{\int_{\mathbb{R}^d} K(u)^2 du}{n h_n^d} + o(n^{-1/2}), \tag{18}$$

[4]More generally (i.e., whether or not B_n is proportional to n), it is shown in the proof of Theorem 3 that the bias expansion is of the form

$$\mathbb{E}[\hat{\theta}_n^{\mathtt{ISD\text{-}LO}}] - \theta_0 = \eta_n \frac{\int_{\mathbb{R}^d} K(u)^2 du}{n h_n^d} + o(n^{-1/2}),$$

where $\eta_n \geq 1$ is bounded.

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

so the nonlinearity bias of $\hat{\theta}_n^{\texttt{ISD-LO}}$ is nonnegligible (and no smaller than that of $\hat{\theta}_n^{\texttt{ISD}}$).

Nevertheless, because θ_0 is a quadratic functional of f_0, the method of "doubly cross-fitting" (in the terminology of Newey and Robins (2018)) can be used to construct an estimator which is free of nonlinearity bias and can be implemented without knowledge of the functional form of the nonlinearity bias. One such estimator is

$$\hat{\theta}_n^{\texttt{ISD-DCF}} = \int_{\mathbb{R}^d} \hat{f}_{1,n}^{\texttt{CF}}(x) \hat{f}_{n,n}^{\texttt{CF}}(x)\, dx,$$

whose bias turns out to be negligible under Conditions D, K, and B.

Under Conditions D, K, and B$^-$, the estimators $\hat{\theta}_n^{\texttt{ISD}}, \hat{\theta}_n^{\texttt{ISD-BC}}, \hat{\theta}_n^{\texttt{ISD-GJ}}, \hat{\theta}_n^{\texttt{ISD-LO}}$, and $\hat{\theta}_n^{\texttt{ISD-DCF}}$ all satisfy (7). As a consequence, we obtain the following integrated squared density counterpart of Theorem 1.

THEOREM 3. *Suppose Conditions D, K, and B are satisfied. Then, $\hat{\theta}_n^{\texttt{ISD-BC}}$, $\hat{\theta}_n^{\texttt{ISD-GJ}}$, and $\hat{\theta}_n^{\texttt{ISD-DCF}}$ satisfy (1). If Condition B is strengthened to Condition B$^+$, then $\hat{\theta}_n^{\texttt{ISD}}$ and $\hat{\theta}_n^{\texttt{ISD-LO}}$ satisfy (1).*

An integrated squared density counterpart of Theorem 2 is also available. Under Conditions D, K, and B, if $\hat{\theta}_n \in \{\hat{\theta}_n^{\texttt{ISD}}, \hat{\theta}_n^{\texttt{ISD-BC}}, \hat{\theta}_n^{\texttt{ISD-LO}}, \hat{\theta}_n^{\texttt{ISD-DCF}}\}$, then its bootstrap analog satisfies (11) and has a bias of the form

$$\mathbb{E}_n^*[\hat{\theta}_n^*] - \hat{\theta}_n = \frac{\int_{\mathbb{R}^d} K(u)^2 du}{nh_n^d} + o_{\mathbb{P}}(n^{-1/2}),$$

so (14) (and therefore also (2)) is satisfied if (and only if)

$$\mathbb{E}[\hat{\theta}_n] - \theta_0 = \frac{\int_{\mathbb{R}^d} K(u)^2 du}{nh_n^d} + o(n^{-1/2}).$$

The latter condition is satisfied by $\hat{\theta}_n^{\texttt{ISD}}$, but violated by $\hat{\theta}_n^{\texttt{ISD-BC}}$ and $\hat{\theta}_n^{\texttt{ISD-DCF}}$. In the case of $\hat{\theta}_n^{\texttt{ISD-LO}}$, it follows from (18) that the condition is satisfied when $B_n = n$ (i.e., when $\hat{\theta}_n^{\texttt{ISD-LO}}$ is a leave-one-out estimator), but violated when B_n is fixed (i.e., when $\hat{\theta}_n^{\texttt{ISD-LO}}$ is a cross-fit estimator).

THEOREM 4. *Suppose Conditions D, K, and B are satisfied. Then, $\hat{\theta}_n^{\texttt{ISD},*}$ and $\hat{\theta}_n^{\texttt{ISD-GJ},*}$ satisfy (2). If $B_n = n$, then $\hat{\theta}_n^{\texttt{ISD-LO},*}$ satisfies (2). If Condition B is strengthened to Condition B$^+$, then $\hat{\theta}_n^{\texttt{ISD-BC},*}, \hat{\theta}_n^{\texttt{ISD-LO},*}$, and $\hat{\theta}_n^{\texttt{ISD-DCF},*}$ satisfy (2).*

In important respects, the results reported in Theorems 3 and 4 are in qualitative agreement with those reported in Theorems 1 and 2. In particular, we find that in spite of being inefficient, the simple plug-in estimator achieves bootstrap consistency under conditions that are weaker than those required for efficient

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

estimators to achieve bootstrap consistency. The most notable difference between the integrated squared density and average derivative estimators is probably that in the case of integrated squared density estimators, the cross-fit estimator is demonstrably worse than the plug-in estimator, satisfying neither (1) nor (2).

As was the case with the average density estimators, the integrated squared density estimators can all be shown to satisfy the bootstrap variance consistency property (5) under Conditions D, K, and B. The estimators $\hat{\theta}_n^{\texttt{ISD-BC}}$, $\hat{\theta}_n^{\texttt{ISD-GJ}}$, and $\hat{\theta}_n^{\texttt{ISD-DCF}}$ therefore enjoy the property that the intervals $\mathsf{CI}_{n,1-\alpha}^{\mathsf{N}}$ based on the bootstrap variance estimator are consistent (indeed, efficient) under Conditions D, K, and B.

For completeness, we conclude this subsection by briefly discussing integrated squared density versions of (15)–(17). In what follows, suppose Conditions D, K, and B are satisfied. A bias-corrected version of $\hat{\theta}_n^{\texttt{ISD-BC},*}$ is given by

$$\tilde{\theta}_n^{\texttt{ISD-BC},*} = \hat{\theta}_n^{\texttt{ISD-BC},*} - \frac{\int_{\mathbb{R}^d} K(u)^2 du}{nh_n^d}.$$

In perfect analogy with $\tilde{\theta}_n^{\texttt{AD-BC},*}$, this estimator satisfies (15), and the associated percentile interval is identical to the percentile interval associated with $\hat{\theta}_n^{\texttt{ISD},*}$. Next,

$$\tilde{\theta}_n^{\texttt{ISD-LO},*} = \frac{1}{n} \sum_{1 \le i \le n} \int_{\mathbb{R}^d} \tilde{f}_{i,n}^{\texttt{LO},*}(x)^2 dx$$

is an integrated squared density counterpart of $\tilde{\theta}_n^{\texttt{AD-LO},*}$. Because $\tilde{\theta}_n^{\texttt{ISD-LO},*} = \hat{\theta}_n^{\texttt{ISD-LO},*}$, this estimator satisfies (15) when $B_n = n$, but not when B_n is fixed. On the other hand, the cross-fit bootstrap can be used when B_n is fixed. As before, suppose $B_n = 2$ for specificity. In that case, $\hat{\theta}_n^{\texttt{ISD-LO}}$ reduces to

$$\hat{\theta}_n^{\texttt{ISD-CF}} = \frac{1}{n} \sum_{1 \le i \le n} \int_{\mathbb{R}^d} \hat{f}_{i,n}^{\texttt{CF}}(x)^2 dx,$$

and it can be shown that

$$\hat{\theta}_n^{\texttt{ISD-CF},\star} = \frac{1}{n} \sum_{1 \le i \le n} \int_{\mathbb{R}^d} \hat{f}_{i,n}^{\texttt{CF},\star}(x)^2 dx$$

satisfies (16). Similarly, the distribution of $\hat{\theta}_n^{\texttt{ISD-DCF}}$ can be approximated using

$$\hat{\theta}_n^{\texttt{ISD-DCF},\star} = \int_{\mathbb{R}^d} \hat{f}_{1,n}^{\texttt{CF},\star}(x) \hat{f}_{n,n}^{\texttt{CF},\star}(x)\, dx,$$

as that estimator satisfies (16). Finally, the property (17) holds for $\hat{\theta}_n^{\texttt{ISD-BC}}, \hat{\theta}_n^{\texttt{ISD-GJ}}$, and $\hat{\theta}_n^{\texttt{ISD-DCF}}$ when $\tilde{\theta}_n = \mathbb{E}_n^*[\hat{\theta}_n^*]$ (and for $\hat{\theta}_n^{\texttt{ISD}}, \hat{\theta}_n^{\texttt{ISD-GJ}}$, and the leave-one-out version of $\hat{\theta}_n^{\texttt{ISD-LO}}$ when $\tilde{\theta}_n = \hat{\theta}_n$).

