Average Density Estimators: Efficiency and Bootstrap Consistency*

Matias D. Cattaneo† Michael Jansson‡

April 19, 2019

Abstract

This paper highlights a tension between semiparametric efficiency and bootstrap consistency in the context of a canonical semiparametric estimation problem. 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, "debiased" estimators that achieve semiparametric efficiency under minimal smoothness conditions do not achieve bootstrap consistency under those same conditions.

Keywords: Semiparametric estimation, efficiency, bootstrap consistency.
JEL: C14.

*For comments and suggestions, we are grateful to seminar participants at the Cowles Foundation conference celebrating Peter Phillips' forty years at Yale and the 2019 LAMES conference. Cattaneo gratefully acknowledges financial support from the National Science Foundation through grant SES-1459931, and Jansson gratefully acknowledges financial support from the National Science Foundation through grant SES-1459967 and the research support of CREATES (funded by the Danish National Research Foundation under grant no. DNRF78).

†Department of Economics and Department of Statistics, University of Michigan.
‡Department of Economics, University of California at Berkeley and CREATES.

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 *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 (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 non-negligible. 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.

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 none of the estimators produced by these approaches also achieve bootstrap consistency under minimal smoothness conditions. In sharp 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 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 estimators are analyzed in Section 5. Finally, Section 6 offers concluding remarks and Section 7 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 studied estimand in this setting is

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

the average density. The average density is of some interest in its own right, but more importantly the problem of estimating θ_0 has attracted considerable interest because it can be viewed as a canonical example of a semiparametric estimation problem.

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 that if f_0 is bounded, then the efficient influence function 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

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

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

Condition D For some $s > d/4$ with $s/2 \notin \mathbb{N}$, 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.

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

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

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 $1-\alpha$ "percentile" bootstrap confidence interval for θ_0 is given by

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

This interval is said to be consistent if

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

and to be efficient if its end points satisfy

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

where $\Phi(\cdot)$ is the standard normal cdf. 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 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 and 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 $\mathsf{CI}_{n,1-\alpha}$ are governed solely by (the density f_0 and) the functional form of $\hat{\theta}_n$, the properties (2) , (3) , and (4) can all be interpreted as a properties of the estimator $\hat{\theta}_n$ and one of the main purposes of this paper is explore the relationship between those properties and the more familiar (efficiency) property (1) .

3 Average Density Estimators: Efficiency

Our discussion of efficiency (or otherwise) of average density estimators $\hat{\theta}_n$ will 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 $\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]$, respectively. If these components satisfy

$$\sqrt{n}(\mathbb{E}[\hat{\theta}_n] - \theta_0) = o(1) \tag{5}$$

and

$$\sqrt{n}(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]) = \frac{1}{\sqrt{n}} \sum_{1\le i\le n} L_0(X_i) + o_{\mathbb{P}}(1), \tag{6}$$

respectively, then (1) holds. Moreover, if (6) holds, then the easy-to-interpret bias condition (5) 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 (6) under very mild conditions.

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

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

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 $\hat{\theta}_n^{\mathtt{AD}}$ and other estimators, we impose the following standard condition on the kernel, in which $\|\cdot\|_p$ denotes the p-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

$$\int_{\mathbb{R}^d} |K(u)| \, (1 + \|u\|_2^P) du < \infty$$

and

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

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

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 $\hat{\theta}_n^{\mathtt{AD}}$ satisfies (6) (under Conditions D and K).[1] As a consequence, under Conditions D, K, and B$^-$, the estimator $\hat{\theta}_n^{\mathtt{AD}}$ is efficient if and only if it satisfies the bias condition (5).

