Econometric Theory, **30**, 2014, 1135–1164.
doi:10.1017/S0266466614000127

BOOTSTRAPPING DENSITY-WEIGHTED AVERAGE DERIVATIVES

Matias D. Cattaneo
Department of Economics, University of Michigan

Richard K. Crump
Capital Markets Function, Federal Reserve Bank of New York

Michael Jansson
Department of Economics, UC Berkeley and CREATES

We investigate the properties of several bootstrap-based inference procedures for semiparametric density-weighted average derivatives. The key innovation in this paper is to employ an alternative asymptotic framework to assess the properties of these inference procedures. This theoretical approach is conceptually distinct from the traditional approach (based on asymptotic linearity of the estimator and Edgeworth expansions), and leads to different theoretical prescriptions for bootstrap-based semiparametric inference. First, we show that the conventional bootstrap-based approximations to the distribution of the estimator and its classical studentized version are both invalid in general. This result shows a fundamental lack of "robustness" of the associated, classical bootstrap-based inference procedures with respect to the bandwidth choice. Second, we present a new bootstrap-based inference procedure for density-weighted average derivatives that is more "robust" to perturbations of the bandwidth choice, and hence exhibits demonstrable superior theoretical statistical properties over the traditional bootstrap-based inference procedures. Finally, we also examine the validity and invalidity of related bootstrap-based inference procedures and discuss additional results that may be of independent interest. Some simulation evidence is also presented.

1. INTRODUCTION

The bootstrap has gained great popularity in modern econometrics and statistics.[1] In semiparametric problems, where estimators of a finite-dimensional parameter of interest involve a nonparametric estimator of an unknown function, the

The authors thank Joel Horowitz, Guido Imbens, Lutz Kilian, Demian Pouzo, Rocio Titiunik, seminar participants at Columbia/NYU, Duke, Harvard, LSE, Michigan, Northwestern, Rochester, and USC, and conference participants at the 2010 Econometric Society World Congress for comments. We also thank the co-editor, Yoon-Jae Whang, and two anonymous referees for their suggestions. The first author gratefully acknowledges financial support from the National Science Foundation (SES 0921505 and SES 1122994). The third author gratefully acknowledges financial support from the National Science Foundation (SES 0920953 and SES 1124174) and the research support of CREATES (funded by the Danish National Research Foundation).

bootstrap is attractive, because of its ability to approximate the distribution of the semiparametric estimator in cases where variance estimation is difficult (e.g., Chen, Linton, and van Keilegom, 2003; Cheng and Huang, 2010). Even when variance estimation is relatively straightforward, the bootstrap is potentially useful in semiparametrics, because it may provide more accurate approximations to the distributions of (asymptotically) pivotal quantities such as studentized estimators, whenever it achieves asymptotic refinements similar to those well-established in parametric problems (e.g., Hall, 1992).

The kernel-based density-weighted average derivative estimator of Powell, Stock, and Stoker (1989) is one of the few semiparametric estimators for which the bootstrap has been shown to offer asymptotic refinements. Nishiyama and Robinson (2005) recently showed that a suitably implemented version of the nonparametric bootstrap provides a distributional approximation for the classical studentized test statistic that is superior to the standard Gaussian approximation. In this paper we revisit this problem and obtain new results that can be viewed as a cautionary tale regarding "the potential for bootstrap-based inference to (...) provide improvements in moderate-sized samples" (Nishiyama and Robinson, 2005, p. 927). We present simulation evidence that appears hard to reconcile with the theoretical results establishing asymptotic refinements of the bootstrap in this semiparametric context and develop an alternative theory-based explanation of this evidence. In addition, we use our theoretical framework to derive results for alternative bootstrap-based inference procedures and to show, among other things, that there exists a valid bootstrap-based inference procedure that dominates the one proposed by Nishiyama and Robinson (2005), a theory-based prediction also borne out in our simulations.

The traditional approach to evaluating the accuracy of bootstrap-based inference procedures (in parametric and semiparametric problems) relies on asymptotic linearity of estimators and employs Edgeworth expansions to elucidate the role of "higher-order" terms in the distributional approximation of the associated test statistics. For the density-weighted average derivative estimator, Nishiyama and Robinson (2005) used this traditional approach to demonstrate the ability of a bootstrap-based inference procedure to deliver asymptotic refinements. In contrast, we propose in this paper to employ an alternative (first-order) distributional approach to examine the properties of bootstrap-based inference procedures, which retains some terms that are asymptotically negligible when the estimator is asymptotically linear but can be first-order otherwise. This alternative approach accommodates, but does not require, certain departures from asymptotic linearity, namely those that occur when the bandwidth of the nonparametric estimator vanishes too rapidly for asymptotic linearity to hold. Thus, we refer to this approach as a "small bandwidth" approach (Cattaneo, Crump and Jansson 2010; 2014).

Although similar in spirit to the Edgeworth expansion approach to improve asymptotic approximations, our small bandwidth approach is conceptually distinct and leads to different theoretical prescriptions for bootstrap-based semiparametric inference. In particular, Theorem 1 finds that the conventional bootstrap-based approximations to the distribution of the kernel-based semiparametric estimator and the associated studentized version of this estimator employing the traditional (jackknife) variance estimator are both invalid in general. On the other hand, Theorem 2 establishes consistency of the bootstrap approximation to the distribution of the semiparametric estimator when studentized by a different, bias-corrected variance estimator. This alternative variance estimator is one for which the resulting studentized statistic is asymptotically standard normal even when asymptotic linearity fails. However, and perhaps surprisingly, Theorem 3 shows that pivotality of the studentized estimator is not sufficient for bootstrap validity: a variance estimator is exhibited which renders the associated studentized statistic asymptotically standard normal even when asymptotic linearity fails, but nonetheless the standard bootstrap provides a valid distributional approximation for this asymptotically pivotal statistic only when asymptotic linearity holds.

These results have some interesting theoretical implications. First, our findings shed new light on the properties of the bootstrap and some of its variants in the context of semiparametric inference, documenting and highlighting, in particular, a fragility of traditional bootstrap-based distributional approximations for kernel-based semiparametric statistics with respect to perturbations of the bandwidth choice (see Section 4.4 for further discussion on this point). Second, our results also include a new bootstrap-based inference procedure for density-weighted average derivatives which is more "robust" to perturbations of the bandwidth choice, and hence exhibiting theoretically demonstrable superior statistical properties over the traditional bootstrap-based inference procedures.

The remainder of the paper is organized as follows. Section 2 introduces the model, summarizes some theoretical results available in the literature, and provides a motivation for our work using a small-scale simulation study. Section 3 reviews our alternative approach based on the small bandwidth framework and develops the main theoretical tools needed to study the bootstrap. Section 4 includes the main results of the paper, while Section 5 concludes and discusses other contexts where our results could be applied. The Appendix contains brief mathematical proofs, but the supplemental appendix includes a detailed development of our results.

2. SETUP AND MOTIVATION

We assume throughout that $z_i = (y_i, x_i')'$, $i = 1, \dots, n$, is a random sample of $z = (y, x')'$, where $y \in \mathbb{R}$ is a dependent variable and $x \in \mathbb{R}^d$ is a continuous explanatory variable with density $f(\cdot)$. The density-weighted average derivative of the regression function $g(x) = \mathbb{E}[y|x]$ is $\theta = \mathbb{E}[f(x)\partial g(x)/\partial x]$; see, e.g., Stoker (1986). (Detailed regularity conditions are given in the following section, but omitted here to ease the discussion.) Models where this estimand is of interest include single-index limited dependent variable models, generalized partially linear models, and other related semilinear single-index generalized additive and nonadditive models. For example, suppose $g(\cdot)$ is of the form $g(x) = G(x_1'\beta, x_2)$ with $G(\cdot)$ unknown and x partitioned as $x = (x_1', x_2')'$. Then, partitioning θ conformably with x as $\theta = (\theta_1', \theta_2')'$, the index parameter β is proportional to θ_1 with proportionality factor $\mathbb{E}[f(x)\dot{G}_1(x_1'\beta, x_2)]$ where $\dot{G}_1(u, x_2) = \partial G(u, x_2)/\partial u$.

Powell, Stock, and Stoker (1989, henceforth PSS) noted that $\theta = -2\mathbb{E}[y\partial f(x)/\partial x]$, and hence proposed the kernel-based estimator

$$\hat{\theta}_n = -2\frac{1}{n}\sum_{i=1}^{n} y_i \frac{\partial}{\partial x}\hat{f}_{n,i}(x_i), \quad \hat{f}_{n,i}(x) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} \frac{1}{h_n^d} K\left(\frac{x_j - x}{h_n}\right),$$

where $K : \mathbb{R}^d \to \mathbb{R}$ is a kernel function and h_n is a vanishing (positive) bandwidth sequence. Having subsequently been studied by Härdle and Tsybakov (1993), Robinson (1995), Powell and Stoker (1996), Nishiyama and Robinson (2000, 2001, 2005), and many others, this estimator is one of the most widely investigated estimators in the semiparametrics literature.

Under conditions similar to those discussed below, PSS showed that $\hat{\theta}_n$ is asymptotically linear with influence function $L(z) = 2(\partial[f(x)g(x)]/\partial x - y\partial f(x)/\partial x - \theta)$; that is,

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L(z_i) + o_p(1) \rightsquigarrow \mathcal{N}(0, \Sigma), \qquad \Sigma = \mathbb{E}\left[L(z)L(z)'\right], \tag{1}$$

where $\rightsquigarrow$ denotes weak convergence. (Throughout the paper limits are taken as $n \to \infty$ unless otherwise noted.) PSS also exhibited a consistent estimator $\hat{\Sigma}_n$ of Σ. Defining $\hat{V}_{0,n} = n^{-1}\hat{\Sigma}_n$, these results imply, in particular, that $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, I_d)$, a result that can be used to construct asymptotically valid and easily implemented confidence intervals for θ.

Although asymptotically valid, the distributional approximation $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta) \overset{a}{\sim} \mathcal{N}(0, I_d)$ might be suspected to be somewhat inaccurate in samples of moderate size due to the presence of the nonparametric estimator of (the derivative of) the density $f(\cdot)$. In particular, folklore and simulation evidence suggests that the distributional properties of kernel-based estimators such as $\hat{\theta}_n$, and studentized versions thereof, can be rather sensitive to the choice of bandwidth h_n. Motivated by concerns of this nature, Nishiyama and Robinson (2000, 2001) developed valid Edgeworth expansions for statistics of the form $\lambda'(\hat{\theta}_n - \theta)/\sqrt{\lambda'\hat{V}_{0,n}\lambda}$ with $\lambda \in \mathbb{R}^d$ and found that in general the magnitude of the error in the distributional approximation $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta) \overset{a}{\sim} \mathcal{N}(0, I_d)$ depends on both the sample size and the bandwidth, this error vanishing at a conventional parametric rate $n^{-1/2}$ only in exceptional circumstances. Subsequently, Nishiyama and Robinson (2005, henceforth NR) developed more detailed expansions and showed that the nonparametric bootstrap provides approximations to the sampling distribution of (a possibly bias-corrected version of) $\lambda'(\hat{\theta}_n - \theta)/\sqrt{\lambda'\hat{V}_{0,n}\lambda}$ that are not merely asymptotically valid, but actually capable of achieving asymptotic refinements.