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

6.2. Locally Robust Estimators

A locally robust kernel-based plug-in estimator of θ_0 is

$$\hat{\theta}_n^{\mathtt{LR}} = \frac{2}{n} \sum_{1 \le i \le n} \hat{f}_n(X_i) - \int_{\mathbb{R}^d} \hat{f}_n(x)^2 \, dx = 2\hat{\theta}_n^{\mathtt{AD}} - \hat{\theta}_n^{\mathtt{ISD}}.$$

Because $\hat{\theta}_n^{\mathtt{LR}}$ is a linear combination of $\hat{\theta}_n^{\mathtt{AD}}$ and $\hat{\theta}_n^{\mathtt{ISD}}$, its properties follow directly from the results obtained in the previous sections, as do the properties of estimators such as

$$\hat{\theta}_n^{\mathtt{LR\text{-}BC}} = 2\hat{\theta}_n^{\mathtt{AD\text{-}BC}} - \hat{\theta}_n^{\mathtt{ISD\text{-}BC}},$$

$$\hat{\theta}_n^{\mathtt{LR\text{-}GJ}} = 2\hat{\theta}_n^{\mathtt{AD\text{-}GJ}} - \hat{\theta}_n^{\mathtt{ISD\text{-}GJ}},$$

and

$$\hat{\theta}_n^{\mathtt{LR\text{-}LO}} = 2\hat{\theta}_n^{\mathtt{AD\text{-}LO}} - \hat{\theta}_n^{\mathtt{ISD\text{-}LO}},$$

the cross-fit version of the latter being the only estimator (in this paper) satisfying both of the defining properties of the "double/debiased machine learning" estimators proposed by Chernozhukov et al. (2018).

Once again, the results are in qualitative agreement with those reported in Theorems 1 and 2.

THEOREM 5. *Suppose Conditions D, K, and B are satisfied. Then, $\hat{\theta}_n^{\mathtt{LR\text{-}BC}}$ and $\hat{\theta}_n^{\mathtt{LR\text{-}GJ}}$ satisfy (1). If Condition B is strengthened to Condition B^+, then $\hat{\theta}_n^{\mathtt{LR}}$ and $\hat{\theta}_n^{\mathtt{LR\text{-}LO}}$ satisfy (1).*

THEOREM 6. *Suppose Conditions D, K, and B are satisfied. Then, $\hat{\theta}_n^{\mathtt{LR},*}$ and $\hat{\theta}_n^{\mathtt{LR\text{-}GJ},*}$ satisfy (2). If Condition B is strengthened to Condition B^+, then $\hat{\theta}_n^{\mathtt{LR\text{-}BC},*}$ and $\hat{\theta}_n^{\mathtt{LR\text{-}LO},*}$ satisfy (2).*

Rather than spelling out those locally robust versions of (15)–(17) that follow directly from our earlier results, it seems more constructive to mention a feature of local robustness that is particularly useful for bootstrap purposes. As pointed out by Belloni et al. (2017), a notable feature of locally robust moment conditions is that in two-step estimation settings, one does not need to recompute the first step estimator in each iteration of the bootstrap. In the case of $\hat{\theta}_n^{\mathtt{LR}}$, this implies that (15) can be achieved with the help of

$$\tilde{\theta}_n^{\mathtt{LR},*} = \frac{2}{n} \sum_{1 \le i \le n} \hat{f}_n(X_{i,n}^*) - \int_{\mathbb{R}^d} \hat{f}_n(x)^2 \, dx,$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

a computationally attractive feature of which is that $\hat{f}_n$ is kept fixed across bootstrap repetitions. Perhaps more importantly (for our purposes at least), the fact that $\hat{f}_n$ is kept fixed actually makes it easier to achieve (15) also in the case of debiased estimators. For instance,

$$\tilde{\theta}_n^{\texttt{LR-BC},*} = \frac{2}{n} \sum_{1 \le i \le n} \hat{f}_n(X_{i,n}^*) - \int_{\mathbb{R}^d} \hat{f}_n(x)^2 \, dx - \frac{2K(0) - \int_{\mathbb{R}^d} K(u)^2 du}{nh_n^d}$$

satisfies (15) under Conditions D, K, and B.

7. CONCLUDING REMARKS

Among other things, this paper has demonstrated by example that the nonparametric bootstrap can fail to provide a consistent approximation to the distribution of debiased versions of two-step semiparametric estimators. Reasonable people can disagree about whether this is a shortcoming of the nonparametric bootstrap and/or popular debiasing methods, but either way this finding has potentially important implications for econometric practice and it would therefore be of interest to explore the extent to which similar results are available for estimators other than those considered in this paper.

In addition to the cautionary tale about debiasing, the paper contains at least three constructive observations. First, the nonparametric bootstrap variance estimator can be consistent even if the corresponding distributional approximation is not, so valid confidence intervals can be obtained by combining a (successful) debiasing method with a standard error computed by means of the nonparametric bootstrap. Second, the (apparently novel) cross-fit bootstrap consistently estimates the distribution of cross-fit estimators in all the cases considered. Third, estimators based on generalized jackknifing are both efficient and satisfy bootstrap consistency under weak conditions. It seems plausible that all three findings generalize well beyond the average density setting, but it is beyond the scope of this paper to substantiate that conjecture.

At a more abstract level, this paper highlights the importance of paying careful attention to first moments (e.g., bias properties) when diagnosing bootstrap success. It seems noteworthy that the "heuristically necessary" bias condition (14) is also sufficient for bootstrap consistency under mild conditions. A similar phenomenon occurs for estimators of maximum score type. For such estimators, Cattaneo, Jansson, and Nagasawa (2020) achieved bootstrap consistency effectively by paying careful attention to the first moment properties of a certain stochastic process. It would appear that similar heuristics can assist the construction of valid bootstrap-based distributional approximations in other contexts (e.g., shape-constrained nonparametric estimation), but again it is beyond the scope of this paper to substantiate that conjecture.

REFERENCES

Belloni, A., V. Chernozhukov, I. Fernández-Val, & C. Hansen (2017) Program evaluation and causal inference with high-dimensional data. *Econometrica* 85, 233–298.

Bickel, P.J. & D.A. Freedman (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics* 9, 1196–1217.

Bickel, P.J. & Y. Ritov (1988) Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā: The Indian Journal of Statistics, Series A* 50, 381–393.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2013) Generalized Jackknife estimators of weighted average derivatives (with discussion and rejoinder). *Journal of the American Statistical Association* 108, 1243–1268.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2014a) Bootstrapping density-weighted average derivatives. *Econometric Theory* 30, 1135–1164.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2014b) Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory* 30, 176–200.