Defining

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

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

$$\mathbb{E}[\hat{\theta}_n^{\mathtt{AD}}] - \theta_0 \approx \frac{K(0)}{nh_n^d} + \int_{\mathbb{R}^d} K(t)[f_0^{\Delta}(h_n t) - f_0^{\Delta}(0)] dt, \tag{7}$$

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 term is a smoothing bias term whose magnitude depends on the order of K and the smoothness of f_0. To be specific, under

[1] Conversely, Condition B$^-$ is minimal in the sense that the methods of Cattaneo, Crump, and Jansson (2014b) can be used to show that (6) can fail if Condition B$^-$ is violated.

Conditions D and K, Giné and Nickl (2008b) can be used to show that if $h \to 0$, then

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

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

Condition $\mathbf{B}^+$ For $S = \min(P/2, s)$, $nh_n^{4S} \to 0$ 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 $\hat{\theta}_n^{\mathtt{AD}}$ 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 $\hat{\theta}_n^{\mathtt{AD}}$ is easily avoidable. One option is to employ a kernel satisfying $K(0) = 0$. Recognizing that no standard kernels satisfy that condition, a more natural option is to use the "bias-corrected" version of $\hat{\theta}_n^{\mathtt{AD}}$ given by

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

By construction, the bias of this estimator satisfies

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

so under Conditions D and K the bias condition (5) is satisfied by $\hat{\theta}_n^{\mathtt{AD-BC}}$ provided $nh_n^{4S} \to 0$, implying in turn that $\hat{\theta}_n^{\mathtt{AD-BC}}$ 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 For $S = \min(P/2, s)$, $nh_n^{4S} \to 0$ and $nh_n^d \to \infty$.

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

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

where $\hat{f}_{i,n}^{\mathtt{LO}}$ is a kernel density estimator constructed using observations belonging to a set that does include X_i. Relative to $\hat{\theta}_n^{\mathtt{AD-BC}}$, an attractive feature of $\hat{\theta}_n^{\mathtt{AD-LO}}$ is that it can be constructed without knowledge of the functional form of the leave-in bias. For concreteness, we shall develop results for $\hat{\theta}_n^{\mathtt{AD-LO}}$ only in the (leading) special case where the sample the sample $X_1, \ldots, X_n$ is partitioned into $B_n \in \{2, \ldots, n\}$ disjoint blocks of (approximately) equal size and $\hat{f}^{\texttt{LO}}_{i,n}$ is constructed using observations from all blocks except the one to which the ith observation belongs. To be specific, we assume that $\hat{f}^{\texttt{LO}}_{i,n}$ is of the form

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

When $B_n = n$, $\hat{f}^{\texttt{LO}}_{i,n}$ is the ith "leave-one-out" estimator of f_0 and the estimator $\hat{\theta}^{\texttt{AD-LO}}_n$ 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 $\hat{\theta}^{\texttt{AD-LO}}_n$ is a "cross-fit" estimator (using an B_n-fold non-random 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 $\hat{\theta}^{\texttt{AD-LO}}_n$ is similar to $\hat{\theta}^{\texttt{AD-BC}}_n$ insofar as it satisfies (6) and has

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

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

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

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

4 Average Density Estimators: Bootstrap Consistency

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

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

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

and

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

respectively. 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:

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \rightsquigarrow_{\mathbb{P}} \mathcal{N}(0, \sigma_0^2), \tag{8}$$

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 (8) 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 $\hat{\theta}_n^* - \mathbb{E}_n^*[\hat{\theta}_n^*]$, respectively, where $\mathbb{E}_n^*[\cdot] = \mathbb{E}[\cdot|X_1, \ldots, X_n]$. The resulting sufficient condition for (8) is given by the pair

$$\sqrt{n}(\mathbb{E}_n^*[\hat{\theta}_n^*] - \hat{\theta}_n) = o_{\mathbb{P}}(1) \tag{9}$$

and

$$\sqrt{n}(\hat{\theta}_n^* - \mathbb{E}_n^*[\hat{\theta}_n^*]) \rightsquigarrow_{\mathbb{P}} \mathcal{N}(0, \sigma_0^2), \tag{10}$$

where (9) is the natural bootstrap analog of (5), (10) is a bootstrap version of the main distributional implication of (6), and where (9) is necessary and sufficient for (8) when (10) holds.