It is tempting to interpret the latter result as evidence that even in samples of moderate size, highly accurate confidence intervals for θ can be constructed using the bootstrap. To investigate the extent to which this interpretation is warranted, we conducted a Monte Carlo experiment to evaluate the performance of the standard normal and bootstrap approximations to the distribution of $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$. Following NR, the simulation study uses a Tobit model $y_i = \tilde{y}_i \mathbf{1}(\tilde{y}_i > 0)$ with $\tilde{y}_i = x_i'\beta + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0,1)$ independent of the bivariate vector x_i, and $\mathbf{1}(\cdot)$ representing the indicator function. We set $\beta = (1,1)'$ and consider two models: Model 1, also used by NR, employs $(x_{1i}, x_{2i})' \sim \mathcal{N}(0, I_2)$, while Model 2 introduces asymmetry in the regressor distribution by employing $x_{1i} \sim (\chi_4 - 4)/\sqrt{8}$, $x_{2i} \sim \mathcal{N}(0,1)$, and $x_{1i} \perp\!\!\!\perp x_{2i}$, where χ_4 denotes a chi-squared random variable with 4 degrees of freedom. The estimator $\hat{\theta}_n$ is implemented using a fourth-order Gaussian product kernel (i.e., $P = 4$ in Assumption K below). We set $\lambda = (1,0)'$ and consider three 95% confidence intervals:

$$\mathrm{CI}_0 = \left[\ \lambda'\hat{\theta}_n - 1.96\sqrt{\lambda'\hat{V}_{0,n}\lambda},\quad \lambda'\hat{\theta}_n + 1.96\sqrt{\lambda'\hat{V}_{0,n}\lambda}\ \right],$$

$$\mathrm{CI}_0^* = \left[\ \lambda'\hat{\theta}_n - c^*_{0,97.5}\sqrt{\lambda'\hat{V}_{0,n}\lambda},\quad \lambda'\hat{\theta}_n - c^*_{0,2.5}\sqrt{\lambda'\hat{V}_{0,n}\lambda}\ \right],$$

$$\mathrm{CI}_{0,BC}^* = \left[\ \lambda'(\hat{\theta}_n - \hat{\mathcal{B}}_n) - c^*_{0,97.5}\sqrt{\lambda'\hat{V}_{0,n}\lambda},\quad \lambda'(\hat{\theta}_n - \hat{\mathcal{B}}_n) - c^*_{0,2.5}\sqrt{\lambda'\hat{V}_{0,n}\lambda}\ \right],$$

where $c^*_{0,\alpha}$ denotes the αth percentile of the bootstrap approximation and $\hat{\mathcal{B}}_n$ denotes a bias-correction estimate, both implemented as in NR. We conducted 3,000 simulations, each with a sample size $n = 1{,}000$ and 2,000 bootstrap replications.

Figure 1 presents a summary of the Monte Carlo results. To investigate the sensitivity of the empirical coverage probabilities with respect to the bandwidth, these results are presented for a grid of possible bandwidth choices. This figure includes two horizontal lines at 0.90 and at the nominal coverage rate 0.95 for reference, and also plots as vertical lines two (infeasible) bandwidth choices available in the literature proposed by Powell and Stoker (1996) and NR, respectively, denoted h_{PS} and h_{NR}.

In perfect agreement with the theoretical findings of NR, the results for Model 1 indicate that the bootstrap-based confidence intervals without bias-correction (CI_0^*) are more accurate than those based on a standard normal approximation (CI_0) and, in particular, that these bootstrap-based confidence intervals are highly accurate across a nontrivial range of bandwidths. ($\mathrm{CI}_{0,BC}^*$ do not perform well when the bias-correction is estimated.) On the other hand, the results for Model 2 are much less encouraging, indicating, in particular, that the impressive findings about the bootstrap in Model 1 are to some extent an artifact of the particular distributional assumption made on the part of the regressors in that model. Specifically, in the case of Model 2 both approximations are inaccurate outside a narrow

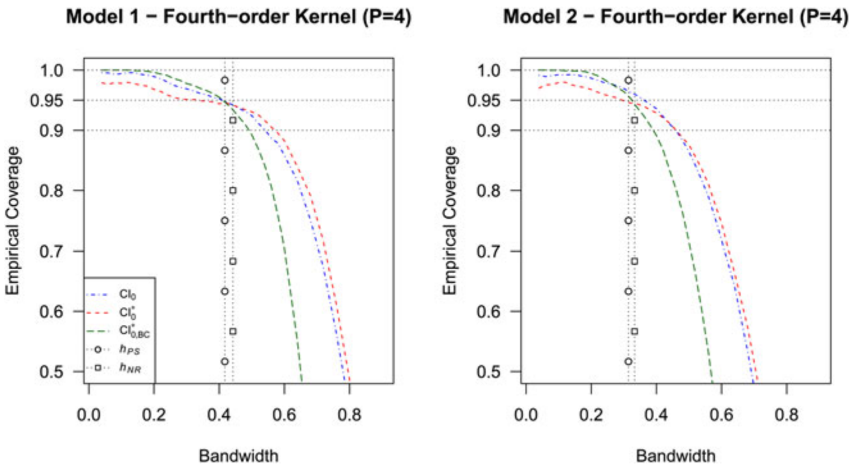

FIGURE 1. Empirical coverage of traditional 95% confidence intervals.

range of bandwidths, although the bootstrap approximation tends to outperform the standard normal approximation.

Particularly noteworthy in the case of Model 2 and, albeit to a somewhat lesser extent in Model 1, are the results for bandwidths that are "small" in the sense that they fall below the optimal bandwidths. Across a wide range of such bandwidths, both confidence intervals are conservative with the degree of conservatism being noticeably larger for the intervals based on the standard normal approximation than for the bootstrap-based intervals. These features appear hard to reconcile with the Edgeworth expansion-based theory of NR and suggest that in the case of the density-weighted average derivative estimator of PSS there is room for improvement when it comes to a theoretical understanding of the properties of the bootstrap in samples of moderate size.

One important objective of this paper is to propose a theory-based explanation of the "small bandwidth" results reported in Figure 1 for the bootstrap, which will be based on the framework of Cattaneo, Crump, and Jansson (2014, henceforth CCJ). (This alternative asymptotic framework was found to deliver predictions consistent with Figure 1's results for the case of the standard normal approximation.) Another goal of the paper is to use this framework to analyze the properties of alternative bootstrap-based procedures. In addition to providing additional novel implications, whose finite-sample relevance will also be present in our simulations, at least one of the theoretical results obtained in pursuit of our goals may be of independent theoretical interest (e.g., Thm. 3).

Remark. For the model and estimator used in the simulations, $h_{PS} \propto n^{-1/6}$ and $h_{NR} \propto n^{-1/6}$, with factors of proportionality that are functionals of the unknown distribution of z. Implementing these selectors with estimated factors of proportionality will likely introduce additional estimation error that will seriously affect the empirical coverage of the resulting data-driven confidence intervals. Cattaneo, Crump, and Jansson (2010) reports results corroborating this conjecture for CI_0 (standard normal approximation). In Section 4.5 we further discuss these implementation issues.

3. PRELIMINARY RESULTS

3.1. Assumptions and Bandwidth Conditions

Throughout the development of our theoretical results we maintain the following standard assumptions.

Assumption M (Model). (a) $\mathbb{E}[y^4] < \infty$, $\mathbb{E}\left[\sigma^2(x)f(x)\right] > 0$ and $\mathbb{V}[\partial e(x)/\partial x - y\partial f(x)/\partial x]$ is positive definite, where $\sigma^2(x) = \mathbb{V}[y|x]$ and $e(x) = f(x)g(x)$.

(b) f is $(Q+1)$ times differentiable, and f and its first $(Q+1)$ derivatives are bounded, for some $Q \geq 2$.

(c) g is twice differentiable, and e and its first two derivatives are bounded.

(d) v is differentiable, and vf and its first derivative are bounded, where $v(x) = \mathbb{E}\left[y^2|x\right]$.

(e) $\lim_{\|x\|\to\infty}[f(x) + |e(x)|] = 0$, where $\|\cdot\|$ is the Euclidean norm.

Assumption K (Kernel). (a) K is even and differentiable, and K and its first derivative are bounded.

(b) $\int_{\mathbb{R}^d} \dot{K}(u)\dot{K}(u)'\mathrm{d}u$ is positive definite, where $\dot{K}(u) = \partial K(u)/\partial u$.

(c) For some $P \geq 2$, $\int_{\mathbb{R}^d} |K(u)|\left(1 + \|u\|^P\right)\mathrm{d}u + \int_{\mathbb{R}^d} \|\dot{K}(u)\|(1 + \|u\|^2)\mathrm{d}u < \infty$, and

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

The purpose of the following assumption is to ensure that the smoothing bias of the estimator $\hat{\theta}_n$ is asymptotically negligible (relative to its standard deviation).

Assumption B. (Bias). $\min\left(nh_n^{d+2}, 1\right)nh_n^{2s} \to 0$, where $s = \min(P, Q)$.

Finally, the following conditions will play a crucial role in our theoretical developments.

Condition AL. (*Asymptotic Linearity*) $nh_n^{d+2} \to \infty$.

Condition AN. (*Asymptotic Normality*) $nh_n^{d/2} \to \infty$.

Conditions AL and AN are nested, the latter being significantly weaker than the former by accommodating bandwidths that are "small" in the sense that the sequence h_n is allowed to converge more rapidly to zero than is permitted by Condition AL. While the traditional Gaussian and bootstrap distributional approximations employ Condition AL, our alternative approximation framework relaxes this condition, employing instead Condition AN.

3.2. Gaussian Approximation

To further appreciate the distinction between Conditions AL and AN, observe that $\hat{\theta}_n = \hat{\theta}_n(h_n)$ admits the (n-varying) U-statistic representation:

$$\hat{\theta}_n(h) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} U\left(z_i, z_j; h\right),$$

$$U(z_i, z_j; h) = -h^{-(d+1)} \dot{K}\left(\frac{x_i - x_j}{h}\right)\left(y_i - y_j\right),$$

which leads to the Hoeffding decomposition $\hat{\theta}_n - \theta = \mathcal{B}_n + \bar{L}_n + \bar{W}_n$, where $\mathcal{B}_n = \theta(h_n) - \theta$ with $\theta(h) = \mathbb{E}\left[U\left(z_i, z_j; h\right)\right]$, $\bar{L}_n = n^{-1}\sum_{i=1}^{n} L\left(z_i; h_n\right)$ with $L(z_i; h) = 2\left[\mathbb{E}[U(z_i, z_j; h)|z_i] - \theta(h)\right]$, and $\bar{W}_n = \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W\left(z_i, z_j; h_n\right)$ with $W\left(z_i, z_j; h\right) = U\left(z_i, z_j; h\right) - \left(L(z_i; h) + L(z_j; h)\right)/2 - \theta(h)$. It can be shown that if Assumptions M and K hold, then

$$\sqrt{n}(\hat{\theta}_n - \theta) = \underbrace{\sqrt{n}\mathcal{B}_n}_{O(\sqrt{n}h_n^s)} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L(z_i; h_n) + \underbrace{\sqrt{n}\bar{W}_n}_{O_p\left(1/\sqrt{nh_n^{d+2}}\right)}$$

and $\bar{L}_n = n^{-1/2}\sum_{i=1}^{n} L(z_i) + o_p(1)$ whenever $h_n \to 0$. As a consequence, Assumption B and Condition AL are sufficient for the asymptotic linearity result (1), as shown by PSS.

Condition AL helps ensure asymptotic linearity of $\hat{\theta}_n$ by rendering the "remainder" term $\bar{W}_n$ asymptotically negligible. In contrast, CCJ showed that if Assumptions M, K, and B hold and if Condition AN is satisfied, then Condition AL can be removed, and obtained the alternative Gaussian approximation

$$V_n^{-1/2}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, I_d), \quad V_n = n^{-1}\Sigma + \binom{n}{2}^{-1} h_n^{-(d+2)}\Delta, \tag{2}$$

where $\Delta = 2\mathbb{E}\left[\sigma^2(x)f(x)\right]\int_{\mathbb{R}^d} \dot{K}(u)\dot{K}(u)'\mathrm{d}u$. This result shows that while failure of Condition AL leads to a failure of asymptotic linearity, asymptotic normality of $\hat{\theta}_n$ holds under the significantly weaker Condition AN, which permits failure not only of asymptotic linearity, but also of $\sqrt{n}$-consistency when $nh_n^{d+2} \to 0$ (and even of consistency when $\underline{\lim}_{n\to\infty} nh_n^{d/2+1} < \infty$).

A key result exploited in the derivation of the asymptotic normality result (2) is that the degenerate U-statistic $\bar{W}_n$ is itself asymptotically normal under the stated conditions: $\sqrt{n^2 h_n^{d+2}}\bar{W}_n \rightsquigarrow \mathcal{N}(0, 2\Delta)$. Therefore, and in sharp contrast to the distributional approximation $\hat{\theta}_n \overset{a}{\sim} \mathcal{N}(\theta, n^{-1}\Sigma)$ suggested by (1), the distributional approximation $\hat{\theta}_n \overset{a}{\sim} \mathcal{N}(\theta, V_n)$ suggested by (2) does not ignore the variability in the "remainder" term $\bar{W}_n$. This latter feature seems desirable when finite sample accuracy of conventional distributional approximations is a concern, as is the case here.