Cattaneo, M.D. & M. Jansson (2018) Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica* 86, 955–995.

Cattaneo, M.D., M. Jansson, & K. Nagasawa (2020) Bootstrap-based inference for cube root asymptotics. *Econometrica* 88, 2203–2219.

Chen, X., O. Linton, & I. Van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71, 1591–1608.

Cheng, G. & J.Z. Huang (2010) Bootstrap consistency for general semiparametric M-estimation. *Annals of Statistics* 38, 2884–2915.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C.B. Hansen, W.K. Newey, & J.M. Robins (2018) Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21, C1–C68.

Chernozhukov, V., J.C. Escanciano, H. Ichimura, W.K. Newey, & J.M. Robins (2020) Locally robust semiparametric estimation. Preprint, arXiv:1608.00033.

Giné, E. & R. Nickl (2008a) A simple adaptive estimator of the integrated square of a density. *Bernoulli* 14, 47–61.

Giné, E. & R. Nickl (2008b) Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields* 141, 333–387.

Hall, P. & J.S. Marron (1987) Estimation of integrated squared density derivatives. *Statistics and Probability Letters* 6, 109–115.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245. North-Holland.

Newey, W.K. & J.M. Robins (2018) Cross-fitting and fast remainder rates for semiparametric estimation. Preprint, arXiv:1801.09138.

Pfanzagl, J. (1982) *Contributions to a General Asymptotic Statistical Theory*. Springer.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.

Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.

Ritov, Y. & P.J. Bickel (1990) Achieving information bounds in non and semiparametric models. *Annals of Statistics* 18, 925–938.

Tsybakov, A.T. (2009) *Introduction to Nonparametric Estimation*. Springer.

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

APPENDIX

A.1. Hoeffding Decompositions

Each of the estimators studied in this paper has a V-statistic-type representation of the form

$$\hat{\theta}_n = \frac{1}{n^2} \sum_{1 \le i,j \le n} V_{ij,n},$$

where $V_{ij,n}$ depends on $X_1, \ldots, X_n$ only through (X_i, X_j). The proofs of Theorems 1, 3, and 5 are based on the associated Hoeffding decomposition of $\hat{\theta}_n - \theta_0$ given by

$$\hat{\theta}_n - \theta_0 = \beta_n + \frac{1}{n} \sum_{1 \le i \le n} L_{i,n} + \frac{2}{n(n-1)} \sum_{1 \le i,j \le n, i<j} W_{ij,n}, \tag{A.1}$$

where, defining $\bar{V}_{ij,n} = (V_{ij,n} + V_{ji,n})/2$,

$$\begin{aligned}
\beta_n &= \mathbb{E}[\hat{\theta}_n] - \theta_0 \\
&= \frac{1}{n} \left\{ \frac{1}{n} \sum_{1 \le i \le n} \mathbb{E}[V_{ii,n}] \right\} + \left(1 - \frac{1}{n}\right) \left\{ \frac{2}{n(n-1)} \sum_{1 \le i,j \le n, i<j} \mathbb{E}[\bar{V}_{ij,n}] \right\} - \theta_0,
\end{aligned}$$

$$\begin{aligned}
L_{i,n} &= n\{\mathbb{E}[\hat{\theta}_n|X_i] - \mathbb{E}[\hat{\theta}_n]\} \\
&= \frac{1}{n}\{V_{ii,n} - \mathbb{E}[V_{ii,n}]\} + \frac{1}{n-1} \sum_{1 \le j \le n, j \ne i} 2\frac{n-1}{n}\{\mathbb{E}[\bar{V}_{ij,n}|X_i] - \mathbb{E}[\bar{V}_{ij,n}]\},
\end{aligned}$$

$$\begin{aligned}
W_{ij,n} &= \frac{n(n-1)}{2}\{\mathbb{E}[\hat{\theta}_n|X_i,X_j] - \mathbb{E}[\hat{\theta}_n|X_i] - \mathbb{E}[\hat{\theta}_n|X_j] + \mathbb{E}[\hat{\theta}_n]\} \\
&= \frac{n-1}{n}\{\bar{V}_{ij,n} - \mathbb{E}[\bar{V}_{ij,n}|X_i] - \mathbb{E}[\bar{V}_{ij,n}|X_j] + \mathbb{E}[\bar{V}_{ij,n}]\}.
\end{aligned}$$

By construction, $L_{i,n}$ and $W_{ij,n}$ depend on $X_1, \ldots, X_n$ only through X_i and (X_i, X_j), respectively, and satisfy, for each $1 \le i,j \le n$ with $i \ne j$,

$$\mathbb{E}[L_{i,n}] = \mathbb{E}[W_{ij,n}|X_i] = \mathbb{E}[W_{ij,n}|X_j] = 0.$$

Moreover, if the $V_{ij,n}$ satisfy $V_{ii,n} = \delta_n$ and $\mathbb{E}[V_{ij,n}] = \theta_n$, then the bias is of the form

$$\beta_n = \frac{\delta_n}{n} + \theta_n - \theta_0 - \frac{\theta_n}{n}.$$

If also $V_{ij,n} = V_{ji,n}$ and $\mathbb{E}[V_{ij,n}|X_i] = f_n(X_i)$, then

$$L_{i,n} = 2\frac{n-1}{n}\{f_n(X_i) - \theta_n\}, \qquad W_{ij,n} = \frac{n-1}{n}\{V_{ij,n} - f_n(X_i) - f_n(X_j) + \theta_n\}.$$

A bootstrap analog of (A.1) will be employed in the proofs of Theorems 2, 4, and 6. To state it, suppose

$$\hat{\theta}_n^* = \frac{1}{n^2} \sum_{1 \le i,j \le n} V_{ij,n}^*,$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

where $V^*_{ij,n}$ depends on $X^*_{1,n},\ldots,X^*_{n,n}$ only through $(X^*_{i,n},X^*_{j,n})$. Then,

$$\hat{\theta}^*_n-\hat{\theta}_n=\beta^*_n+\frac{1}{n}\sum_{1\le i\le n}L^*_{i,n}+\frac{2}{n(n-1)}\sum_{1\le i,j\le n,\,i<j}W^*_{ij,n}, \tag{A.2}$$

where, defining $\bar{V}^*_{ij,n}=(V^*_{ij,n}+V^*_{ji,n})/2$,

$$\begin{aligned}\beta^*_n&=\mathbb{E}^*_n[\hat{\theta}^*_n]-\hat{\theta}_n\\&=\frac{1}{n}\left\{\frac{1}{n}\sum_{1\le i\le n}\mathbb{E}^*_n[V^*_{ii,n}]\right\}+\left(1-\frac{1}{n}\right)\left\{\frac{2}{n(n-1)}\sum_{1\le i,j\le n,\,i<j}\mathbb{E}^*_n[\bar{V}^*_{ij,n}]\right\}-\hat{\theta}_n,\end{aligned}$$

$$\begin{aligned}L^*_{i,n}&=n\{\mathbb{E}^*_n[\hat{\theta}^*_n|X^*_{i,n}]-\mathbb{E}^*_n[\hat{\theta}^*_n]\}\\&=\frac{1}{n}\{V^*_{ii,n}-\mathbb{E}^*_n[V^*_{ii,n}]\}+\frac{1}{n-1}\sum_{1\le j\le n,\,j\ne i}2\frac{n-1}{n}\{\mathbb{E}^*_n[\bar{V}^*_{ij,n}|X^*_{i,n}]-\mathbb{E}^*_n[\bar{V}^*_{ij,n}]\},\end{aligned}$$

$$\begin{aligned}W^*_{ij,n}&=\frac{n(n-1)}{2}\{\mathbb{E}^*_n[\hat{\theta}^*_n|X^*_{i,n},X^*_{j,n}]-\mathbb{E}^*_n[\hat{\theta}^*_n|X^*_{i,n}]-\mathbb{E}^*_n[\hat{\theta}^*_n|X^*_{j,n}]+\mathbb{E}^*_n[\hat{\theta}^*_n]\}\\&=\frac{n-1}{n}\{\bar{V}^*_{ij,n}-\mathbb{E}^*_n[\bar{V}^*_{ij,n}|X^*_{i,n}]-\mathbb{E}^*_n[\bar{V}^*_{ij,n}|X^*_{j,n}]+\mathbb{E}^*_n[\bar{V}^*_{ij,n}]\}.\end{aligned}$$