In perfect analogy with (6), it turns out that (10) holds under very mild bandwidth conditions. Indeed, under Conditions D and K, the estimators $\hat{\theta}_n^{\texttt{AD},*}$, $\hat{\theta}_n^{\texttt{AD-BC},*}$, and $\hat{\theta}_n^{\texttt{AD-LO},*}$ all satisfy (10) 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 $\hat{\theta}_n^{\texttt{AD},*}$ satisfies

$$\mathbb{E}_n^*[\hat{\theta}_n^{\texttt{AD},*}] - \hat{\theta}_n^{\texttt{AD}} = \frac{K(0)}{nh_n^d} - \frac{1}{n}\hat{\theta}_n^{\texttt{AD}} = \frac{K(0)}{nh_n^d} + O_{\mathbb{P}}(n^{-1}). \tag{11}$$

Therefore, the bias condition (9) is satisfied by $\hat{\theta}_n^{\texttt{AD},*}$ provided $nh_n^{2d} \to \infty$. In other words, $\hat{\theta}_n^{\texttt{AD},*}$ satisfies (2) (and therefore also (3) and (4)) under Conditions D, K, and B^+.

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

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

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.

In sharp contrast, it turns out that $\hat{\theta}_n^{\texttt{AD},*}$ satisfies (2), (3), and (4) under conditions that are weaker than the conditions under which $\hat{\theta}_n^{\texttt{AD}}$ is efficient. In generic notation, suppose the estimators

[2] Conversely, Condition B^- is minimal in the sense that the methods of Cattaneo, Crump, and Jansson (2014a) can be used to show that (10) can fail if Condition B^- is violated.

$\hat{\theta}_n$ and $\hat{\theta}_n^*$ satisfy (6) and (10), 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 (5) and (9):

$$\sqrt{n}(\mathbb{E}^*[\hat{\theta}_n^*] - \hat{\theta}_n) = \sqrt{n}(\mathbb{E}[\hat{\theta}_n] - \theta_0) + o_{\mathbb{P}}(1). \tag{13}$$

Now, as discussed above, the estimators $\hat{\theta}_n^{\texttt{AD}}$ and $\hat{\theta}_n^{\texttt{AD},*}$ satisfy (6) and (10), respectively, under Conditions D, K, and B. Under the same conditions, it follows from (7) and (11) that (13) is satisfied. As consequence, we obtain the first part of the following result.

Theorem 2 *Suppose Conditions D, K, and B are satisfied. Then $\hat{\theta}_n^{\texttt{AD},*}$ satisfies (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 may be a fundamental 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 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 construction of an estimator is simply a means to the end of conducting bootstrap-based inference.

In light of Theorem 2 it is of interest to construct bootstrap-based approximations to the distributions of $\hat{\theta}_n^{\texttt{AD-BC}}$ and $\hat{\theta}_n^{\texttt{AD-LO}}$ 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

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \le t] - \mathbb{P}_n^*[\sqrt{n}(\tilde{\theta}_n^* - \hat{\theta}_n) \le t] \right| = o_{\mathbb{P}}(1). \tag{14}$$

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

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

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 (12) 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 (14) 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 (14) 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 (14) if and only if $\hat{\theta}_n^{\texttt{AD},*}$ satisfies (2). As a consequence, the fact $\tilde{\theta}_n^{\texttt{AD-BC},*}$ satisfies (14) 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

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

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

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

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

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

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

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

In fact, it can be shown that (14) 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

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

where

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

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,

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

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

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

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

5 Alternative Estimators

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

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

while 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, Escanciano, Ichimura, Newey, and Robins (2018)).