Because the distributional approximation suggested by (2) is normal, asymptotic standard normality of studentized estimators can be achieved also when Condition AL is replaced by Condition AN, provided that the variance estimator $\hat{V}_n$ (say) used for studentization purposes satisfies $V_n^{-1}\hat{V}_n \to_p I_d$ under Condition AN. PSS's estimator $\hat{\Sigma}_n$ of Σ mentioned in Section 2 is (proportional to) the jackknife variance estimator of $\hat{\theta}_n(h)$, being of the form

$$\hat{\Sigma}_n = \hat{\Sigma}_n(h_n) = \frac{1}{n}\sum_{i=1}^{n} \hat{L}_{n,i}(h_n)\hat{L}_{n,i}(h_n)',$$

$$\hat{L}_{n,i}(h) = 2\left(\frac{1}{n-1}\sum_{j=1, j\neq i}^{n} U(z_i, z_j; h) - \hat{\theta}_n(h)\right).$$

It was shown by CCJ that

$$\hat{V}_{0,n} = n^{-1}\hat{\Sigma}_n(h_n) = n^{-1}\left[\Sigma + o_p(1)\right] + 2\binom{n}{2}^{-1} h_n^{-(d+2)}\left[\Delta + o_p(1)\right].$$

This expansion, which will play an important role in the present study of the bootstrap, implies, in particular, that validity of $\hat{V}_{0,n}$ requires Condition AL. The lack of "robustness" of $\hat{V}_{0,n}$ with respect to h_n can be avoided by employing either of the variance estimators

$$\hat{V}_{1,n} = \hat{V}_{0,n} - \binom{n}{2}^{-1} h_n^{-(d+2)}\hat{\Delta}_n(h_n) \quad \text{and} \quad \hat{V}_{2,n} = n^{-1}\hat{\Sigma}_n\left(2^{1/(d+2)}h_n\right),$$

where $\hat{\Delta}_n(h) = h^{d+2}\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\hat{W}_{n,ij}(h)\hat{W}_{n,ij}(h)'$ with $\hat{W}_{n,ij}(h) = U\left(z_i, z_j; h\right) - \left(\hat{L}_{n,i}(h) + \hat{L}_{n,j}(h)\right)/2 - \hat{\theta}_n(h)$.

Remark. It can be shown that the adjustment employed in the construction of $\hat{V}_{1,n}$ is asymptotically equivalent to the bias-correction proposed by Efron and Stein (1981). The multiplicative factor $2^{1/(d+2)}$ involved in the construction of $\hat{V}_{2,n}$ is designed to yield equality between the terms premultiplying Δ in the expansions of V_n and $\hat{V}_{2,n}$.

The following result is adapted from CCJ and formulated in a manner that facilitates comparison with the main theorems given below.

LEMMA 1. *Suppose Assumptions M, K, and B hold and suppose Condition AN is satisfied.*

(a) *The following are equivalent:*

i. *Condition AL is satisfied.*

ii. $V_n^{-1}\hat{V}_{0,n} \to_p I_d$.

iii. $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, I_d)$.

(b) *If* nh_n^{d+2} *is convergent in* $\bar{\mathbb{R}}_+ = [0,\infty]$*, then* $\hat{V}_{0,n}^{-1/2}\left(\hat{\theta}_n - \theta\right) \rightsquigarrow \mathcal{N}(0,\Omega_0)$*, where*

$$\Omega_0 = \lim_{n\to\infty}\left(nh_n^{d+2}\Sigma + 4\Delta\right)^{-1/2}\left(nh_n^{d+2}\Sigma + 2\Delta\right)\left(nh_n^{d+2}\Sigma + 4\Delta\right)^{-1/2}.$$

(c) *For* $k \in \{1,2\}$*,* $V_n^{-1}\hat{V}_{k,n} \to_p I_d$ *and* $\hat{V}_{k,n}^{-1/2}\left(\hat{\theta}_n - \theta\right) \rightsquigarrow \mathcal{N}(0, I_d)$.

Part (a) is a qualitative result highlighting the crucial role played by Condition AL in connection with asymptotic validity of inference procedures based on $\hat{V}_{0,n}$. The equivalence between (i) and (iii) shows that Condition AL is necessary and sufficient for the test statistic $\hat{V}_{0,n}^{-1/2}\left(\hat{\theta}_n - \theta\right)$ proposed by PSS to be asymptotically pivotal. In turn, this equivalence is a special case of part (b), which is a quantitative result that can furthermore be used to characterize the consequences of relaxing Condition AL. Specifically, part (b) shows that also under departures from Condition AL the statistic $\hat{V}_{0,n}^{-1/2}\left(\hat{\theta}_n - \theta\right)$ can be asymptotically normal with mean zero, but with a variance matrix Ω_0 whose value depends on the limiting value of nh_n^{d+2}. This matrix satisfies $I_d/2 \le \Omega_0 \le I_d$ (in a positive semidefinite sense) and takes on the limiting values $I_d/2$ and I_d when $\lim_{n\to\infty} nh_n^{d+2}$ equals 0 and ∞, respectively. By implication, part (b) indicates that inference procedures based on the test statistic proposed by PSS will be conservative across a nontrivial range of bandwidths. In contrast, part (c) shows that studentization by means of $\hat{V}_{1,n}$ and $\hat{V}_{2,n}$ achieves asymptotic pivotality across the full range of bandwidth sequences allowed by Condition AN, suggesting, in particular, that coverage probabilities of confidence intervals constructed using these variance estimators will be close to their nominal level across a nontrivial range of bandwidths.

3.3. Bootstrap Approximation

We study two variants of the m-out-of-n replacement bootstrap with $m = m(n) \to \infty$: the standard nonparametric bootstrap ($m = n$) and the variant where m is a vanishing fraction of n (i.e., $m/n \to 0$), calling the latter "m-out-of-n bootstrap" for short. (Here, and elsewhere in the sequel, the dependence of $m(n)$ on n will often be suppressed to achieve notational economy.) Specifically, to describe the bootstrap procedure(s), let z_i^*, $i = 1,\ldots,m$, be a random sample with replacement from the observed sample $\mathcal{Z}_n = \{z_1,\ldots,z_n\}$. The bootstrap analog of $\hat{\theta}_n$ is

$$\hat{\theta}_n^* = \hat{\theta}_n^*(h_m) = \binom{m}{2}^{-1}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} U\left(z_i^*, z_j^*; h_m\right),$$

while the bootstrap analogs of $\hat{\Sigma}_n$ and $\hat{\Delta}_n$ are $\hat{\Sigma}_n^* = \hat{\Sigma}_n^*(h_m)$ and $\hat{\Delta}_n^* = \hat{\Delta}_n^*(h_m)$, respectively, where

$$\hat{\Sigma}_n^*(h) = \frac{1}{m}\sum_{i=1}^{m} \hat{L}_{n,i}^*(h)\hat{L}_{n,i}^*(h)',$$

$$\hat{L}_{n,i}^*(h) = 2\left(\frac{1}{m-1}\sum_{j=1,j\neq i}^{m} U\left(z_i^*, z_j^*; h\right) - \hat{\theta}_n^*(h)\right),$$

and, defining $\hat{W}_{n,ij}^*(h) = U\left(z_i^*, z_j^*; h\right) - \left(\hat{L}_{n,i}^*(h) + \hat{L}_{n,j}^*(h)\right)/2 - \hat{\theta}_n^*(h)$,

$$\hat{\Delta}_n^*(h) = \binom{m}{2}^{-1} h^{d+2} \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \hat{W}_{n,ij}^*(h)\hat{W}_{n,ij}^*(h)'.$$

Finally, the bootstrap analogs of $\hat{V}_{0,n}$, $\hat{V}_{1,n}$, and $\hat{V}_{2,n}$ are $\hat{V}_{0,n}^* = m^{-1}\hat{\Sigma}_n^*(h_m)$,

$$\hat{V}_{1,n}^* = \hat{V}_{0,n}^* - \binom{m}{2}^{-1} h_m^{-(d+2)}\hat{\Delta}_n^*(h_m), \quad \text{and} \quad \hat{V}_{2,n}^* = m^{-1}\hat{\Sigma}_n^*\left(2^{1/(d+2)}h_m\right).$$

Remark. The m-out-of-n bootstrap is closely related to subsampling (i.e., the m-out-of-n nonreplacement bootstrap). The properties of subsampling are immediate consequences of Lemma 1(b) and (c) and Politis and Romano (1994). In particular, for $k \in \{1, 2\}$ consistency of the subsampling approximation to the distribution of $\hat{V}_{k,n}^{-1/2}(\hat{\theta}_n - \theta)$ is automatic (under the assumptions of Lemma 1) whenever $m/n \to 0$ and the following (mild) additional assumption holds: If $nh_n^{d+2} \to 0$, then $(m/n)^2 (h_m/h_n)^{d+2} \to 0$. Also, under the same assumptions the subsampling approximation to the distribution of $\hat{V}_{0,n}^{-1/2}\left(\hat{\theta}_n - \theta\right)$ is consistent whenever nh_n^{d+2} is convergent in $\bar{\mathbb{R}}_+$. As will be shown in Theorem 1(c), Theorem 2, and Theorem 3(b) below, these properties are shared by m-out-of-n bootstrap studied in this paper.

Let $\mathbb{P}^*$, $\mathbb{E}^*$, or $\mathbb{V}^*$ denote a probability or moment computed under the bootstrap distribution conditional on $\mathcal{Z}_n$, and let $\rightsquigarrow_p$ denote weak convergence in probability (e.g., Gine and Zinn, 1990). Also, define $\theta_n^* = \theta^*(h_m)$, where $\theta^*(h) = \mathbb{E}^*\left[U(z_i^*, z_j^*; h)\right] = (n-1)\hat{\theta}_n(h)/n$. The main results of this paper follow from Lemma 1 and the following lemma.

LEMMA 2. *Suppose Assumptions M and K hold, suppose Condition AN is satisfied, and suppose $h_n \to 0$, $m \to \infty$, and $\overline{\lim}_{n\to\infty} m/n < \infty$.*

(a) $V_n^{*-1}\mathbb{V}^*\left[\hat{\theta}_n^*\right] \to_p I_d$, *where*

$$V_n^* = m^{-1}\Sigma + \left(1 + 2\frac{m}{n}\right)\binom{m}{2}^{-1} h_m^{-(d+2)}\Delta.$$

(b) ${\Sigma_n^*}^{-1}\hat{\Sigma}_n^* \to_p I_d$ *and* $\Delta^{-1}\hat{\Delta}_n^* \to_p I_d$, *where*

$$\Sigma_n^* = \Sigma + 2m\left(1+\frac{m}{n}\right)\binom{m}{2}^{-1} h_m^{-(d+2)}\Delta.$$

(c) ${V_n^*}^{-1/2}\left(\hat{\theta}_n^* - \theta_n^*\right) \rightsquigarrow_p \mathcal{N}(0, I_d)$.

The (conditional on $\mathcal{Z}_n$) Hoeffding decomposition gives $\hat{\theta}_n^* = \theta^*(h_m) + \bar{L}_n^* + \bar{W}_n^*$, where

$$\bar{L}_n^* = m^{-1}\sum_{i=1}^{m} L^*\left(z_i^*; h_m\right), \quad \bar{W}_n^* = \binom{m}{2}^{-1}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} W^*\left(z_i^*, z_j^*; h_m\right),$$

with $L^*\left(z_i^*; h\right) = 2\left(\mathbb{E}^*\left[U\left(z_i^*, z_j^*; h\right)|z_i^*\right] - \theta^*(h)\right)$ and $W^*\left(z_i^*, z_j^*; h\right) = U\left(z_i^*, z_j^*; h\right) - \left(L^*\left(z_i^*; h\right) + L^*\left(z_j^*; h\right)\right)/2 - \theta^*(h)$. Lemma 2 (a) is obtained from this decomposition by noting that

$$\mathbb{V}^*\left[\hat{\theta}_n^*\right] = m^{-1}\mathbb{V}^*\left[L^*\left(z_i^*; h_m\right)\right] + \binom{m}{2}^{-1}\mathbb{V}^*\left[W^*\left(z_i^*, z_j^*; h_m\right)\right],$$

where, with "$A_n \approx B_n$" being shorthand for $A_n^{-1}B_n \to_p I_d$,

$$\mathbb{V}^*\left[L^*(z_i^*; h_m)\right] \approx \hat{\Sigma}_n(h_m) \approx \Sigma + 2\frac{m^2}{n}\binom{m}{2}^{-1} h_m^{-(d+2)}\Delta$$

and $\mathbb{V}^*\left[W^*(z_i^*, z_j^*; h_m)\right] \approx h_m^{-(d+2)}\hat{\Delta}_n(h_m) \approx h_m^{-(d+2)}\Delta$.