By construction, $L^*_{i,n}$ and $W^*_{ij,n}$ depend on $X^*_{1,n},\ldots,X^*_{n,n}$ only through $X^*_{i,n}$ and $(X^*_{i,n},X^*_{j,n})$, respectively, and satisfy, for each $1\le i,j\le n$ with $i\ne j$,

$$\mathbb{E}^*_n[L^*_{i,n}]=\mathbb{E}^*_n[W^*_{ij,n}|X^*_{i,n}]=\mathbb{E}^*_n[W^*_{ij,n}|X^*_{j,n}]=0.$$

Moreover, if the $V^*_{ij,n}$ satisfy $V^*_{ii,n}=\delta^*_n$ and $\mathbb{E}^*_n[V^*_{ij,n}]=\theta^*_n$, then the bootstrap bias is of the form

$$\beta^*_n=\frac{\delta^*_n}{n}+\theta^*_n-\hat{\theta}_n-\frac{\theta^*_n}{n}.$$

If also $V^*_{ij,n}=V^*_{ji,n}$ and $\mathbb{E}^*_n[V^*_{ij,n}|X^*_{i,n}]=f^*_n(X^*_{i,n})$, then

$$L^*_{i,n}=2\frac{n-1}{n}\{f^*_n(X^*_{i,n})-\theta^*_n\},\qquad W^*_{ij,n}=\frac{n-1}{n}\{V^*_{ij,n}-f^*_n(X^*_{i,n})-f^*_n(X^*_{j,n})+\theta^*_n\}.$$

A.2. Proof of Theorem 1

The estimators $\hat{\theta}^{\mathtt{AD}}_n$ and $\hat{\theta}^{\mathtt{AD-LO}}_n$ both have Hoeffding decompositions of the form (A.1), with

$$L_{i,n}=\lambda_{i,n}L^{\mathtt{AD}}_n(X_i)\qquad\text{and}\qquad W_{ij,n}=\omega_{ij,n}W^{\mathtt{AD}}_n(X_i,X_j),$$

where $\lambda_{i,n}$ and $\omega_{ij,n}$ are (nonrandom) estimator-specific weights, while

$$L_n^{\mathsf{AD}}(x) = 2\{f_n^{\mathsf{AD}}(x) - \theta_n^{\mathsf{AD}}\},$$

$$W_n^{\mathsf{AD}}(x_1,x_2) = K_n(x_1 - x_2) - f_n^{\mathsf{AD}}(x_1) - f_n^{\mathsf{AD}}(x_2) + \theta_n^{\mathsf{AD}},$$

where

$$f_n^{\mathsf{AD}}(x) = \mathbb{E}[K_n(x - X)] = \int_{\mathbb{R}^d} K(u) f_0(x + uh_n)du,$$

$$\theta_n^{\mathsf{AD}} = \mathbb{E}[f_n^{\mathsf{AD}}(X)] = \int_{\mathbb{R}^d} f_n^{\mathsf{AD}}(x) f_0(x)dx.$$

To be specific, in the case of

$$\hat{\theta}_n^{\mathsf{AD}} = \frac{1}{n}\sum_{1\le i\le n} \hat{f}_n(X_i) = \frac{1}{n^2}\sum_{1\le i,j\le n} K_n(X_i - X_j),$$

each of $\lambda_{i,n}$ and $\omega_{ij,n}$ is given by $1 - n^{-1}$, while the weights for

$$\hat{\theta}_n^{\mathsf{AD\text{-}LO}} = \frac{1}{n}\sum_{1\le i\le n} \hat{f}_{i,n}^{\mathsf{LO}}(X_i) = \frac{1}{n^2}\sum_{1\le i,j\le n} n w_{ij,n} K_n(X_i - X_j)$$

are of the form

$$\lambda_{i,n} = \sum_{1\le j\le n} \bar{w}_{ij,n}, \qquad \omega_{ij,n} = (n-1)\bar{w}_{ij,n}, \qquad \bar{w}_{ij,n} = (w_{ij,n} + w_{ji,n})/2.$$

In both cases, the weights satisfy

$$\max_{1\le i\le n} (\lambda_{i,n} - 1)^2 = o(1) \tag{A.3}$$

and

$$\max_{1\le i<j\le n} \omega_{ij,n}^2 = O(1). \tag{A.4}$$

It therefore follows from simple moment calculations that the estimators satisfy (7) if

$$\frac{1}{n}\mathbb{E}[W_n^{\mathsf{AD}}(X_1,X_2)^2] \to 0 \tag{A.5}$$

and if

$$\mathbb{E}[\{L_n^{\mathsf{AD}}(X) - L_0(X)\}^2] \to 0. \tag{A.6}$$

Suppose Conditions D and K are satisfied. Then, (A.5) holds if $nh_n^d \to \infty$, because then

$$\begin{aligned}\frac{1}{n}\mathbb{E}[W_n^{\mathsf{AD}}(X_1,X_2)^2] &\le \frac{1}{nh_n^d}\left\{h_n^d \mathbb{E}[K_n(X_1 - X_2)^2]\right\}\\ &= \frac{1}{nh_n^d}\left\{h_n^d \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K_n(u-v)^2 f_0(u) f_0(v)dudv\right\}\end{aligned}$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

$$
\begin{aligned}
&= \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(t)^2 f_0(v+h_n t) f_0(v) dt dv \\
&\leq \frac{1}{nh_n^d} \left\{ \sup_{u \in \mathbb{R}^d} |K(u)| \right\} \left\{ \sup_{x \in \mathbb{R}^d} f_0(x) \right\} \int_{\mathbb{R}^d} |K(u)| du \to 0.
\end{aligned}
$$

Also, because

$$
\mathbb{E}[\{L_n^{\mathsf{AD}}(X) - L_0(X)\}^2] \leq 4\mathbb{E}\left[\{f_n^{\mathsf{AD}}(X) - f_0(X)\}^2\right],
$$

a sufficient condition for (A.6) to hold is that

$$
\mathbb{E}\left[\{f_n^{\mathsf{AD}}(X) - f_0(X)\}^2\right] \to 0.
$$

As in Proposition 1(c) of Giné and Nickl (2008b), the displayed condition is satisfied if $h_n \to 0$. To summarize, each estimator satisfies (7) under Conditions D, K, and B$^-$.