5.1 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 $\hat{\theta}_n^{\mathtt{AD}}$, 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 "leave in" bias term (in the terminology of Cattaneo, Crump, and Jansson (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

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

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

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

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

$$\mathbb{E}[\hat{\theta}_n^{\texttt{ISD-LO}}] - \theta_0 = \frac{1}{1 - B_n^{-1}} \frac{\int_{\mathbb{R}^d} K(u)^2 du}{n h_n^d} + o(n^{-1/2}), \tag{16}$$

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

On the other hand, 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 $\hat{\theta}_n^{\texttt{ISD}}, \hat{\theta}_n^{\texttt{ISD-BC}}, \hat{\theta}_n^{\texttt{ISD-LO}}$, and $\hat{\theta}_n^{\texttt{ISD-DCF}}$ all satisfy (6). 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* $\hat{\theta}_n^{\texttt{ISD-BC}}$ *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 $\hat{\theta}_n$ is one of the four abovementioned estimators, then its bootstrap analog satisfies (10) and has a bias of the form

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

so (2) is satisfied if (and only if)

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

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 (16) 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).

[4] More generally, the bias expansion is of the form

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

where $\eta_n \geq 1$ is bounded.

Theorem 4 *Suppose Conditions D, K, and B are satisfied. Then* $\hat{\theta}_n^{\texttt{ISD},*}$ *satisfies* (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 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).

For completeness, we conclude this subsection by briefly discussing integrated squared density versions of (14) and (15). 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 (14) 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 (14) 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 (15). 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 (15).

5.2 Locally Robust Estimators

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

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

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

$$\hat{\theta}_n^{\texttt{LR-BC}} = 2\hat{\theta}_n^{\texttt{AD-BC}} - \hat{\theta}_n^{\texttt{ISD-BC}}$$

and

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

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, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (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 $\hat{\theta}_n^{\texttt{LR-BC}}$ satisfies (1). If Condition B is strengthened to Condition B^+, then $\hat{\theta}_n^{\texttt{LR}}$ and $\hat{\theta}_n^{\texttt{LR-LO}}$ satisfy (1).*

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

6 Concluding Remarks

Our investigation of average density and integrated squared density estimators lead us to draw two main conclusions. First, in spite of their inability to achieve efficiency under minimal conditions, simple plug-in estimators are attractive from the perspective of inference because the percentile intervals associated with these estimators are asymptotically valid (indeed, efficient) under minimal conditions. Second, although the estimand is sufficiently simple to permit estimators of cross-fit type to be efficient under minimal conditions, care must be exercised when using such estimators for inference. In particular, percentile intervals based on the nonparametric bootstrap are not asymptotically valid (let alone efficient) under minimal conditions. On the other hand, a carefully constructed bootstrap procedure (namely, the cross-fit bootstrap) turns out to give rise to inference procedures that are asymptotically valid (indeed, efficient) under minimal conditions.

The positive results about percentile intervals associated with simple plug-in estimators are driven by the ability of the bootstrap to "endogenously" perform a bias-correction when approximating the distribution of the estimator(s). Cattaneo and Jansson (2018) found that the same mechanism enables the nonparametric bootstrap to be consistent under weak conditions in a much more general class of two-step semiparametric estimators whose first-step estimator is kernel-based (in a sense made precise in that paper). It would be of interest to investigate whether similar results can be obtained also for two-step semiparametric estimators whose first-step is not kernel-based. Some results in that direction have been obtained for first step series estimators by Cattaneo, Jansson, and Ma (2018), but more work is needed to fully understand the extent to which the findings of this paper are representative. Likewise, it would be of interest to explore whether our (positive and negative) results about the cross-fit-type estimators and bootstrap procedures generalize to more complicated settings.