The bootstrap estimator of the variance of $\hat{\theta}_n$ is $\mathbb{V}^*\left[\hat{\theta}_n^*\right]$ with $m = n$. In view of the foregoing, this estimator exceeds $n^{-1}\mathbb{V}^*[L^*(z_i^*; h_n)] \approx n^{-1}\hat{\Sigma}_n(h_n) = \hat{V}_{0,n}$, implying that the bootstrap variance estimator exhibits an upward bias even greater than that of $\hat{V}_{0,n}$. In particular, the bootstrap variance estimator is inconsistent whenever PSS's variance estimator is, a result also contained in Theorem 1 below. This failure of the bootstrap is attributable solely to its inability to consistently estimate the variability of the term $\bar{L}_n$ in the Hoeffding decomposition of $\hat{\theta}_n$, since $\mathbb{V}^*\left[W^*(z_i^*, z_j^*; h_n)\right] \approx h_n^{-(d+2)}\Delta$ implies that the variability of $\bar{W}_n$ is estimated consistently.

The proof of Lemma 2(b) shows that

$$\hat{\Sigma}_n^* \approx \hat{\Sigma}_n(h_m) + 2m\binom{m}{2}^{-1} h_m^{-(d+2)}\hat{\Delta}_n(h_m),$$

implying that the asymptotic behavior of $\hat{\Sigma}_n^*$ differs from that of $\hat{\Sigma}_n(h_m)$ whenever Condition AL fails. Finally, Lemma 2(c) is a bootstrap counterpart of (2), giving a weak convergence in probability result for $\hat{\theta}_n^*$ without requiring asymptotic linearity.

Remark. By continuity of the d-variate standard normal cdf $\Phi_d(\cdot)$ and Polya's theorem for weak convergence in probability (e.g., Xiong and Li, 2008, Thm. 3.5), Lemma 2(c) is equivalent to the statement that

$$\sup_{t\in\mathbb{R}^d}\left|\mathbb{P}^*\left[V_n^{*-1/2}(\hat{\theta}_n^*-\theta_n^*)\le t\right]-\Phi_d(t)\right|\to_p 0. \tag{3}$$

By arguing along subsequences, it can be shown that a sufficient condition for (3) is the following (uniform) Cramér-Wold-type condition:

$$\sup_{\lambda\in\Lambda_d}\sup_{t\in\mathbb{R}^d}\left|\mathbb{P}^*\left[\lambda'(\hat{\theta}_n^*-\theta_n^*)/\sqrt{\lambda' V_n^*\lambda}\le t\right]-\Phi_1(t)\right|\to_p 0, \tag{4}$$

where $\Lambda_d=\left\{\lambda\in\mathbb{R}^d:\lambda'\lambda=1\right\}$ denotes the unit sphere in $\mathbb{R}^d$. The proof of Lemma 2(c) uses the theorem of Heyde and Brown (1970) to verify (4). In contrast to the case of unconditional joint weak convergence, it would appear to be an open question whether a pointwise Cramér-Wold condition such as

$$\sup_{t\in\mathbb{R}^d}\left|\mathbb{P}^*\left[\lambda'(\hat{\theta}_n^*-\theta_n^*)/\sqrt{\lambda' V_n^*\lambda}\le t\right]-\Phi_1(t)\right|\to_p 0,\quad\forall\lambda\in\Lambda_d,$$

implies weak convergence in probability of $V_n^{*-1/2}\left(\hat{\theta}_n^*-\theta_n^*\right)$, and for this reason we establish the stronger result (4) in the Appendix.

4. MAIN RESULTS

4.1. Bootstrapping PSS's Estimator and Test Statistic

To anticipate our findings, notice that Lemma 1 gives

$$\mathbb{V}[\hat{\theta}_n]\approx n^{-1}\Sigma+\binom{n}{2}^{-1}h_n^{-(d+2)}\Delta\quad\text{and}\quad\hat{V}_{0,n}\approx n^{-1}\Sigma+2\binom{n}{2}^{-1}h_n^{-(d+2)}\Delta,$$

whereas in the case of the nonparametric bootstrap (when $m=n$) Lemma 2 gives

$$\mathbb{V}^*\left[\hat{\theta}_n^*\right]\approx n^{-1}\Sigma+3\binom{n}{2}^{-1}h_n^{-(d+2)}\Delta\quad\text{and}\quad\hat{V}_{0,n}^*\approx n^{-1}\Sigma+4\binom{n}{2}^{-1}h_n^{-(d+2)}\Delta,$$

strongly indicating that Condition AL is crucial for consistency of the bootstrap. On the other hand, in the case of the m-out-of-n bootstrap (when $m/n\to 0$), Lemma 2 gives

$$\mathbb{V}^*[\hat{\theta}_n^*]\approx m^{-1}\Sigma+\binom{m}{2}^{-1}h_m^{-(d+2)}\Delta\quad\text{and}\quad\hat{V}_{0,n}^*\approx m^{-1}\Sigma+2\binom{m}{2}^{-1}h_m^{-(d+2)}\Delta,$$

suggesting that consistency of the m-out-of-n bootstrap might hold even if Condition AL fails, at least in those cases where $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n-\theta)$ converges in distribution.

(By Lemma 1(b), convergence in distribution of $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$ occurs when nh_n^{d+2} is convergent in $\bar{\mathbb{R}}_+$.)

The following result, which follows from Lemmas 1 and 2 and the continuous mapping theorem for weak convergence in probability (e.g., Xiong and Li, 2008, Thm. 3.1), makes the preceding heuristics precise.

THEOREM 1. *Suppose the assumptions of Lemma 1 hold.*

(a) *If $m = n$, then the following are equivalent:*

i. *Condition AL is satisfied.*
ii. $V_n^{-1}\mathbb{V}^*[\hat{\theta}_n^*] \to_p I_d$.
iii. $\sup_{t\in\mathbb{R}^d}\left|\mathbb{P}^*\left[\hat{\theta}_n^* - \theta_n^* \le t\right] - \mathbb{P}\left[\hat{\theta}_n - \theta \le t\right]\right| \to_p 0$.
iv. $\sup_{t\in\mathbb{R}^d}\left|\mathbb{P}^*\left[\hat{V}_{0,n}^{*-1/2}(\hat{\theta}_n^* - \theta_n^*) \le t\right] - \mathbb{P}\left[\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta) \le t\right]\right| \to_p 0$.

(b) *If $m = n$ and if nh_n^{d+2} is convergent in $\bar{\mathbb{R}}_+$, then $\hat{V}_{0,n}^{*-1/2}\left(\hat{\theta}_n^* - \theta_n^*\right) \rightsquigarrow_p \mathcal{N}\left(0, \Omega_0^*\right)$, where*

$$\Omega_0^* = \lim_{n\to\infty}\left(nh_n^{d+2}\Sigma + 8\Delta\right)^{-1/2}\left(nh_n^{d+2}\Sigma + 6\Delta\right)\left(nh_n^{d+2}\Sigma + 8\Delta\right)^{-1/2}.$$

(c) *If $m^{-1} + m/n \to 0$ and if nh_n^{d+2} is convergent in $\bar{\mathbb{R}}_+$, then $\hat{V}_{0,n}^{*-1/2}\left(\hat{\theta}_n^* - \theta_n^*\right) \rightsquigarrow_p \mathcal{N}(0, \Omega_0)$.*

In an obvious way, Theorem 1(a) and (b) can be viewed as a bootstrap analog of Lemma 1(a) and (b). In particular, Theorem 1(a) shows that Condition AL is necessary and sufficient for consistency of the nonparametric bootstrap and therefore implies that the nonparametric bootstrap is inconsistent whenever the estimator is not asymptotically linear (when $\overline{\lim}_{n\to\infty} nh_n^{d+2} < \infty$), including, in particular, the knife-edge case $nh_n^{d+2} \to \kappa \in (0,\infty)$ where the estimator is $\sqrt{n}$-consistent and asymptotically normal (we discuss this issue further in Section 4.4). The implication (i) $\Rightarrow$ (iv) in Theorem 1(a) is essentially due to NR. (Their results are obtained under slightly stronger assumptions than those of Lemma 1 and require $nh_n^{d+3}/(\log n)^9 \to \infty$.) On the other hand, the result that Condition AL is necessary for bootstrap consistency would appear to be new.

Theorem 1(b) can be used to quantify the severity of the bootstrap inconsistency under departures from Condition AL. The extent of the failure of the bootstrap to approximate the asymptotic distribution of the test statistic is captured by the variance matrix Ω_0^*, which satisfies $3I_d/4 \le \Omega_0^* \le I_d$ and takes on the limiting values $3I_d/4$ and I_d when $\lim_{n\to\infty} nh_n^{d+2}$ equals 0 and ∞, respectively. Interestingly, comparing Theorem 1(b) with Lemma 1(b), the nonparametric bootstrap approximation to the distribution of $\hat{V}_{0,n}^{-1/2}\left(\hat{\theta}_n - \theta\right)$ is seen to be superior to the standard normal approximation, because $\Omega_0 \le \Omega_0^* \le I_d$, both inequalities being strict when Condition AL fails. In other words, replacing Condition AL by

Condition AN yields the prediction that bootstrap-based confidence intervals "should" be conservative (albeit less so than confidence intervals based on standard normal approximations) when bandwidths are "small". In combination, Theorem 1(b) with Lemma 1(b), therefore, provide a theory-based explanation of the simulation evidence in Figure 1.

Theorem 1(c) shows that a sufficient condition for consistency of m-out-of-n bootstrap is convergence of nh_n^{d+2} in $\bar{\mathbb{R}}_+$. To illustrate what can happen when the latter condition fails, suppose nh_n^{d+2} is "large" when n is even and "small" when n is odd. Specifically, suppose that $nh_{2n}^{d+2} \to \infty$ and $nh_{2n+1}^{d+2} \to 0$. Then, if m is even for every n, it follows from Theorem 1(c) that $\hat{V}_{0,n}^{*-1/2}(\hat{\theta}_n^* - \theta_n^*) \rightsquigarrow_p \mathcal{N}(0, I_d)$, whereas, by Lemma 1(b), $\hat{V}_{0,2n+1}^{-1/2}(\hat{\theta}_{2n+1} - \theta) \rightsquigarrow \mathcal{N}(0, I_d/2)$.

Remark. (i) The example just given is intentionally extreme, but the qualitative message that consistency of m-out-of-n bootstrap can fail when $\lim_{n\to\infty} nh_n^{d+2}$ does not exist is valid more generally. Indeed, Theorem 1(c) admits the following partial converse: if nh_n^{d+2} is not convergent in $\bar{\mathbb{R}}_+$, then there exists a sequence $m = m(n)$ such that ($m \to \infty$, $m/n \to 0$, and)

$$\sup_{t\in\mathbb{R}^d} \left| \mathbb{P}^*\left[\hat{V}_{0,n}^{*-1/2}(\hat{\theta}_n^* - \theta_n^*) \le t\right] - \mathbb{P}\left[\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta) \le t\right] \right| \not\to_p 0.$$

In other words, employing critical values obtained by means of the m-out-of-n bootstrap does not automatically "robustify" an inference procedure based on PSS's statistic.

(ii) Applying Lemma 1(b) and Politis and Romano (1994), it can be shown that the previous remark also applies to subsampling. In other words, the subsampling approximation to the distribution of $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$ can be inconsistent whenever nh_n^{d+2} is not convergent in $\bar{\mathbb{R}}_+$.

4.2. Bootstrapping "Robust" Test Statistics

Because $\hat{V}_{1,n}^{-1/2}(\hat{\theta}_n - \theta)$ and $\hat{V}_{2,n}^{-1/2}(\hat{\theta}_n - \theta)$ are both asymptotically standard normal under the assumptions of Lemma 1, folklore suggests that the bootstrap should be capable of consistently estimating their distributions. In the case of the statistic studentized by means of $\hat{V}_{1,n}$, this conjecture turns out to be correct, essentially because it follows from Lemma 2 that

$$\hat{V}_{1,n}^* \approx m^{-1}\Sigma + \left(1 + 2\frac{m}{n}\right)\binom{m}{2}^{-1} h_m^{-(d+2)}\Delta \approx \mathbb{V}^*[\hat{\theta}_n^*].$$

More precisely, an application of Lemma 2 and the continuous mapping theorem for weak convergence in probability yields the following result.

THEOREM 2. *If the assumptions of Lemma 1 hold, $m \to \infty$, and if $\overline{\lim}_{n\to\infty} m/n < \infty$, then $\hat{V}_{1,n}^{*-1/2}\left(\hat{\theta}_n^* - \theta_n^*\right) \rightsquigarrow_p \mathcal{N}(0, I_d)$.*

Theorem 2 demonstrates by example that even if Condition AL fails it is possible, by proper choice of variance estimator, to achieve consistency of the nonparametric bootstrap estimator of the distribution of a studentized version of PSS's estimator.