The proof will be completed by giving conditions under which the estimators satisfy (6). As before, suppose Conditions D and K are satisfied. In the notation introduced above, the biases of $\hat{\theta}_n^{\mathsf{AD}}$ and $\hat{\theta}_n^{\mathsf{AD\text{-}LO}}$ are given by

$$
\beta_n^{\mathsf{AD}} = \frac{K(0)}{nh_n^d} + \theta_n^{\mathsf{AD}} - \theta_0 - \frac{\theta_n^{\mathsf{AD}}}{n}
$$

and

$$
\beta_n^{\mathsf{AD\text{-}LO}} = \theta_n^{\mathsf{AD}} - \theta_0,
$$

respectively. Following Giné and Nickl (2008a), we base our analysis of the smoothing bias $\theta_n^{\mathsf{AD}} - \theta_0$ on the representation

$$
\begin{aligned}
\theta_n^{\mathsf{AD}} &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_n(u-v) f_0(v) f_0(u) du dv \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(t) f_0(u-h_n t) f_0(u) du dt \\
&= \int_{\mathbb{R}^d} K(t) f_0^{\Delta}(h_n t) dt,
\end{aligned}
$$

where the last equality uses the fact that K is even. By Lemma 12 of Giné and Nickl (2008b), the function f_0^{Δ} belongs to the Hö lder space $\mathbf{C}^{2s}(\mathbb{R}^d)$. As a consequence, it follows from standard arguments (e.g., Tsybakov, 2009, Prop. 1.2) that if Condition B is satisfied, then

$$
\theta_n^{\mathsf{AD}} - \theta_0 = \int_{\mathbb{R}^d} K(t)[f_0^{\Delta}(h_n t) - f_0^{\Delta}(0)] dt = O(h_n^S) = o(n^{-1/2}).
$$

In particular, $\hat{\theta}_n^{\mathsf{AD\text{-}LO}}$ satisfies (6) under Conditions D, K, and B. Under the same conditions, θ_n^{AD} is bounded, so

$$
\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\mathsf{AD}}] - \theta_0) = \frac{K(0)}{\sqrt{nh_n^{2d}}} + o(1),
$$

implying in particular that Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\mathsf{AD}}$ to satisfy (6) (unless $K(0) = 0$).

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

Finally, the results for $\hat{\theta}_n^{\mathtt{AD\text{-}BC}}$ and $\hat{\theta}_n^{\mathtt{AD\text{-}GJ}}$ follow from those for $\hat{\theta}_n^{\mathtt{AD}}$. To be specific, $\hat{\theta}_n^{\mathtt{AD\text{-}BC}}$ differs from $\hat{\theta}_n^{\mathtt{AD}}$ by an additive constant, so it satisfies (7) under Conditions D, K, and B$^-$. Also, the additive constant is designed to ensure that (6) is satisfied by $\hat{\theta}_n^{\mathtt{AD\text{-}BC}}$ under Conditions D, K, and B. Similarly, because

$$\frac{1}{1-c^d} - \frac{c^d}{1-c^d} = 1,$$

the estimator $\hat{\theta}_n^{\mathtt{AD\text{-}GJ}}$ satisfies (7) under Conditions D, K, and B$^-$, while the fact that

$$\frac{1}{1-c^d}\frac{1}{nh_n^d} - \frac{c^d}{1-c^d}\frac{1}{n(ch_n)^d} = 0$$

ensures that (6) is satisfied by $\hat{\theta}_n^{\mathtt{AD\text{-}GJ}}$ under Conditions D, K, and B.

A.3. Proof of Theorem 2

The estimators $\hat{\theta}_n^{\mathtt{AD},*}$ and $\hat{\theta}_n^{\mathtt{AD\text{-}LO},*}$ both have Hoeffding decompositions of the form (A.2), with

$$L_{i,n}^* = \lambda_{i,n}\hat{L}_n^{\mathtt{AD}}(X_{i,n}^*) \qquad \text{and} \qquad W_{ij,n}^* = \omega_{ij,n}\hat{W}_n^{\mathtt{AD}}(X_{i,n}^*, X_{j,n}^*),$$

where $\lambda_{i,n}$ and $\omega_{ij,n}$ are the same as those for $\hat{\theta}_n^{\mathtt{AD}}$ and $\hat{\theta}_n^{\mathtt{AD\text{-}LO}}$, while

$$\hat{L}_n^{\mathtt{AD}}(x) = 2\{\hat{f}_n(x) - \hat{\theta}_n^{\mathtt{AD}}\},$$

$$\hat{W}_n^{\mathtt{AD}}(x_1,x_2) = K_n(x_1-x_2) - \hat{f}_n(x_1) - \hat{f}_n(x_2) + \hat{\theta}_n^{\mathtt{AD}}.$$

Because the weights satisfy (A.3) and (A.4), it follows from simple moment calculations that the estimators satisfy

$$\sqrt{n}(\hat{\theta}_n^* - \mathbb{E}_n^*[\hat{\theta}_n^*]) = \frac{1}{\sqrt{n}}\sum_{1\le i\le n}\{L_0(X_{i,n}^*) - \mathbb{E}_n^*[L_0(X_{i,n}^*)]\} + o_{\mathbb{P}}(1) \rightsquigarrow_{\mathbb{P}} \mathcal{N}(0,\sigma_0^2)$$

if

$$\frac{1}{n}\mathbb{E}_n^*[\hat{W}_n^{\mathtt{AD}}(X_{1,n}^*, X_{2,n}^*)^2] \to_{\mathbb{P}} 0 \tag{A.7}$$

and if (A.6) and (A.8) hold, where

$$\mathbb{E}_n^*[\{\hat{L}_n^{\mathtt{AD}}(X_{1,n}^*) - L_n^{\mathtt{AD}}(X_{1,n}^*)\}^2] \to_{\mathbb{P}} 0. \tag{A.8}$$

Suppose Conditions D and K are satisfied. Then, (A.7) holds if $nh_n^d \to \infty$, because then

$$\begin{aligned}\frac{1}{n}\mathbb{E}_n^*[\hat{W}_n^{\mathtt{AD}}(X_{1,n}^*, X_{2,n}^*)^2] &\le \frac{1}{n}\mathbb{E}_n^*[K_n(X_{1,n}^* - X_{2,n}^*)^2] \\ &= \frac{1}{n^3}\sum_{1\le i,j\le n} K_n(X_i - X_j)^2\end{aligned}$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

$$= \frac{1}{n^3} \sum_{1 \le i \le n} K_n(0)^2 + \frac{2}{n^3} \sum_{1 \le i,j \le n, i<j} K_n(X_i - X_j)^2$$
$$= \frac{1}{n} \left(\frac{K(0)}{nh_n^d} \right)^2 + O_{\mathbb{P}} \left(\frac{1}{n} \mathbb{E}[K_n(X_1 - X_2)^2] \right) \to_{\mathbb{P}} 0,$$

where the convergence result follows from the proof of Theorem 2. In that same proof, it was shown that (A.6) holds when $h_n \to 0$. Finally, because

$$\mathbb{E}_n^*[\{\hat{L}_n^{\mathsf{AD}}(X_{1,n}^*) - L_n^{\mathsf{AD}}(X_{1,n}^*)\}^2] = \frac{1}{n} \sum_{1 \le i \le n} \{\hat{L}_n^{\mathsf{AD}}(X_i) - L_n^{\mathsf{AD}}(X_i)\}^2,$$

a sufficient condition for (A.8) to hold is that

$$\mathbb{E}[\{\hat{L}_n^{\mathsf{AD}}(X_1) - L_n^{\mathsf{AD}}(X_1)\}^2] \to 0.$$

It follows from a direct calculation this condition is satisfied when $h_n \to 0$ and $nh_n^d \to \infty$. To summarize, each estimator satisfies (11) under Conditions D, K, and B$^-$.

The proof will be completed by giving conditions under which the estimators satisfy (14). Suppose Conditions D, K, and B are satisfied. By the proof of Theorem 1,

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\mathsf{AD}}] - \theta_0) = \frac{K(0)}{\sqrt{nh_n^{2d}}} + o(1),$$

and

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\mathsf{AD\text{-}LO}}] - \theta_0) = o(1),$$

while it follows from (A.2) and Theorem 1 that

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\mathsf{AD},*}] - \hat{\theta}_n^{\mathsf{AD}}) = \frac{K(0)}{\sqrt{nh_n^{2d}}} - \frac{\hat{\theta}_n^{\mathsf{AD}}}{\sqrt{n}} = \frac{K(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1),$$

and

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\mathsf{AD\text{-}LO},*}] - \hat{\theta}_n^{\mathsf{AD\text{-}LO}}) = \sqrt{n}(\hat{\theta}_n^{\mathsf{AD}} - \hat{\theta}_n^{\mathsf{AD\text{-}LO}}) = \frac{K(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1).$$

As a consequence, $\hat{\theta}_n^{\mathsf{AD},*}$ satisfies (14) under Conditions D, K, and B, whereas Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\mathsf{AD\text{-}LO},*}$ to satisfy (14) (unless $K(0) = 0$).