7 Proofs

7.1 Hoeffding Decompositions

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

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

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

$$\hat{\theta}_n - \theta_0 = \beta_n + \frac{1}{n} \sum_{1 \leq i \leq n} L_{i,n} + \frac{2}{n(n-1)} \sum_{1 \leq i,j \leq n, i<j} W_{ij,n}, \qquad (17)$$

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

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

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

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

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 $i \neq j$,

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

Moreover, if the $V_{ij,n}$ satisfy $V_{ii,n} = \delta_n, V_{ij,n} = V_{ji,n}$, 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 $\mathbb{E}[V_{ij,n}|X_i] = f_n(X_i)$, 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 (17) will be employed in the proofs of Theorems 2, 4, and 6. To state it, suppose

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

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

$$\hat{\theta}_n^* - \hat{\theta}_n = \beta_n^* + \frac{1}{n} \sum_{1 \le i \le n} L_{i,n}^* + \frac{2}{n(n-1)} \sum_{1 \le i,j \le n, i<j} W_{ij,n}^*, \tag{18}$$

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

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

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

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

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 $i \neq j$,

$$\mathbb{E}_n^*[L_{i,n}^*] = \mathbb{E}_n^*[W_{ij,n}^*|X_{i,n}^*] = \mathbb{E}_n^*[W_{ij,n}^*|X_{j,n}^*] = 0.$$

Moreover, if the $V^*_{ij,n}$ satisfy $V^*_{ii,n} = \delta^*_n$, $\mathbb{E}^*_n[V^*_{ij,n}] = \theta^*_n$, and $V^*_{ij,n} = V^*_{ji,n}$, 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 $\mathbb{E}^*_n[V^*_{ij,n}|X^*_{i,n}] = f^*_n(X^*_{i,n})$, then

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

7.2 Proof of Theorem 1

The estimators $\hat{\theta}^{\mathtt{AD}}_n$, $\hat{\theta}^{\mathtt{AD-BC}}_n$, and $\hat{\theta}^{\mathtt{AD-LO}}_n$ all have Hoeffding decompositions of the form (17), with

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

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

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

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

where

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

$$\theta^{\mathtt{AD}}_n = \mathbb{E}[f^{\mathtt{AD}}_n(X)] = \int_{\mathbb{R}^d} f^{\mathtt{AD}}_n(x) f_0(x) dx.$$

To be specific, in the case of

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

each $\lambda_{i,n}$ and $\omega_{i,n}$ is given by $1 - n^{-1}$. Because, $\hat{\theta}^{\mathtt{AD-BC}}_n$ differs from $\hat{\theta}^{\mathtt{AD}}_n$ by an additive constant, its $\lambda_{i,n}$ and $\omega_{ij,n}$ are of the same form. On the other hand, for

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

we have

$$\lambda_{i,n} = \sum_{1\le j\le n} \bar{w}_{ij,n}, \qquad \omega_{ij,n} = (n-1)\bar{w}_{ij,n},$$

where $\bar{w}_{ij,n} = (w_{ij,n} + w_{ji,n})/2$.

In all cases, the weights satisfy

$$\max_{1\leq i\leq n}(\lambda_{i,n}-1)^2 = o(1) \tag{19}$$

and

$$\max_{1\leq i<j\leq n}\omega_{ij,n}^2 = O(1). \tag{20}$$

It therefore follows from simple moment calculations that the estimators satisfy (6) if

$$\frac{1}{n}\mathbb{E}[W_n^{\mathtt{AD}}(X_1,X_2)^2] \to 0 \tag{21}$$

and if

$$\mathbb{E}[\{L_n^{\mathtt{AD}}(X)-L_0(X)\}^2] \to 0. \tag{22}$$

Suppose Conditions D and K are satisfied. Then (21) holds if $nh_n^d \to \infty$, because then

$$\begin{aligned}\frac{1}{n}\mathbb{E}[W_n^{\mathtt{AD}}(X_1,X_2)^2] &\leq \frac{1}{nh_n^d}\left\{h_n^d\mathbb{E}[K_n(X_1-X_2)^2]\right\}\\ &= \frac{1}{nh_n^d}\left\{h_n^d\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K_n(u-v)^2 f_0(u)f_0(v)dudv\right\}\\ &= \frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K(t)^2 f_0(v+h_nt)f_0(v)dtdv\\ &\leq \frac{1}{nh_n^d}\left\{\sup_{u\in\mathbb{R}^d}|K(u)|\right\}\left\{\sup_{x\in\mathbb{R}^d}f_0(x)\right\}\int_{\mathbb{R}^d}|K(u)|du \to 0.\end{aligned}$$

Also, because

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

a sufficient condition for (22) to hold is that

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

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 (6) under Conditions D, K, and B$^-$.