In the case of the m-out-of-n bootstrap, consistency of the approximation to the distribution of $\hat{V}_{1,n}^{-1/2}(\hat{\theta}_n - \theta)$ is unsurprising in light of its asymptotic pivotality, and it is natural to expect an analogous result holds for $\hat{V}_{2,n}^{-1/2}\left(\hat{\theta}_n - \theta\right)$. On the other hand, in the case of the nonparametric bootstrap it follows from Lemma 2 that

$$\hat{V}_{2,n}^* \approx n^{-1}\Sigma + 2\binom{n}{2}^{-1} h_n^{-(d+2)}\Delta \approx \mathbb{V}^*\left[\hat{\theta}_n^*\right] - \binom{n}{2}^{-1} h_n^{-(d+2)}\Delta,$$

suggesting that Condition AL will be required for consistency in the case of $\hat{V}_{2,n}^{-1/2}\left(\hat{\theta}_n - \theta\right)$.

THEOREM 3. *Suppose the assumptions of Lemma 1 hold.*

(a) *If $m = n$ and if nh_n^{d+2} is convergent in $\bar{\mathbb{R}}_+$, then $\hat{V}_{2,n}^{*-1/2}\left(\hat{\theta}_n^* - \theta_n^*\right) \rightsquigarrow_p \mathcal{N}\left(0, \Omega_2^*\right)$, where*

$$\Omega_2^* = \lim_{n\to\infty}\left(nh_n^{d+2}\Sigma + 4\Delta\right)^{-1/2}\left(nh_n^{d+2}\Sigma + 6\Delta\right)\left(nh_n^{d+2}\Sigma + 4\Delta\right)^{-1/2}.$$

In particular, $\hat{V}_{2,n}^{*-1/2}(\hat{\theta}_n^* - \theta_n^*) \rightsquigarrow_p \mathcal{N}(0, I_d)$ if and only if Condition AL is satisfied.

(b) *If $m^{-1} + m/n \to 0$, then $\hat{V}_{2,n}^{*-1/2}(\hat{\theta}_n^* - \theta_n^*) \rightsquigarrow_p \mathcal{N}(0, I_d)$.*

While there is no shortage of examples of bootstrap failure in the literature, it seems surprising that the nonparametric bootstrap fails to approximate the distribution of the asymptotically pivotal statistic $\hat{V}_{2,n}^{-1/2}\left(\hat{\theta}_n - \theta\right)$ whenever Condition AL is violated. (Counter-example 1 of Bickel and Freedman (1981) is also concerned with U-statistics, but the bootstrap failure reported there is due to a violation of their (von Mises) condition (6.5) whose natural counterpart is automatically satisfied here.) Intuitively, the failure of the nonparametric bootstrap for this statistic follows naturally from the results of Theorem 1. The logic underpinning the form of $\hat{V}_{2,n}$ is that we can scale h_n up by the appropriate constant, $2^{1/(d+2)}$, to offset the bias of the untransformed estimator $\hat{V}_{0,n}$. However, by Theorem 1, $\Omega_0 < \Omega_0^* < I_d$ when Condition AL fails and so the closer approximation of the bootstrap-based statistic to the standard normal distribution implies that the factor $2^{1/(d+2)}$ overcompensates. This leads directly to the invalidity result in Theorem 3(a). The degree of this overcompensation is measured by the variance matrix Ω_2^*, which satisfies $I_d \leq \Omega_2^* \leq 3I_d/2$, implying that inference based on the bootstrap approximation to the distribution of $\hat{V}_{2,n}^{-1/2}(\hat{\theta}_n - \theta)$ will be asymptotically conservative.

Remark. In light of the above discussion, a variation on the idea underlying the construction of $\hat{V}_{2,n}$ can be used to construct a test statistic whose bootstrap distribution validly approximates the distribution of PSS's statistic under the assumptions of Lemma 1. Specifically, because it follows from Lemmas 1 and 2 that $\mathbb{V}^*[\hat{\theta}_n^*(3^{1/(d+2)}h_n)] \approx \mathbb{V}[\hat{\theta}_n]$ and $\hat{V}_{2,n}^* \approx \hat{V}_{0,n}$, it can be shown that if the assumptions of Lemma 1 hold, then

$$\sup_{t\in\mathbb{R}^d}\Big|\mathbb{P}^*\left[\hat{V}_{2,n}^{*-1/2}\left(\hat{\theta}_n^*\left(3^{1/(d+2)}h_n\right) - \theta_n^*\left(3^{1/(d+2)}h_n\right)\right) \leq t\right]$$
$$-\mathbb{P}\left[\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta) \leq t\right]\Big| \to_p 0,$$

even if nh_n^{d+2} does not converge. Admittedly, this construction is mainly of theoretical interest, but it does seem noteworthy that this resampling procedure works even in the case where the m-out-of-n bootstrap might fail.

4.3. Summary of Theoretical Results

The main results of this paper are summarized in Table 1, which describe the limiting distributions of the three test statistics $\hat{V}_{k,n}^{-1/2}(\hat{\theta}_n - \theta)$ $(k = 1, 2, 3)$ as well as the limiting distributions (in probability) of their bootstrap analogs. Each panel corresponds to one test statistic and includes three rows corresponding to each approximation used (large sample distribution, nonparametric bootstrap, and m-out-of-n bootstrap, respectively). Each column analyzes a subset of possible bandwidth sequences, which leads to different approximations in general. The only statistic that remains valid in all cases is $\hat{V}_{1,n}^{-1/2}(\hat{\theta}_n - \theta)$. For PSS's statistic $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$ both the nonparametric bootstrap and the m-out-of-n bootstrap (and subsampling) are invalid in general, while for $\hat{V}_{2,n}^{-1/2}(\hat{\theta}_n - \theta)$ only the m-out-of-n bootstrap (and subsampling) is valid in general. As discussed above, the direction and "worst case" magnitude of the "bias" of the bootstrap can be extracted from the $\kappa = 0$ column of Table 1. Finally, the " $-$" entries in the last column of Table 1 serve as remainders that that when nh_n^{d+2} is not convergent in $\bar{\mathbb{R}}_+$, weak convergence (in probability) of bootstrap distribution estimators is not guaranteed in general.

4.4. Implications and Further Discussion

To further describe the key implications of our theoretical work, we consider two of the most common approaches to conduct bootstrap-based inference in empirical work: (i) Efron-type confidence intervals and (ii) bootstrap-based variance-covariance estimators.[2] For simplicity, we focus on conducting inference on $\lambda'\theta$, with $\lambda \in \mathbb{R}^d$. Let $\{\hat{\theta}_{n,b}^*(h_n) : b = 1, 2, \cdots, B\}$ be a nonparametric $(m = n)$ bootstrap sample of size B of the semiparametric estimator $\hat{\theta}_n(h_n)$, and set

TABLE 1. Summary of main results

Test statistic	Distributional approximation	$\lim_{n\to\infty} nh_n^{d+2} = \kappa \in \bar{\mathbb{R}}_+$ $\kappa = \infty$	$\kappa \in (0,\infty)$	$\kappa = 0$	$\underline{\lim}_{n\to\infty} nh_n^{d+2} \neq \overline{\lim}_{n\to\infty} nh_n^{d+2}$
$\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$	Large-sample distribution	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, \Omega_0)$	$\mathcal{N}(0, I_d/2)$	–
	Nonparametric bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, \Omega_0^*)$	$\mathcal{N}(0, 3I_d/4)$	–
	m-out-of-n bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, \Omega_0)$	$\mathcal{N}(0, I_d/2)$	–
$\hat{V}_{1,n}^{-1/2}(\hat{\theta}_n - \theta)$	Large-sample distribution	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$
	Nonparametric bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$
	m-out-of-n bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$
$\hat{V}_{2,n}^{-1/2}(\hat{\theta}_n - \theta)$	Large-sample distribution	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$
	Nonparametric bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, \Omega_2^*)$	$\mathcal{N}(0, 3I_d/2)$	–
	m-out-of-n bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$

Notes:
(i) Ω_0, Ω_0^*, and Ω_2^* are defined in Lemma 1(b), Theorem 1(b), and Theorem 3(a), respectively.
(ii) Lemmas 1 and 2 specify other assumptions and conditions imposed.

$\hat{F}_n^*(t) = B^{-1}\sum_{b=1}^{B} \mathbf{1}\big(\lambda'\hat{\theta}_{n,b}^*(h_n) \leq t\big)$. To simplify the exposition we assume throughout this section that $nh_n^{d+2} \to \kappa \in (0,\infty]$, but our discussion also applies to the case $\kappa = 0$ (albeit the scaling factor must be changed). Note that $\kappa = \infty$ corresponds to the conventional, asymptotically linear case.

The popular, easy-to-implement Efron-type $100\alpha\%$ confidence intervals are

$$\mathrm{CI} = \Big[\ q_{\alpha/2}^*,\ \ q_{1-\alpha/2}^*\ \Big], \quad q_\alpha^* = \inf\Big\{t \in \mathbb{R} : \hat{F}_n^*(t) \geq \alpha\Big\},$$

where B is chosen large enough, so that the bootstrap distribution is well approximated. Our theoretical results have important implications for this popular approach, showing, in particular, that asymptotic linearity is a fundamental feature for (at least) this semiparametric estimator. Specifically, whenever $nh_n^{d+2} \to \kappa \in (0,\infty]$,

$$\sqrt{n}(\hat{\theta}_n(h_n) - \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{L(z_i) + \frac{1}{\sqrt{\kappa}}\sqrt{\frac{h_n^{d+2}}{n-1}}\sum_{j=1, j\neq i}^{n} W(z_i, z_j; h_n)\right\} + o_p(1)$$
$$\rightsquigarrow \mathcal{N}\left(0, \Sigma + \frac{2}{\kappa}\Delta\right),$$

but, in contrast,

$$\sqrt{n}\left(\hat{\theta}_n^*(h_n)-\hat{\theta}_n(h_n)\right) \rightsquigarrow_p \mathcal{N}\left(0, \Sigma+\frac{6}{\kappa}\Delta\right).$$

Consequently, our results show that even the "vanilla" nonparametric Efron-type confidence intervals are valid *if and only if* the semiparametric estimator is asymptotically linear (i.e., $\kappa=\infty$). This result shows that the nonparametric bootstrap fails in a fundamental way, as this result holds separately from any results involving standard-error estimators.

The previous result shows that even the simplest of the bootstrap approaches fails in one of the simplest semiparametric inference contexts, in the sense that perturbations in the choice of h_n may lead to invalid confidence intervals. An alternative approach also many times employed in empirical work is to estimate the variance-covariance matrix using the bootstrap, as an alternative to employing an analytic standard-errors estimator. In cases where the analytic standard-errors are believed to be difficult to estimate (e.g., quantile regression), this approach may offer a useful empirical alternative. In our semiparametric context, this approach leads to the following $100\alpha\%$ confidence intervals:

$$\mathrm{CI}=\left[\lambda'\hat{\theta}_n-q_{1-\alpha/2}\sqrt{\lambda'\hat{V}_n^*\lambda},\quad \lambda'\hat{\theta}_n+q_{1-\alpha/2}\sqrt{\lambda'\hat{V}_n^*\lambda}\right],\quad q_\alpha=\Phi_1^{-1}(\alpha),$$

with

$$\hat{V}_n^*=\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_{n,b}^*-\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{n,b}^*\right)\left(\hat{\theta}_{n,b}^*-\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{n,b}^*\right)',$$

where here again B is chosen large enough, so that $\hat{V}_n^*$ approximates $\mathbb{V}^*[\hat{\theta}_n^*]$ well. Our results also show that this approach leads to biased confidence intervals, because

$$\hat{V}_n^*\approx\frac{1}{n}\left(\Sigma+\frac{6}{\kappa}\Delta\right)\neq\frac{1}{n}\left(\Sigma+\frac{2}{\kappa}\Delta\right)\approx\mathbb{V}[\hat{\theta}_n],$$

whenever $\kappa\neq\infty$, that is, when asymptotic linearity fails.