Finally, the results for $\hat{\theta}_n^{\mathsf{AD\text{-}BC},*}$ and $\hat{\theta}_n^{\mathsf{AD\text{-}GJ},*}$ follow from those for $\hat{\theta}_n^{\mathsf{AD},*}$. To be specific, $\hat{\theta}_n^{\mathsf{AD\text{-}BC},*}$ satisfies (11) under Conditions D, K, and B$^-$, because $\hat{\theta}_n^{\mathsf{AD},*}$ does. Moreover,

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\mathsf{AD\text{-}BC},*}] - \hat{\theta}_n^{\mathsf{AD\text{-}BC}}) = \sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\mathsf{AD},*}] - \hat{\theta}_n^{\mathsf{AD}}),$$

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

so under Conditions D and K, Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{AD-BC},*}$ to satisfy (14) (unless $K(0) = 0$). Similarly, because

$$\frac{1}{1-c^d} - \frac{c^d}{1-c^d} = 1,$$

the estimator $\hat{\theta}_n^{\texttt{AD-GJ},*}$ satisfies (11) under Conditions D, K, and B$^-$, while the fact that

$$\frac{1}{1-c^d}\frac{1}{nh_n^d} - \frac{c^d}{1-c^d}\frac{1}{n(ch_n)^d} = 0$$

ensures that (14) is satisfied by $\hat{\theta}_n^{\texttt{AD-GJ},*}$ under Conditions D, K, and B.

A.4. Proof of Theorem 3

The proof is similar to that of Theorem 1. The estimators $\hat{\theta}_n^{\texttt{ISD}}$, $\hat{\theta}_n^{\texttt{ISD-LO}}$, and $\hat{\theta}_n^{\texttt{ISD-CF}}$ all have Hoeffding decompositions of the form (A.1), with

$$L_{i,n} = \lambda_{i,n} L_n^{\texttt{ISD}}(X_i), \qquad W_{ij,n} = \omega_{ij,n} W_n^{\texttt{ISD}}(X_i, X_j),$$

where $\lambda_{i,n}$ and $\omega_{ij,n}$ are (nonrandom) estimator-specific weights, while

$$L_n^{\texttt{ISD}}(x) = 2\{f_n^{\texttt{ISD}}(x) - \theta_n^{\texttt{ISD}}\},$$

$$W_n^{\texttt{ISD}}(x_1, x_2) = K_n^{\Delta}(x_1 - x_2) - f_n^{\texttt{ISD}}(x_1) - f_n^{\texttt{ISD}}(x_2) + \theta_n^{\texttt{ISD}},$$

where

$$f_n^{\texttt{ISD}}(x) = \mathbb{E}[K_n^{\Delta}(x - X)] = \int_{\mathbb{R}^d} K^{\Delta}(u) f_0(x + uh_n) du,$$

$$\theta_n^{\texttt{ISD}} = \mathbb{E}[f_n^{\texttt{ISD}}(X)] = \int_{\mathbb{R}^d} f_n^{\texttt{ISD}}(x) f_0(x) dx,$$

$$K_n^{\Delta}(x) = \frac{1}{h_n^d} K^{\Delta}\left(\frac{x}{h_n}\right), \qquad K^{\Delta}(x) = \int_{\mathbb{R}^d} K(u) K(x+u) du.$$

To be specific, in the case of

$$\begin{aligned}
\hat{\theta}_n^{\texttt{ISD}} &= \int_{\mathbb{R}^d} \hat{f}_n(x)^2 \, dx \\
&= \int_{\mathbb{R}^d} \left[\frac{1}{n} \sum_{1 \le j_1 \le n} K_n(x - X_{j_1})\right] \left[\frac{1}{n} \sum_{1 \le j_2 \le n} K_n(x - X_{j_2})\right] dx \\
&= \frac{1}{n^2} \sum_{1 \le i,j \le n} K_n^{\Delta}(X_i - X_j),
\end{aligned}$$

each of $\lambda_{i,n}$ and $\omega_{ij,n}$ is given by $1-n^{-1}$. For

$$\hat{\theta}_n^{\texttt{ISD-LO}} = \frac{1}{n}\sum_{1\le i\le n}\int_{\mathbb{R}^d}\hat{f}_{i,n}^{\texttt{LO}}(x)^2dx$$
$$= \frac{1}{n}\sum_{1\le i\le n}\int_{\mathbb{R}^d}\left[\sum_{1\le j_1\le n} w_{ij_1,n}K_n(x-X_{j_1})\right]\left[\sum_{1\le j_2\le n} w_{ij_2,n}K_n(x-X_{j_2})\right]dx$$
$$= \frac{1}{n^2}\sum_{1\le i,j\le n}\left[n\sum_{1\le k\le n} w_{ki,n}w_{kj,n}\right]K_n^{\Delta}(X_i-X_j),$$

the weights are given by

$$\lambda_{i,n} = \sum_{1\le j,k\le n, j\ne i} w_{ki,n}w_{kj,n}, \qquad \omega_{ij,n} = (n-1)\sum_{1\le k\le n} w_{ki,n}w_{kj,n},$$

while the weights for

$$\hat{\theta}_n^{\texttt{ISD-DCF}} = \int_{\mathbb{R}^d}\hat{f}_{1,n}^{\texttt{CF}}(x)\hat{f}_{n,n}^{\texttt{CF}}(x)dx$$
$$= \int_{\mathbb{R}^d}\left[\sum_{1\le j_1\le n} w_{1j_1,n}K_n(x-X_{j_1})\right]\left[\sum_{1\le j_2\le n} w_{nj_2,n}K_n(x-X_{j_2})\right]dx$$
$$= \frac{1}{n^2}\sum_{1\le i,j\le n}[n^2 w_{1i,n}w_{nj,n}]K_n^{\Delta}(X_i-X_j)$$

can be shown to be given by

$$\lambda_{i,n} = \frac{n/2}{\sum_{1\le j\le n}\mathbb{1}(\lceil 2i/n\rceil = \lceil 2j/n\rceil)},$$
$$\omega_{ij,n}^{\texttt{ISD-DCF}} = \frac{n(n-1)/2}{(n-\lfloor n/2\rfloor)\lfloor n/2\rfloor}\mathbb{1}(\lceil 2i/n\rceil \ne \lceil 2j/n\rceil).$$

In all cases, the weights satisfy (A.3) and (A.4), so the estimators satisfy (7) if

$$\frac{1}{n}\mathbb{E}[W_n^{\texttt{ISD}}(X_1,X_2)^2]\to 0 \tag{A.9}$$

and if

$$\mathbb{E}[\{L_n^{\texttt{ISD}}(X)-L_0(X)\}^2]\to 0. \tag{A.10}$$

Proceeding as in the proof of Theorem 1, it can be shown that (A.9) and (A.10) hold under Conditions D, K, and B$^-$.