The proof will be completed by giving conditions under which the estimators satisfy (5). As before, suppose Conditions D and K are satisfied. In the notation introduced above, the biases of $\hat{\theta}_n^{\mathtt{AD}}$, $\hat{\theta}_n^{\mathtt{AD-BC}}$, and $\hat{\theta}_n^{\mathtt{AD-LO}}$ are given by

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

$$\beta_n^{\mathtt{AD-BC}} = \theta_n^{\mathtt{AD}} - \theta_0 - \frac{\theta_n^{\mathtt{AD}}}{n},$$

and

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

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

$$\begin{aligned}\theta_n^{\texttt{AD}} &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K_n(u-v)f_0(v)f_0(u)dudv \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(t)f_0(u-h_nt)f_0(u)dudt \\ &= \int_{\mathbb{R}^d} K(t)f_0^{\Delta}(h_nt)dt,\end{aligned}$$

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 Hölder space $\mathbf{C}^{2s}(\mathbb{R}^d)$. As a consequence, it follows from standard arguments that if Condition B is satisfied, then

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

In particular, $\hat{\theta}_n^{\texttt{AD-LO}}$ satisfies (5) under Conditions D, K, and B. Under the same conditions, $\theta_n^{\texttt{AD}}$ is bounded, so $\hat{\theta}_n^{\texttt{AD-BC}}$ satisfies (5) , whereas

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

so Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{AD}}$ to satisfy (5) (unless $K(0) = 0$).

7.3 Proof of Theorem 2

The estimators $\hat{\theta}_n^{\texttt{AD},*}$, $\hat{\theta}_n^{\texttt{AD-BC},*}$, and $\hat{\theta}_n^{\texttt{AD-LO},*}$ all have Hoeffding decompositions of the form (18) , with

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

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

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

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

Because the weights satisfy (19) and (20) , it follows from simple moment calculations that the estimators satisfy

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

if

$$\frac{1}{n}\mathbb{E}_n^*[\hat{W}_n^{\mathtt{AD}}(X_{1,n}^*, X_{2,n}^*)^2] \to_{\mathbb{P}} 0 \tag{23}$$

and if (22) and (24) hold, where

$$\mathbb{E}_n^*[\{\hat{L}_n^{\mathtt{AD}}(X_{1,n}^*) - L_n^{\mathtt{AD}}(X_{1,n}^*)\}^2] \to_{\mathbb{P}} 0. \tag{24}$$

Suppose Conditions D and K are satisfied. Then (23) holds if $nh_n^d \to \infty$, because then

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

where the convergence result follow from the proof of Theorem 2. In that same proof it was shown that (22) holds when $h_n \to 0$. Finally, because

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

a sufficient condition for (24) to hold is that

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

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 (10) under Conditions D, K, and B$^-$.

The proof will be completed by giving conditions under which the estimators satisfy (13). Suppose Conditions D, K, and B are satisfied. By the proof of Theorem 1,

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

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

and

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

while it follows from (18) and Theorem 1 that

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

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

and

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

As a consequence, $\hat{\theta}_n^{\texttt{AD},*}$ satisfies (13) under Conditions D, K, and B, whereas Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{AD-BC},*}$ and $\hat{\theta}_n^{\texttt{AD-LO},*}$ to satisfy (13) (unless $K(0) = 0$).