4.5. Further Simulation Evidence

To evaluate the small sample relevance of our theoretical results, we revisit the Monte Carlo experiment from Section 2. For brevity we focus on the "robustness" of the nonparametric bootstrap with respect to the choice of bandwidth. We employ exactly the same simulation setup as described above and compare the performance of the confidence intervals $\mathrm{CI}_k^*=\left[\lambda'\hat{\theta}_n-c_{k,97.5}^*\sqrt{\lambda'\hat{V}_{k,n}\lambda},\ \lambda'\hat{\theta}_n-c_{k,2.5}^*\sqrt{\lambda'\hat{V}_{k,n}\lambda}\right]$ across a range of bandwidths

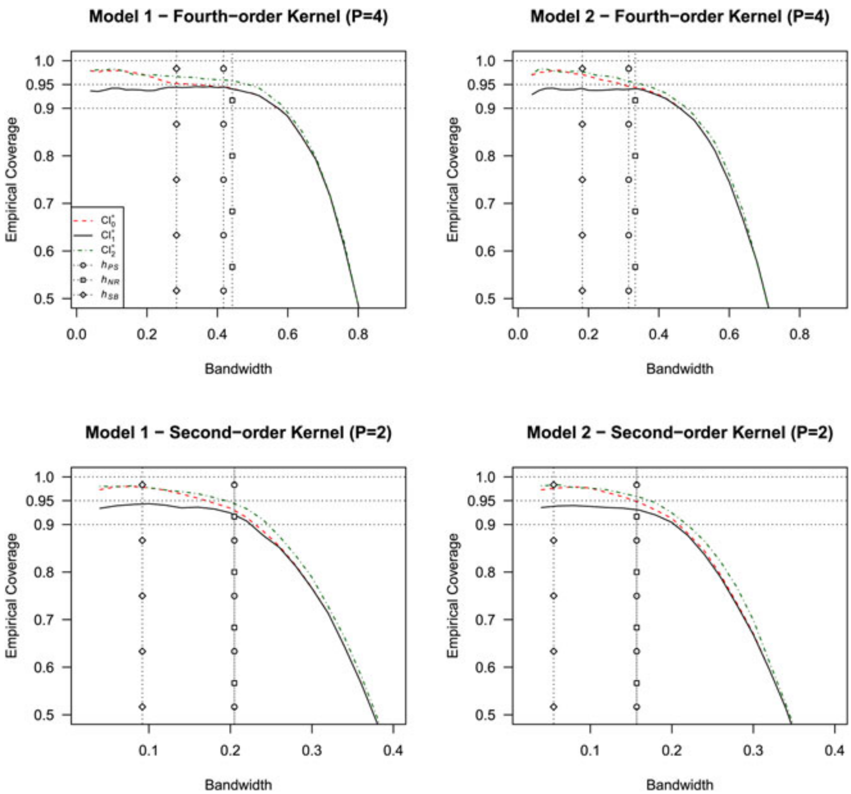

FIGURE 2. Empirical coverage of 95% bootstrap confidence intervals.

and for intervals constructed using estimated bandwidths (further discussed below), where $c^*_{k,\alpha}$ denotes the αth percentile of the distribution of $\lambda'(\hat{\theta}^*_n - n\hat{\theta}_n/(n-1))/\sqrt{\lambda'\hat{V}^*_{k,n}\lambda}$ for $k \in \{0,1,2\}$.

The main results from the simulation study are reported in Figure 2 and Table 2. As before, the figure includes the infeasible bandwidth choices h_{PS} and h_{NR}, but now we also include a third infeasible bandwidth choice, denoted h_{SB}, which is compatible with the small bandwidth asymptotic framework. These are the main "optimal" bandwidth choices available in the literature for $\hat{\theta}_n(h_n)$ and take the form

$$h_{PS} = \mathcal{C}_{PS}\, n^{-\frac{2}{2s+d+2}}, \quad h_{NR} = \mathcal{C}_{NR}\, n^{-\frac{2}{2s+d+2}}, \quad h_{SB} = \mathcal{C}_{SB}\, n^{-\frac{2}{2s+d}},$$

where $\mathcal{C}_{PS}$, $\mathcal{C}_{NR}$, and $\mathcal{C}_{SB}$ are fixed constants depending on the population parameter of interest and the underlying data generating process. The exact form of these constants, as well as a detailed discussion and comparison of these bandwidth selectors, is available in Cattaneo, Crump, and Jansson (2010). The

TABLE 2. Empirical coverage and interval length of 95% confidence intervals

	Bandwidth		Empirical coverage (bootstrap approx.)			Empirical coverage (Gaussian approx.)			Interval length (bootstrap approx.)			Interval length (Gaussian approx.)		
		h_n	CI_0^*	$\mathrm{CI}_{0,BC}^*$	CI_1^*	CI_0	$\mathrm{CI}_{0,BC}$	CI_1	CI_0^*	$\mathrm{CI}_{0,BC}^*$	CI_1^*	CI_0	$\mathrm{CI}_{0,BC}$	CI_1
Model 1														
$P=2$	h_{PS}	0.205	0.930	0.809	0.917	0.942	0.825	0.902	0.013	0.013	0.013	0.014	0.014	0.013
	h_{NR}	0.205	0.930	0.809	0.917	0.942	0.825	0.902	0.013	0.013	0.013	0.014	0.014	0.013
	h_{SB}	0.092	0.976	0.843	0.944	0.994	0.897	0.942	0.041	0.041	0.034	0.047	0.047	0.034
	$\hat{h}_{PS}$	0.209	0.874	0.764	0.855	0.890	0.786	0.845	0.014	0.014	0.014	0.016	0.016	0.014
	$\hat{h}_{NR}$	0.209	0.874	0.764	0.855	0.890	0.786	0.845	0.014	0.014	0.014	0.016	0.016	0.014
	$\hat{h}_{SB}$	0.088	0.973	0.837	0.942	0.987	0.890	0.948	0.071	0.071	0.058	0.083	0.083	0.059
Model 1														
$P=4$	h_{PS}	0.417	0.951	0.952	0.949	0.956	0.951	0.944	0.012	0.012	0.012	0.012	0.012	0.012
	h_{NR}	0.442	0.950	0.943	0.949	0.952	0.935	0.941	0.012	0.012	0.012	0.012	0.012	0.012
	h_{SB}	0.284	0.956	0.983	0.953	0.975	0.991	0.947	0.015	0.015	0.014	0.016	0.016	0.014
	$\hat{h}_{PS}$	0.242	0.961	0.987	0.917	0.975	0.993	0.926	0.019	0.019	0.017	0.021	0.021	0.017
	$\hat{h}_{NR}$	0.257	0.956	0.986	0.918	0.970	0.991	0.926	0.018	0.018	0.016	0.020	0.020	0.016
	$\hat{h}_{SB}$	0.148	0.983	0.999	0.946	0.996	1.000	0.954	0.045	0.045	0.037	0.052	0.052	0.038

Table continues on overleaf

TABLE 2. *continued*

	Bandwidth		Empirical coverage (bootstrap approx.)			Empirical coverage (Gaussian approx.)			Interval length (bootstrap approx.)			Interval length (Gaussian approx.)		
		h_n	CI_0^*	$\mathrm{CI}_{0,BC}^*$	CI_1^*	CI_0	$\mathrm{CI}_{0,BC}$	CI_1	CI_0^*	$\mathrm{CI}_{0,BC}^*$	CI_1^*	CI_0	$\mathrm{CI}_{0,BC}$	CI_1
Model 2														
$P=2$	h_{PS}	0.157	0.948	0.845	0.930	0.965	0.878	0.928	0.017	0.017	0.016	0.019	0.019	0.016
	h_{NR}	0.157	0.948	0.845	0.930	0.965	0.878	0.928	0.017	0.017	0.016	0.019	0.019	0.016
	h_{SB}	0.055	0.981	0.826	0.947	0.995	0.881	0.957	0.105	0.105	0.086	0.123	0.123	0.087
	$\hat{h}_{PS}$	0.172	0.902	0.796	0.875	0.921	0.822	0.875	0.018	0.018	0.017	0.020	0.020	0.017
	$\hat{h}_{NR}$	0.172	0.902	0.796	0.875	0.921	0.822	0.875	0.018	0.018	0.017	0.020	0.020	0.017
	$\hat{h}_{SB}$	0.075	0.983	0.839	0.955	0.992	0.887	0.961	0.093	0.093	0.077	0.109	0.109	0.078
Model 2														
$P=4$	h_{PS}	0.314	0.946	0.954	0.938	0.958	0.963	0.940	0.015	0.015	0.014	0.016	0.016	0.014
	h_{NR}	0.333	0.943	0.945	0.938	0.955	0.954	0.937	0.014	0.014	0.014	0.015	0.015	0.014
	h_{SB}	0.183	0.971	0.997	0.952	0.987	1.000	0.947	0.026	0.026	0.023	0.030	0.030	0.023
	$\hat{h}_{PS}$	0.197	0.965	0.986	0.920	0.979	0.991	0.928	0.025	0.025	0.022	0.029	0.029	0.023
	$\hat{h}_{NR}$	0.209	0.962	0.984	0.918	0.976	0.989	0.923	0.023	0.023	0.021	0.027	0.027	0.021
	$\hat{h}_{SB}$	0.124	0.976	0.996	0.945	0.992	0.998	0.951	0.061	0.061	0.051	0.071	0.071	0.051

Notes:
(i) "Empirical Coverage" columns report average coverage rate for each confidence interval.
(ii) "Interval Length" columns report average interval length for each confidence interval.
(iii) "Bandwidth" columns report either population "optimal" bandwidth or average of estimated bandwidths.

Monte Carlo experiment considers both the infeasible choices h_{PS}, h_{NR}, h_{SB}, as well as their feasible fully data-driven versions, which are denoted by $\hat{h}_{PS}$, $\hat{h}_{NR}$, $\hat{h}_{SB}$. The latter estimators for the bandwidth h_n are constructed as described in Cattaneo, Crump, and Jansson (2010), but we do not provide the details here to avoid unnecessary repetition. Table 2 reports results for confidence intervals constructed employing the infeasible bandwidths and their estimators, thus providing simulation evidence for fully data-driven inference procedures. This table also reports results for the Gaussian-based confidence intervals for completeness.

Figure 2 shows the following results. As predicted by Theorem 1, the interval CI_0^* is conservative for small bandwidths, having a coverage probability exceeding 0.95. In contrast, and in (almost perfect) agreement with Theorem 2, one of the new bootstrap-based confidence intervals introduced in this paper, CI_1^*, provides close-to-correct empirical coverage for a substantial range of small bandwidth choices. More precisely, in this simulation study, the confidence intervals CI_1^* exhibit slight undercoverage, which we conjecture is due to sampling ($n = 1,000$), bootstrap replication ($B = 2,000$), and simulation ($S = 3,000$) errors.[3] In terms of bandwidth selection, the Monte Carlo experiment shows that h_{SB} falls clearly inside the "robust" range of bandwidths in all cases. Interestingly, and because of the large "robust" range of bandwidths for CI_1^*, the bandwidth selectors h_{PS} and h_{NR} also appear to be "valid" when used to construct CI_1^*. Finally, as predicted by Theorem 3, the interval CI_2^* is also conservative for small bandwidths.

Remark. Unlike Condition AL, Condition AN can be satisfied (under Condition B) even when $s = 2$. Consistent with our theory, the bottom half of Figure 2 shows that the bootstrap-based interval CI_1^* is reasonably accurate also when a second-order kernel is used (i.e., when $P = 2$).

The results reported in Table 2 are in general consistent with the findings reported above, showing also how the estimation of the bandwidths translate into the performance of the different confidence intervals. The bootstrapped confidence intervals CI_1^* perform well across all designs considered, and on par with the confidence intervals based on the Gaussian approximation CI_1. The competing (classical) confidence intervals do not exhibit correct empirical coverage when $P = 2$ (Table 1), especially in the empirically important case when the bandwidth is estimated. When $P = 4$, however, the performance of these confidence intervals improves (as theoretically expected), especially when the bandwidth is estimated. Nonetheless, they never outperform the confidence intervals based on the bootstrap and Gaussian approximations in terms of empirical coverage. As for the (average) interval length of these intervals, we find that bootstrapping does not improve their performance in any case. Specifically, the nonparametric bootstrap leads to confidence intervals with essentially the same interval length as those constructed using the Gaussian approximation for all the inference procedures considered here.

5. CONCLUSION

Using an alternative asymptotic framework that removes the bandwidth conditions implying asymptotic linearity, we obtained new theory-based predictions about the finite-sample behavior of a variety of bootstrap-based inference procedures associated with the density-weighted averaged derivative estimator of PSS. In important respects, the predictions and methodological prescriptions emerging from the analysis presented here differ from those obtained by NR, who employed traditional bandwidth conditions and Edgeworth expansions.