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

Finally, the biases of $\hat{\theta}_n^{\mathtt{ISD}}$, $\hat{\theta}_n^{\mathtt{ISD\text{-}LO}}$, and $\hat{\theta}_n^{\mathtt{ISD\text{-}DCF}}$ are given by

$$\beta_n^{\mathtt{ISD}} = \frac{K^{\Delta}(0)}{nh_n^d} + \theta_n^{\mathtt{ISD}} - \theta_0 - \frac{\theta_n^{\mathtt{ISD}}}{n},$$

$$\beta_n^{\mathtt{ISD\text{-}LO}} = \eta_n \frac{K^{\Delta}(0)}{nh_n^d} + \theta_n^{\mathtt{ISD}} - \theta_0 - \eta_n \frac{\theta_n^{\mathtt{ISD}}}{n},$$

and

$$\beta_n^{\mathtt{ISD\text{-}DCF}} = \theta_n^{\mathtt{ISD}} - \theta_0,$$

respectively, where

$$\eta_n = \sum_{1 \le i \le n} \frac{1}{\sum_{1 \le j \le n} \mathbb{1}(\lceil iB_n/n \rceil \neq \lceil jB_n/n \rceil)},$$

and where

$$\theta_n^{\mathtt{ISD}} - \theta_0 = \int_{\mathbb{R}^d} K^{\Delta}(t)[f_0^{\Delta}(h_n t) - f_0^{\Delta}(0)]dt = O(h_n^S) = o(n^{-1/2})$$

under Conditions D, K, and B.

As a consequence, $\hat{\theta}_n^{\mathtt{ISD\text{-}DCF}}$ satisfies (6) under Conditions D, K, and B, whereas

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\mathtt{ISD}}] - \theta_0) = \frac{K^{\Delta}(0)}{\sqrt{nh_n^{2d}}} + o(1),$$

so Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\mathtt{ISD}}$ to satisfy (6). Finally, $\eta_n \ge 1$ is bounded, so Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\mathtt{ISD\text{-}LO}}$ to satisfy (6).

Finally, the results for $\hat{\theta}_n^{\mathtt{ISD\text{-}BC}}$ and $\hat{\theta}_n^{\mathtt{ISD\text{-}GJ}}$ follow from those for $\hat{\theta}_n^{\mathtt{ISD}}$. To be specific, $\hat{\theta}_n^{\mathtt{ISD\text{-}BC}}$ differs from $\hat{\theta}_n^{\mathtt{ISD}}$ by an additive constant, so it satisfies (7) under Conditions D, K, and B$^-$. Also, the additive constant is designed to ensure that (6) is satisfied by $\hat{\theta}_n^{\mathtt{ISD\text{-}BC}}$ under Conditions D, K, and B. Similarly, because

$$\frac{1}{1-c^d} - \frac{c^d}{1-c^d} = 1,$$

the estimator $\hat{\theta}_n^{\mathtt{ISD\text{-}GJ}}$ satisfies (7) under Conditions D, K, and B$^-$, while the fact that

$$\frac{1}{1-c^d}\frac{1}{nh_n^d} - \frac{c^d}{1-c^d}\frac{1}{n(ch_n)^d} = 0$$

ensures that (6) is satisfied by $\hat{\theta}_n^{\mathtt{ISD\text{-}GJ}}$ under Conditions D, K, and B.

A.5. Proof of Theorem 4

The proof is similar to that of Theorem 2. The estimators $\hat{\theta}_n^{\mathtt{ISD},*}$, $\hat{\theta}_n^{\mathtt{ISD\text{-}LO},*}$, and $\hat{\theta}_n^{\mathtt{ISD\text{-}DCF},*}$ all have Hoeffding decompositions of the form (A.2), with

$$L_{i,n}^* = \lambda_{i,n}\hat{L}_n^{\mathtt{ISD}}(X_{i,n}^*), \qquad W_{ij,n}^* = \omega_{ij,n}\hat{W}_n^{\mathtt{ISD}}(X_{i,n}^*, X_{j,n}^*),$$

where $\lambda_{i,n}$ and $\omega_{ij,n}$ are the same as those for $\hat{\theta}_n^{\mathtt{ISD}}$, $\hat{\theta}_n^{\mathtt{ISD\text{-}LO}}$, and $\hat{\theta}_n^{\mathtt{ISD\text{-}DCF}}$, while

$$\hat{L}_n^{\mathtt{ISD}}(x) = 2\{\hat{f}_n^{\mathtt{ISD}}(x) - \hat{\theta}_n^{\mathtt{ISD}}\}, \qquad \hat{f}_n^{\mathtt{ISD}}(x) = \frac{1}{n}\sum_{1\le j\le n} K_n^{\Delta}(x - X_j),$$

$$\hat{W}_n^{\mathtt{ISD}}(x_1,x_2) = K_n^{\Delta}(x_1 - x_2) - \hat{f}_n^{\mathtt{ISD}}(x_1) - \hat{f}_n^{\mathtt{ISD}}(x_2) + \hat{\theta}_n^{\mathtt{ISD}}.$$

Because the weights satisfy (A.3) and (A.4), it follows from simple moment calculations that the estimators satisfy

$$\sqrt{n}(\hat{\theta}_n^* - \mathbb{E}_n^*[\hat{\theta}_n^*]) = \frac{1}{\sqrt{n}}\sum_{1\le i\le n}\{L_0(X_{i,n}^*) - \mathbb{E}_n^*[L_0(X_{i,n}^*)]\} + o_{\mathbb{P}}(1) \rightsquigarrow_{\mathbb{P}} \mathcal{N}(0,\sigma_0^2)$$

if

$$\frac{1}{n}\mathbb{E}_n^*[\hat{W}_n^{\mathtt{ISD}}(X_{1,n}^*, X_{2,n}^*)^2] \to_{\mathbb{P}} 0 \tag{A.11}$$

and if (A.10) and (A.12) hold, where

$$\mathbb{E}_n^*[\{\hat{L}_n^{\mathtt{ISD}}(X_{1,n}^*) - L_n^{\mathtt{ISD}}(X_{1,n}^*)\}^2] \to_{\mathbb{P}} 0. \tag{A.12}$$

Suppose Conditions D and K are satisfied. Then, (A.11) holds if $nh_n^d \to \infty$, because then

$$\begin{aligned}
\frac{1}{n}\mathbb{E}_n^*[\hat{W}_n^{\mathtt{ISD}}(X_{1,n}^*, X_{2,n}^*)^2] &\le \frac{1}{n}\mathbb{E}_n^*[K_n^{\Delta}(X_{1,n}^* - X_{2,n}^*)^2] \\
&= \frac{1}{n^3}\sum_{1\le i,j\le n} K_n^{\Delta}(X_i - X_j)^2 \\
&= \frac{1}{n^3}\sum_{1\le i\le n} K_n^{\Delta}(0)^2 + \frac{2}{n^3}\sum_{1\le i,j\le n, i<j} K_n^{\Delta}(X_i - X_j)^2 \\
&= \frac{1}{n}\left(\frac{K^{\Delta}(0)}{nh_n^d}\right)^2 + O_{\mathbb{P}}\left(\frac{1}{n}\mathbb{E}[K_n^{\Delta}(X_1 - X_2)^2]\right) \to_{\mathbb{P}} 0.
\end{aligned}$$

Also, (A.10) holds when $h_n \to 0$. Finally, because

$$\mathbb{E}_n^*[\{\hat{L}_n^{\mathtt{ISD}}(X_{1,n}^*) - L_n^{\mathtt{ISD}}(X_{1,n}^*)\}^2] = \frac{1}{n}\sum_{1\le i\le n}\{\hat{L}_n^{\mathtt{ISD}}(X_i) - L_n^{\mathtt{ISD}}(X_i)\}^2,$$

a sufficient condition for (A.12) to hold is that

$$\mathbb{E}[\{\hat{L}_n^{\mathtt{ISD}}(X_1) - L_n^{\mathtt{ISD}}(X_1)\}^2] \to 0.$$

It follows from a direct calculation that this condition is satisfied when $h_n \to 0$ and $nh_n^d \to \infty$. To summarize, each estimator satisfies (11) under Conditions D, K, and B^-.