7.4 Proof of Theorem 3

The proof is similar to that of Theorem 1. The estimators $\hat{\theta}_n^{\texttt{ISD}}, \hat{\theta}_n^{\texttt{ISD-BC}}, \hat{\theta}_n^{\texttt{ISD-LO}}$, and $\hat{\theta}_n^{\texttt{ISD-CF}}$ all have Hoeffding decompositions of the form (17) , 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 (non-random) 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 $\lambda_{i,n}$ and $\omega_{i,n}$ is given by $1 - n^{-1}$. Because, $\hat{\theta}_n^{\texttt{ISD-BC}}$ differs from $\hat{\theta}_n^{\texttt{ISD}}$ by an additive constant, its $\lambda_{i,n}$ and $\omega_{ij,n}$ are of the same form. On the other hand, for

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

we have

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

while the weights for

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

can be shown to be given by

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

In all cases, the weights satisfy (19) and (20), so the estimators satisfy (6) if

$$\frac{1}{n}\mathbb{E}[W_n^{\texttt{ISD}}(X_1,X_2)^2] \to 0 \tag{25}$$

and if

$$\mathbb{E}[\{L_n^{\texttt{ISD}}(X) - L_0(X)\}^2] \to 0. \tag{26}$$

Proceeding as in the proof of Theorem 1 it can be shown that (25) and (26) hold under Conditions D, K, and B$^-$.

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

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

$$\beta_n^{\texttt{ISD-BC}} = \theta_n^{\texttt{ISD}} - \theta_0 - \frac{\theta_n^{\texttt{ISD}}}{n},$$

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

and

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

respectively, where

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

and where

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

under Conditions D, K, and B.

As a consequence $\hat{\theta}_n^{\texttt{ISD-BC}}$ and $\hat{\theta}_n^{\texttt{ISD-DCF}}$ satisfy (5) under Conditions D, K, and B, whereas

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

so Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{ISD}}$ to satisfy (5). Finally, $\eta_n \geq 1$ is bounded, so Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{ISD-LO}}$ to satisfy (5).

7.5 Proof of Theorem 4

The proof is similar to that of Theorem 2. The estimators $\hat{\theta}_n^{\texttt{ISD},*}, \hat{\theta}_n^{\texttt{ISD-BC},*}, \hat{\theta}_n^{\texttt{ISD-LO},*}$, and $\hat{\theta}_n^{\texttt{ISD-DCF},*}$ all have Hoeffding decompositions of the form (18), with

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

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

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

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

Because the weights satisfy (19) and (20), it follows from simple moment calculations that the estimators satisfy

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

if

$$\frac{1}{n}\mathbb{E}_n^*[\hat{W}_n^{\texttt{ISD}}(X_{1,n}^*, X_{2,n}^*)^2] \to_{\mathbb{P}} 0 \tag{27}$$

and if (26) and (28) hold, where

$$\mathbb{E}_n^*[\{\hat{L}_n^{\mathtt{ISD}}(X_{1,n}^*) - L_n^{\mathtt{ISD}}(X_{1,n}^*)\}^2] \to_{\mathbb{P}} 0. \tag{28}$$

Suppose Conditions D and K are satisfied. Then (27) holds if $nh_n^d \to \infty$, because then

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

Also, (26) holds when $h_n \to 0$. Finally, because

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

a sufficient condition for (28) 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 this condition is satisfied when $h_n \to 0$ and $nh_n^d \to \infty$. To summarize, each estimator satisfies (10) under Conditions D, K, and B$^-$.