The main qualitative findings obtained herein for the density-weighted average derivative estimator of PSS should extend to other kernel-based statistics that are asymptotically equivalent to n-varying second-order U-statistics when "small" bandwidths are also employed. Examples of statistics having the latter property include density-weighted averages (see Newey, Hsieh, and Robins, 2004, Section 2, and references therein), certain functionals of U-processes (see Aradillas-Lopéz, Honoré, and Powell, 2007 and references therein), and kernel-based specification test statistics (see Li and Racine, 2007, Ch. 12 and references therein). However, and perhaps surprisingly, in recent work (Cattaneo, Crump, and Jansson, 2013) we found that our results are not applicable to kernel-based (nondensity-)weighted average derivative estimators, as these estimators are not asymptotically equivalent to n-varying second-order U-statistics when smaller-than-usual bandwidths are employed.

NOTES

1. For reviews, see, e.g., Politis, Romano, and Wolf (1999) and Horowitz (2001).
2. The discussion below also applies immediately to the centered version of Efron-type confidence intervals, usually known as the "percentile method".
3. We did not explore further this point because of computational limitations.

REFERENCES

Aradillas-Lopéz, A., B.E. Honoré, & J.L. Powell (2007) Pairwise difference estimation with nonparametric control variables. *International Economic Review* 48, 1119–1158.

Bickel, P.J. & D.A. Freedman (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics* 9(6), 1196–1217.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2010) Robust data-driven inference for density-weighted average derivatives. *Journal of the American Statistical Association* 105(491), 1070–1083.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2014) Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory* 30(1), 176–200.

Cattaneo, M.D., R.K. Crump, & M. Jansson (2013) Generalized jackknife estimators of weighted average derivatives (with comments and rejoinder). *Journal of the American Statistical Association* 108(504), 1243–1268.

Chen, X., O. Linton, & I. van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71(5), 1591–1608.

Cheng, G. & J.Z. Huang (2010) Bootstrap consistency for general semiparametric M-estimation. *Annals of Statistics* 38(5), 2884–2915.

Efron, B. & C. Stein (1981) The jackknife estimate of variance. *Annals of Statistics* 9(3), 586–596.
Gine, E. & J. Zinn (1990) Bootstrapping general empirical measures. *Annals of Probability* 18(2), 851–869.
Hall, P. (1992) *The Bootstrap and Edgeworth Expansions*. Springer.
Härdle, W. & A. Tsybakov (1993) How sensitive are average derivatives? *Journal of Econometrics* 58, 31–48.
Heyde, C.C. & B.M. Brown (1970) On the departure from normality of a certain class of martingales. *Annals of Mathematical Statistics* 41(6), 2161–2165.
Horowitz, J. (2001) The bootstrap. In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. V, pp. 3159–3228. Elsevier Science B.V.
Li, Q. & S. Racine (2007) *Nonparametric Econometrics*. Princeton University Press.
Newey, W.K., F. Hsieh, & J.M. Robins (2004) Twicing kernels and a small bias property of semiparametric estimators. *Econometrica* 72, 947–962.
Nishiyama, Y. & P.M. Robinson (2000) Edgeworth expansions for semiparametric averaged derivatives. *Econometrica* 68(4), 931–979.
Nishiyama, Y. & P.M. Robinson (2001) Studentization in Edgeworth expansions for estimates of semiparametric index models. In C. Hsiao, K. Morimune, & J.L. Powell (eds.), *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*, pp. 197–240. Cambridge University Press.
Nishiyama, Y. & P.M. Robinson (2005) The bootstrap and the Edgeworth correction for semiparametric averaged derivatives. *Econometrica* 73(3), 197–240.
Politis, D. & J. Romano (1994) Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* 22(4), 2031–2050.
Politis, D., J. Romano, & M. Wolf (1999) *Subsampling*. Springer.
Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57(6), 1403–1430.
Powell, J.L. & T.M. Stoker (1996) Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics* 75(2), 291–316.
Robinson, P.M. (1995) The normal approximation for semiparametric averaged derivatives. *Econometrica* 63, 667–680.
Stoker, T.M. (1986) Consistent estimation of scaled coefficients. *Econometrica* 54(6), 1461–1481.
Xiong, S. & G. Li (2008) Some results on the convergence of conditional distributions. *Statistics and Probability Letters* 78(18), 3249–3253.

APPENDIX

Auxiliary Lemmas. For $\lambda \in \mathbb{R}^d$, let $\tilde{U}_{ij,n}(\lambda) = \lambda'\left[U(z_i, z_j; h_n) - \theta(h_n)\right]$ and define

$$T_{1,n}(\lambda) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \tilde{U}_{ij,n}(\lambda), \quad T_{2,n}(\lambda) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \tilde{U}_{ij,n}(\lambda)^2,$$

$$T_{3,n}(\lambda) = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} \frac{\tilde{U}_{ij,n}(\lambda)\tilde{U}_{ik,n}(\lambda) + \tilde{U}_{ij,n}(\lambda)\tilde{U}_{jk,n}(\lambda) + \tilde{U}_{ik,n}(\lambda)\tilde{U}_{jk,n}(\lambda)}{3},$$

$$T_{4,n}(\lambda) = \binom{n}{4}^{-1} \sum_{1 \le i < j < k < l \le n} \frac{\tilde{U}_{ij,n}(\lambda)\tilde{U}_{kl,n}(\lambda) + \tilde{U}_{ik,n}(\lambda)\tilde{U}_{jl,n}(\lambda) + \tilde{U}_{il,n}(\lambda)\tilde{U}_{jk,n}(\lambda)}{3},$$

as well as their bootstrap analogs

$$T_{1,n}^{*}(\lambda)=\binom{m}{2}^{-1}\sum_{1\le i<j\le m}\tilde{U}_{ij,n}^{*}(\lambda),\quad T_{2,n}^{*}(\lambda)=\binom{m}{2}^{-1}\sum_{1\le i<j\le m}\tilde{U}_{ij,n}^{*}(\lambda)^{2},$$

$$T_{3,n}^{*}(\lambda)=\binom{m}{3}^{-1}\sum_{1\le i<j<k\le m}\frac{\tilde{U}_{ij,n}^{*}(\lambda)\tilde{U}_{ik,n}^{*}(\lambda)+\tilde{U}_{ij,n}^{*}(\lambda)\tilde{U}_{jk,n}^{*}(\lambda)+\tilde{U}_{ik,n}^{*}(\lambda)\tilde{U}_{jk,n}^{*}(\lambda)}{3},$$

$$T_{4,n}^{*}(\lambda)=\binom{m}{4}^{-1}\sum_{1\le i<j<k<l\le m}\frac{\tilde{U}_{ij,n}^{*}(\lambda)\tilde{U}_{kl,n}^{*}(\lambda)+\tilde{U}_{ik,n}^{*}(\lambda)\tilde{U}_{jl,n}^{*}(\lambda)+\tilde{U}_{il,n}^{*}(\lambda)\tilde{U}_{jk,n}^{*}(\lambda)}{3},$$

where $\tilde{U}_{ij,n}^{*}(\lambda)=\lambda'\left[U(z_i^*,z_j^*;h_m)-\theta^*(h_m)\right]$.

The proof of Lemma 2 uses four technical lemmas, proofs of which are provided in the supplemental appendix. Lemma A.1 relates $\hat{\Sigma}_n$ and $\hat{\Delta}_n$ (and their bootstrap analogs) to $T_{1,n}$, $T_{2,n}$, $T_{3,n}$, and $T_{4,n}$ (and their bootstrap analogs), Lemma A.2 gives some asymptotic properties of $T_{1,n}$, $T_{2,n}$, $T_{3,n}$, and $T_{4,n}$ (and their bootstrap analogs), while Lemmas A.3 and A.4 are used to establish a pointwise version of (4) and to deduce (4) from its pointwise counterpart, respectively. Let $\eta_n=1/\min\left(1,nh_n^{d+2}\right)$.

LEMMA A.1. *If the assumptions of Lemma 2 hold and if $\lambda\in\mathbb{R}^d$, then*

(a) $\lambda'\hat{\Sigma}_n(h_n)\lambda=4[1+o(1)]n^{-1}T_{2,n}(\lambda)+4[1+o(1)]T_{3,n}(\lambda)-4T_{1,n}(\lambda)^2$,

(b) $h_n^{-(d+2)}\lambda'\hat{\Delta}_n(h_n)\lambda=[1+o(1)]T_{2,n}(\lambda)-T_{1,n}(\lambda)^2-2[1+o(1)]T_{3,n}(\lambda)+2[1+o(1)]T_{4,n}(\lambda)$,

(c) $\lambda'\hat{\Sigma}_n^*(h_m)\lambda=4[1+o(1)]m^{-1}T_{2,n}^*(\lambda)+4[1+o(1)]T_{3,n}^*(\lambda)-4T_{1,n}^*(\lambda)^2$,

(d) $h_m^{-(d+2)}\lambda'\hat{\Delta}_n^*(h_m)\lambda=[1+o(1)]T_{2,n}^*(\lambda)-T_{1,n}^*(\lambda)^2-2[1+o(1)]T_{3,n}^*(\lambda)+2[1+o(1)]T_{4,n}^*(\lambda)$.

LEMMA A.2. *If the assumptions of Lemma 2 hold and if $\lambda\in\mathbb{R}^d$, then*

(a) $T_{1,n}(\lambda)=o_p\left(\sqrt{\eta_n}\right)$,

(b) $T_{2,n}(\lambda)=\mathbb{E}\left[\tilde{U}_{ij,n}(\lambda)^2\right]+o_p\left(h_n^{-(d+2)}\right)$,

(c) $T_{3,n}(\lambda)=\mathbb{E}\left[(\mathbb{E}[\tilde{U}_{ij,n}(\lambda)|z_i])^2\right]+o_p(\eta_n)$,

(d) $T_{4,n}(\lambda)=o_p(\eta_n)$,

(e) $h_n^{d+2}\mathbb{E}\left[\tilde{U}_{ij,n}(\lambda)^2\right]\to\lambda'\Delta\lambda$ *and* $\mathbb{E}\left[(\mathbb{E}[\tilde{U}_{ij,n}(\lambda)|z_i])^2\right]\to\lambda'\Sigma\lambda/4$,

(f) $T_{1,n}^*(\lambda)=o_p\left(\sqrt{\eta_m}\right)$,

(g) $T_{2,n}^*(\lambda)=\mathbb{E}^*\left[\tilde{U}_{ij,n}^*(\lambda)^2\right]+o_p\left(h_m^{-(d+2)}\right)$,

(h) $T_{3,n}^*(\lambda)=\mathbb{E}^*\left[(\mathbb{E}^*[\tilde{U}_{ij,n}^*(\lambda)|z_i^*])^2\right]+o_p(\eta_m)$,

(i) $T_{4,n}^*(\lambda)=o_p(\eta_m)$,

(j) $h_m^{d+2}\mathbb{E}^*\left[\tilde{U}_{ij,n}^*(\lambda)^2\right]\to_p\lambda'\Delta\lambda$ *and*
$\mathbb{E}^*\left[(\mathbb{E}^*[\tilde{U}_{ij,n}^*(\lambda)|z_i^*])^2\right]-\lambda'\hat{\Sigma}_n(h_m)\lambda/4\to_p 0$.

LEMMA A.3. *If the assumptions of Lemma 2 hold and if* $\lambda \in \mathbb{R}^d$, *then*

(a) $\mathbb{E}\big[(\mathbb{E}^*[\tilde{U}^*_{ij,n}(\lambda)|z^*_i])^4\big] = O\big(\eta^2_m + h^2_m \eta^3_m\big)$,

(b) $\mathbb{E}\big[\tilde{U}^*_{ij,n}(\lambda)^4\big] = O\big(h_m^{-(3d+4)}\big)$,

(c) $\mathbb{E}\big[(\mathbb{E}^*[\tilde{U}^*_{ij,n}(\lambda)^2|z^*_i])^2\big] = O\big(m^{-1}h_m^{-(3d+4)} + h_m^{-(2d+4)}\big)$,

(d) $\mathbb{E}\big[(\mathbb{E}^*[\tilde{U}^*_{ij,n}(\lambda)\tilde{U}^*_{ik,n}(\lambda)|z^*_j, z^*_k])^2\big] = O\big(h_m^{-(d+4)} + m^{-2}h_m^{-(3d+4)}\big)$,

(e) $\mathbb{E}\big[(\mathbb{E}^*[\mathbb{E}^*[\tilde{U}^*_{ij,n}(\lambda)|z^*_i]\tilde{U}^*_{ij,n}(\lambda)|z^*_j])^2\big] = O\big(1 + m^{-1}h_m^{-(d+4)} + m^{-3}h_m^{-(3d+4)}\big)$.