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press

The proof will be completed by giving conditions under which the estimators satisfy (14). Suppose Conditions D, K, and B are satisfied. By the proof of Theorem 3,

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\texttt{ISD}}] - \theta_0) = \frac{K^{\Delta}(0)}{\sqrt{nh_n^{2d}}} + o(1),$$

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\texttt{ISD-LO}}] - \theta_0) = \eta_n \frac{K^{\Delta}(0)}{\sqrt{nh_n^{2d}}} + o(1),$$

and

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\texttt{ISD-DCF}}] - \theta_0) = o(1),$$

while it follows from (A.2) and Theorem 3 that

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{ISD},*}] - \hat{\theta}_n^{\texttt{ISD}}) = \frac{K^{\Delta}(0)}{\sqrt{nh_n^{2d}}} - \frac{\hat{\theta}_n^{\texttt{ISD}}}{\sqrt{n}} = \frac{K^{\Delta}(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1),$$

$$\begin{aligned}\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{ISD-LO},*}] - \hat{\theta}_n^{\texttt{ISD-LO}}) &= \eta_n \frac{K^{\Delta}(0)}{\sqrt{nh_n^{2d}}} + \sqrt{n}(\hat{\theta}_n^{\texttt{ISD}} - \hat{\theta}_n^{\texttt{ISD-LO}}) - \eta_n \frac{\hat{\theta}_n^{\texttt{ISD}}}{\sqrt{n}} \\ &= \frac{K^{\Delta}(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1),\end{aligned}$$

and

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{ISD-DCF},*}] - \hat{\theta}_n^{\texttt{ISD-DCF}}) = \sqrt{n}(\hat{\theta}_n^{\texttt{ISD}} - \hat{\theta}_n^{\texttt{ISD-DCF}}) = \frac{K^{\Delta}(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1).$$

As a consequence, $\hat{\theta}_n^{\texttt{ISD},*}$ satisfies (14) under Conditions D, K, and B, whereas Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{ISD-DCF},*}$ to satisfy (14). Finally, if $B_n = n$, then

$$\eta_n = \frac{n}{n-1} = 1 + O(n^{-1}),$$

so $\hat{\theta}_n^{\texttt{ISD-LO},*}$ satisfies (14) under Conditions D, K, and B. On the other hand, Condition B must be strengthened to Condition B$^+$ for the cross-fit version of $\hat{\theta}_n^{\texttt{ISD-LO},*}$ to satisfy (14), because if $B_n = B$, for all n, then

$$\eta_n \to \frac{B}{B-1} \neq 1.$$

Finally, the results for $\hat{\theta}_n^{\texttt{ISD-BC},*}$ and $\hat{\theta}_n^{\texttt{ISD-GJ},*}$ follow from those for $\hat{\theta}_n^{\texttt{ISD},*}$. To be specific, $\hat{\theta}_n^{\texttt{ISD-BC},*}$ satisfies (11) under Conditions D, K, and B$^-$, because $\hat{\theta}_n^{\texttt{ISD},*}$ does. Moreover,

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{ISD-BC},*}] - \hat{\theta}_n^{\texttt{ISD-BC}}) = \sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{ISD},*}] - \hat{\theta}_n^{\texttt{ISD}}),$$

so under Conditions D and K, Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\mathtt{ISD\text{-}BC},*}$ to satisfy (14) (unless $K(0)=0$). Similarly, because

$$\frac{1}{1-c^d}-\frac{c^d}{1-c^d}=1,$$

the estimator $\hat{\theta}_n^{\mathtt{ISD\text{-}GJ},*}$ satisfies (11) under Conditions D, K, and B$^-$, while the fact that

$$\frac{1}{1-c^d}\frac{1}{nh_n^d}-\frac{c^d}{1-c^d}\frac{1}{n(ch_n)^d}=0$$

ensures that (14) is satisfied by $\hat{\theta}_n^{\mathtt{ISD\text{-}GJ},*}$ under Conditions D, K, and B.

A.6. Proof of Theorem 5

It follows from the proofs of Theorems 1 and 3 that the estimators $\hat{\theta}_n^{\mathtt{LR}}, \hat{\theta}_n^{\mathtt{LR\text{-}BC}}, \hat{\theta}_n^{\mathtt{LR\text{-}GJ}}$, and $\hat{\theta}_n^{\mathtt{LR\text{-}LO}}$ satisfy (7) under Conditions D, K, and B$^-$ and have biases of the form

$$\beta_n^{\mathtt{LR}}=2\beta_n^{\mathtt{AD}}-\beta_n^{\mathtt{ISD}}=\frac{2K(0)-K^{\Delta}(0)}{nh_n^d}+o(n^{-1/2}),$$

$$\beta_n^{\mathtt{LR\text{-}BC}}=o(n^{-1/2}),\qquad \beta_n^{\mathtt{LR\text{-}GJ}}=o(n^{-1/2}),$$

and

$$\beta_n^{\mathtt{LR\text{-}LO}}=2\beta_n^{\mathtt{AD\text{-}LO}}-\beta_n^{\mathtt{ISD\text{-}LO}}=-\eta_n\frac{K^{\Delta}(0)}{nh_n^d}+o(n^{-1/2}),$$

respectively, under Conditions D, K, and B.

As a consequence, $\hat{\theta}_n^{\mathtt{LR\text{-}BC}}$ and $\hat{\theta}_n^{\mathtt{LR\text{-}GJ}}$ satisfy (6) under Conditions D, K, and B, whereas Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\mathtt{LR\text{-}LO}}$ to satisfy (6). Likewise, Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\mathtt{LR}}$ to satisfy (6) unless $2K(0)=K^{\Delta}(0)$.

A.7. Proof of Theorem 6

It follows from the proofs of Theorems 2 and 4 that the estimators $\hat{\theta}_n^{\mathtt{LR},*}, \hat{\theta}_n^{\mathtt{LR\text{-}BC},*}$, $\hat{\theta}_n^{\mathtt{LR\text{-}GJ},*}$, and $\hat{\theta}_n^{\mathtt{LR\text{-}LO},*}$ satisfy (11) under Conditions D, K, and B$^-$. The proof will be completed by giving conditions under which the estimators satisfy (14). Suppose Conditions D, K, and B are satisfied. By the proof of Theorem 5,

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\mathtt{LR}}]-\theta_0)=\frac{2K(0)-K^{\Delta}(0)}{\sqrt{nh_n^{2d}}}+o(1),$$

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\mathtt{LR\text{-}BC}}]-\theta_0)=o(1),$$

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\mathtt{LR\text{-}GJ}}]-\theta_0)=o(1),$$

and

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n^{\texttt{LR-LO}}] - \theta_0) = -\eta_n \frac{K^\Delta(0)}{nh_n^d} + o(1),$$

while it follows from the proofs of Theorems 2 and 4 that

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{LR},*}] - \hat{\theta}_n^{\texttt{LR}}) = \frac{2K(0) - K^\Delta(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1),$$

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{LR-BC},*}] - \hat{\theta}_n^{\texttt{LR-BC}}) = \frac{2K(0) - K^\Delta(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1),$$

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{LR-GJ},*}] - \hat{\theta}_n^{\texttt{LR-GJ}}) = o_{\mathbb{P}}(1),$$

and

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{LR-LO},*}] - \hat{\theta}_n^{\texttt{LR-LO}}) = \frac{2K(0) - K^\Delta(0)}{\sqrt{nh_n^{2d}}} + o_{\mathbb{P}}(1).$$

As a consequence, $\hat{\theta}_n^{\texttt{LR},*}$ and $\hat{\theta}_n^{\texttt{LR-GJ},*}$ satisfy (14) under Conditions D, K, and B, whereas Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{LR-BC},*}$ and $\hat{\theta}_n^{\texttt{LR-LO},*}$ to satisfy (14).

https://doi.org/10.1017/S0266466621000530 Published online by Cambridge University Press