The proof will be completed by giving conditions under which the estimators satisfy (13). Suppose Conditions D, K, and B are satisfied. By the proof of Theorem 3,

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

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

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

and

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

while it follows from (18) and Theorem 3 that

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

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

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

and

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

As a consequence, $\hat{\theta}_n^{\texttt{ISD},*}$ satisfies (13) under Conditions D, K, and B, whereas Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{ISD-BC},*}$ and $\hat{\theta}_n^{\texttt{ISD-DCF},*}$ to satisfy (13). Finally, if $B_n = n$, then

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

so $\hat{\theta}_n^{\texttt{ISD-LO},*}$ satisfies (13) under Conditions D, K, and B. Under 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 (13) because if $B_n = B$ for all n, then

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

7.6 Proof of Theorem 5

It follows from the proofs of Theorems 1 and 3 that the estimators $\hat{\theta}_n^{\texttt{LR}}, \hat{\theta}_n^{\texttt{LR-BC}}$, and $\hat{\theta}_n^{\texttt{LR-LO}}$ satisfy (6) under Conditions D, K, and B$^-$ and have biases of the form

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

$$\beta_n^{\texttt{LR-BC}} = 2\beta_n^{\texttt{AD-BC}} - \beta_n^{\texttt{ISD-BC}} = o(n^{-1/2}),$$

and

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

respectively, under Conditions D, K, and B.

As a consequence $\hat{\theta}_n^{\texttt{LR-BC}}$ satisfies (5) under Conditions D, K, and B, whereas Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{LR-LO}}$ to satisfy (5). Likewise, Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\texttt{LR}}$ to satisfy (5) unless $2K(0) - K^{\Delta}(0) = 0$.

7.7 Proof of Theorem 6

It follows from the proofs of Theorems 2 and 4 that the estimators $\hat{\theta}_n^{\texttt{LR},*}, \hat{\theta}_n^{\texttt{LR-BC},*}$, and $\hat{\theta}_n^{\texttt{LR-LO},*}$ satisfy (10) under Conditions D, K, and B$^-$.

The proof will be completed by giving conditions under which the estimators satisfy (13).

Suppose Conditions D, K, and B are satisfied. By the proof of Theorem 5,

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

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

and

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

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

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

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

and

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

As a consequence, $\hat{\theta}_n^{\mathtt{LR},*}$ satisfies (13) under Conditions D, K, and B, whereas Condition B must be strengthened to Condition B$^+$ for $\hat{\theta}_n^{\mathtt{LR-BC},*}$ and $\hat{\theta}_n^{\mathtt{LR-LO},*}$ to satisfy (13).

References

Bickel, P. J., and 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, and M. Jansson (2013): "Generalized Jackknife Estimators of Weighted Average Derivatives (With Discussion and Rejoinder)," *Journal of the American Statistical Association*, 108, 1243–1268.

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

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

Cattaneo, M. D., and M. Jansson (2018): "Kernel-Based Semiparametric Estimators: Small Bandwidth Asymptotics and Bootstrap Consistency," *Econometrica*, 86, 955–995.

Cattaneo, M. D., M. Jansson, and X. Ma (2018): "Two-Step Estimation and Inference with Possibly Many Included Covariates," Review of Economic Studies, forthcoming.

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

Cheng, G., and 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, and J. M. Robins (2018): "Double/Debiased Machine Learning for Treatment and Structural Parameters," *Econometrics Journal*, 21, C1–C68.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2018): "Locally Robust Semiparametric Estimation," working paper.

Giné, E., and R. Nickl (2008a): "A Simple Adaptive Estimator of the Integrated Square of a Density," *Bernoulli*, 14, 47–61.

——— (2008b): "Uniform Central Limit Theorems for Kernel Density Estimators," *Probability Theory and Related Fields*, 141, 333–387.

Hall, P., and J. S. Marron (1987): "Estimation of Integrated Squared Density Derivatives," *Statistics and Probability Letters*, 6, 109–115.

Newey, W. K., and J. M. Robins (2018): "Cross-Fitting and Fast Remainder Rates for Semiparametric Estimation," working paper.

Phillips, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

Phillips, P. C. B., and S. N. Durlauf (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

Phillips, P. C. B., and 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., and P. Perron (1988): "Testing For a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

Ritov, Y., and P. J. Bickel (1990): "Achieving Information Bounds in Non and Semiparametric Models," *Annals of Statistics*, 18, 925–938.