LEMMA A.4. *There exist constants* C *and* J *(only depending on* d*) and a collection* $l_1, \ldots, l_J \in \Lambda_d$ *such that for every* $d \times d$ *matrix* M, $\sup_{\lambda \in \Lambda_d} \left(\lambda' M \lambda\right)^2 \le C \sum_{j=1}^J \left(l_j' M l_j\right)^2$.

Proof of Lemma 2. As noted in the discussion of Lemma 2,

$$\mathbb{V}^*\left[\hat{\theta}^*_n\right] = m^{-1}\mathbb{V}^*\left[L^*(z^*_i; h_m)\right] + \binom{m}{2}^{-1} \mathbb{V}^*\left[W^*(z^*_i, z^*_j; h_m)\right],$$

where, using Lemmas A.1 and A.2,

$$\mathbb{V}^*[L^*(z^*_i; h_m)] = \left(\frac{n-1}{n}\right)^2 \hat{\Sigma}_n(h_m) = \Sigma + 2\frac{m^2}{n}\binom{m}{2}^{-1} h_m^{-(d+2)}\Delta + o_p(\eta_m).$$

The proof of part (a) can be completed by using Lemmas A.1 and A.2 to show that

$$\begin{aligned}\lambda'\mathbb{V}^*[W^*(z^*_i, z^*_j; h_m)]\lambda &= h_m^{-(d+2)}\left(\frac{n-1}{n}\right)\left[\lambda'\hat{\Delta}_n(h_m)\lambda + o_p(1)\right] \\ &\quad - \frac{3}{2}\left(\frac{n-1}{n}\right)^2 \lambda'\hat{\Sigma}_n(h_m)\lambda \\ &= h_m^{-(d+2)}\lambda'\Delta\lambda + o_p(m\eta_m) \quad \left(\forall\lambda \in \mathbb{R}^d\right).\end{aligned}$$

Next, part (b) can be established by using Lemmas A.1 and A.2 to show that

$$\begin{aligned}\lambda'\hat{\Sigma}^*_n(h_m)\lambda &= 4[1+o(1)]m^{-1}T^*_{2,n}(\lambda) + 4[1+o(1)]T^*_{3,n}(\lambda) - 4T^*_{1,n}(\lambda)^2 \\ &= \lambda'\hat{\Sigma}_n(h_m)\lambda + 4m^{-1}h_m^{-(d+2)}\lambda'\Delta\lambda + o_p(\eta_m) \\ &= \lambda'\Sigma^*_n\lambda + o_p(\eta_m) \quad \left(\forall\lambda \in \mathbb{R}^d\right).\end{aligned}$$

Finally, to establish part (c), the theorem of Heyde and Brown (1970) is employed to prove the following condition, which is equivalent to (4) in view of part (a):

$$\sup_{\lambda \in \Lambda_d} \sup_{t \in \mathbb{R}^d} \left|\mathbb{P}^*\left[\lambda'(\hat{\theta}^*_n - \theta^*_n)/\sqrt{\lambda'\mathbb{V}^*[\hat{\theta}^*_n]\lambda} \le t\right] - \Phi_1(t)\right| \to_p 0.$$

For any $\lambda \in \Lambda_d$,

$$\lambda'\left(\hat{\theta}^*_n - \theta^*_n\right)/\sqrt{\lambda'\mathbb{V}^*[\hat{\theta}^*_n]\lambda} = \sum_{1 \le i \le m} Y^*_{i,m}(\lambda),$$

where, defining $L_{i,m}^*(\lambda) = \lambda' L^*\left(z_i^*; h_m\right)$ and $W_{ij,m}^*(\lambda) = \lambda' W^*\left(z_j^*, z_i^*; h_m\right)$,

$$Y_{i,m}^*(\lambda) = \left[m^{-1} L_{i,m}^*(\lambda) + \sum_{1 \le j < i} \binom{m}{2}^{-1} W_{ij,m}^*(\lambda) \right] / \sqrt{\lambda' \mathbb{V}^*[\hat{\theta}_n^*]\lambda}.$$

For any n, $\left(Y_{i,m}^*(\lambda), \mathcal{F}_{i,n}^*\right)$ is a martingale difference sequence, where $\mathcal{F}_{i,n}^* = \sigma\left(\mathcal{Z}_n, z_1^*, \ldots, z_i^*\right)$. Therefore, by the theorem of Heyde and Brown (1970), there exists a constant C such that

$$\begin{aligned}
&\sup_{\lambda \in \Lambda_d} \sup_{t \in \mathbb{R}^d} \left| \mathbb{P}^* \left[\lambda'(\hat{\theta}_n^* - \theta_n^*) / \sqrt{\lambda' \mathbb{V}^*[\hat{\theta}_n^*]\lambda} \le t \right] - \Phi_1(t) \right| \\
&\quad \le C \sup_{\lambda \in \Lambda_d} \left\{ \sum_{1 \le i \le m} \mathbb{E}^* \left[Y_{i,m}^*(\lambda)^4 \right] + \mathbb{E}^* \left[\left(\sum_{1 \le i \le m} \mathbb{E}\left[Y_{i,m}^*(\lambda)^2 \middle| \mathcal{F}_{i-1,n}^* \right] - 1 \right)^2 \right] \right\}^{1/5}.
\end{aligned}$$

Moreover, by Lemma A.4,

$$\sup_{\lambda \in \Lambda_d} \left\{ \sum_{1 \le i \le m} \mathbb{E}^* \left[Y_{i,m}^*(\lambda)^4 \right] + \mathbb{E}^* \left[\left(\sum_{1 \le i \le m} \mathbb{E}\left[Y_{i,m}^*(\lambda)^2 \middle| \mathcal{F}_{i-1,n}^* \right] - 1 \right)^2 \right] \right\} \to_p 0$$

if (and only if) (A.1)–(A.2) hold for every $\lambda \in \Lambda_d$, where

$$\sum_{1 \le i \le m} \mathbb{E}^* \left[Y_{i,m}^*(\lambda)^4 \right] \to_p 0, \tag{A.1}$$

$$\mathbb{E}^* \left[\left(\sum_{1 \le i \le m} \mathbb{E}\left[Y_{i,m}^*(\lambda)^2 \middle| \mathcal{F}_{i-1,n}^* \right] - 1 \right)^2 \right] \to_p 0. \tag{A.2}$$

The proof of part (c) will be completed by fixing $\lambda \in \Lambda_d$ and verifying $(A.1)$–$(A.2)$. First, using $\left(\lambda' V^*[\hat{\theta}_n^*]\lambda\right)^{-1} = O_p\left(m\eta_m^{-1}\right)$ and basic inequalities, it can be shown that $(A.1)$ holds if

$$R_{1,m} = m^{-2}\eta_m^{-2} \sum_{1 \le i \le m} \mathbb{E}\left[L_{i,m}^*(\lambda)^4 \right] \to 0,$$

$$R_{2,m} = m^{-6}\eta_m^{-2} \sum_{1 \le i \le m} \mathbb{E}\left[\left(\sum_{1 \le j < i} W_{ij,m}^*(\lambda) \right)^4 \right] \to 0.$$

Both conditions are satisfied because, using Lemma A.3,

$$\begin{aligned}
R_{1,m} &= O\left(m^{-1}\eta_m^{-2} \mathbb{E}[(\mathbb{E}^*[\tilde{U}_{ij,n}^*(\lambda) | z_i^*])^4] \right) = O\left(m^{-1} + m^{-1} h_m^2 \eta_m \right) \\
&= O\left(m^{-1} + m^{-2} h_m^{-d} \right) \to 0
\end{aligned}$$

and

$$R_{2,m} = O\left(m^{-4}\eta_m^{-2}\mathbb{E}[\tilde{U}^*_{ij,n}(\lambda)^4] + m^{-3}\eta_m^{-2}\mathbb{E}[(\mathbb{E}^*[\tilde{U}^*_{ij,n}(\lambda)^2|z_i^*])^2]\right)$$
$$= O\left(m^{-4}\eta_m^{-2}h_m^{-(3d+4)} + m^{-3}\eta_m^{-2}h_m^{-(2d+4)}\right) = O\left(m^{-2}h_m^{-d} + m^{-1}\right) \to 0.$$

Next, consider $(A.2)$. Because

$$(\lambda'\mathbb{V}^*[\hat{\theta}_n^*]\lambda)\left[\sum_{1\le i\le m}\mathbb{E}\left[Y^*_{i,m}(\lambda)^2\middle|\mathcal{F}^*_{i-1,n}\right]-1\right] = \binom{m}{2}^{-2}$$
$$\times\sum_{1\le i\le m}\left(\mathbb{E}\left[\left(\sum_{1\le j<i}W^*_{ij,m}(\lambda)\right)^2\middle|\mathcal{F}^*_{i-1,n}\right] - \sum_{1\le j<i}\mathbb{E}^*\left[W^*_{ij,m}(\lambda)^2\right]\right)$$
$$+2m^{-1}\binom{m}{2}^{-1}\sum_{1\le j<i\le m}\mathbb{E}\left[L^*_{i,m}(\lambda)W^*_{ij,m}(\lambda)|\mathcal{F}^*_{i-1,n}\right],$$

it suffices to show that

$$R_{3,m} = m^{-6}\eta_m^{-2}\mathbb{E}\left[\left(\sum_{1\le j<i\le m}\left\{\mathbb{E}\left[W^*_{ij,m}(\lambda)^2\middle|\mathcal{F}^*_{i-1,n}\right]\mathbb{E}^*\left[W^*_{ij,m}(\lambda)^2\right]\right\}\right)^2\right] \to 0,$$

$$R_{4,m} = m^{-6}\eta_m^{-2}\mathbb{E}\left[\left(\sum_{1\le k<j<i\le m}\mathbb{E}\left[W^*_{ij,m}(\lambda)W^*_{ik,m}(\lambda)\middle|\mathcal{F}^*_{i-1,n}\right]\right)^2\right] \to 0,$$

$$R_{5,m} = m^{-4}\eta_m^{-2}\mathbb{E}\left[\left(\sum_{1\le j<i\le m}\mathbb{E}\left[L^*_{i,m}(\lambda)W^*_{ij,m}(\lambda)|\mathcal{Z}_n, z_j^*\right]\right)^2\right] \to 0.$$

By simple calculations and Lemma A.3,

$$R_{3,m} = O\left(m^{-4}\eta_m^{-2}\mathbb{E}[W^*_{ij,m}(\lambda)^4]\right) = O\left(m^{-4}\eta_m^{-2}\mathbb{E}[\tilde{U}^*_{ij,n}(\lambda)^4]\right)$$
$$= O\left(m^{-4}\eta_m^{-2}h_m^{-(3d+4)}\right) = O\left(m^{-2}h_m^{-d}\right) \to 0,$$

$$R_{4,m} = O\left(m^{-2}\eta_m^{-2}\mathbb{E}\left[\left(\mathbb{E}^*[W^*_{ij,m}(\lambda)W^*_{ik,m}(\lambda)|z_j^*, z_k^*]\right)^2\right]\right)$$
$$= O\left(m^{-2}\eta_m^{-2}\mathbb{E}\left[\left(\mathbb{E}^*[\tilde{U}^*_{ij,n}(\lambda)\tilde{U}^*_{ik,n}(\lambda)|z_j^*, z_k^*]\right)^2\right]\right)$$
$$= O\left(m^{-2}\eta_m^{-2}h_m^{-(d+4)} + m^{-4}\eta_m^{-2}h_m^{-(3d+4)}\right) = O\left(h_m^d + m^{-2}h_m^{-d}\right) \to 0,$$

$$
\begin{aligned}
R_{5,m} &= O\left(m^{-1}\eta_m^{-2}\mathbb{E}\left[\left(\mathbb{E}^*\left[L^*_{i,m}(\lambda)W^*_{ij,m}(\lambda)|z^*_j\right]\right)^2\right]\right) \\
&= O\left(m^{-1}\eta_m^{-2}\mathbb{E}\left[\left(\mathbb{E}^*\left[\mathbb{E}^*[\tilde{U}^*_{ij,n}(\lambda)|z^*_i]\tilde{U}^*_{ij,n}(\lambda)|z^*_j\right]\right)^2\right]\right) \\
&= O\left(m^{-1}\eta_m^{-2} + m^{-2}\eta_m^{-2}h_m^{-(d+4)} + m^{-4}\eta_m^{-2}h_m^{-(3d+4)}\right) \\
&= O\left(m^{-1} + h_m^d + m^{-2}h_m^d\right) \to 0,
\end{aligned}
$$

as was to be shown. ■