Federal Reserve Bank of New York
Staff Reports

Bootstrapping Density-Weighted Average Derivatives

Matias D. Cattaneo
Richard K. Crump
Michael Jansson

Staff Report no. 452
May 2010

Bootstrapping Density-Weighted Average Derivatives

Matias D. Cattaneo, Richard K. Crump, and Michael Jansson

Federal Reserve Bank of New York Staff Reports, no. 452
May 2010
JEL classification: C12, C14, C21, C24

Abstract

Employing the "small-bandwidth" asymptotic framework of Cattaneo, Crump, and Jansson (2009), this paper studies the properties of several bootstrap-based inference procedures associated with a kernel-based estimator of density-weighted average derivatives proposed by Powell, Stock, and Stoker (1989). In many cases, the validity of bootstrap-based inference procedures is found to depend crucially on whether the bandwidth sequence satisfies a particular (asymptotic linearity) condition. An exception to this rule occurs for inference procedures involving a studentized estimator that employs a "robust" variance estimator derived from the "small-bandwidth" asymptotic framework. The results of a small-scale Monte Carlo experiment are found to be consistent with the theory and indicate in particular that sensitivity with respect to the bandwidth choice can be ameliorated by using the "robust" variance estimator.

Key words: averaged derivatives, bootstrap, small-bandwidth asymptotics

---

Cattaneo: University of Michigan (e-mail: cattaneo@umich.edu). Crump: Federal Reserve Bank of New York (e-mail: richard.crump@ny.frb.org). Jansson: University of California at Berkeley and Center for Research in Econometric Analysis of Time Series (e-mail: mjansson@econ.berkeley.edu). The authors are grateful for comments from Joel Horowitz, Lutz Kilian, Demian Pouzo, Rocio Titiunik, and seminar participants at Duke University, Harvard University, the University of Michigan, Northwestern University, and the University of Rochester. Cattaneo thanks the National Science Foundation for financial support (SES 0921505). Jansson gratefully acknowledges financial support from the National Science Foundation (SES 0920953) and research support from the Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation.

1. Introduction

Semiparametric estimators involving functionals of nonparametric estimators have been studied widely in econometrics. In particular, considerable effort has been devoted to characterizing conditions under which such estimators are asymptotically linear (see, e.g., Newey and McFadden (1994), Chen (2007), and the references therein). Moreover, although the asymptotic variance of an asymptotically linear semiparametric estimator can in principle be obtained by means of the pathwise derivative formula of Newey (1994a), it is desirable from a practical point of view to be able to base inference procedures on measures of dispersion that are "automatic" in the sense that they can be constructed without knowledge (or derivation) of the influence function (e.g., Newey (1994b)).

Perhaps the most natural candidates for such measures of dispersion are variances and/or percentiles obtained using the bootstrap.[1] Consistency of the nonparametric bootstrap has been established for a large class of semiparametric estimators by Chen, Linton, and van Keilegom (2003). Moreover, in the important special case of the density-weighted average derivative estimator of Powell, Stock, and Stoker (1989, henceforth PSS), a suitably implemented version of the nonparametric bootstrap was shown by Nishiyama and Robinson (2005, henceforth NR) to provide asymptotic refinements. The analysis in NR is conducted within the asymptotic framework of Nishiyama and Robinson (2000, 2001). Using the alternative asymptotic framework of Cattaneo, Crump, and Jansson (2009, henceforth CCJ), this paper revisits the large sample behavior of bootstrap-based inference procedures for density-weighted average derivatives and obtains (analytical and Monte Carlo) results that could be interpreted as a cautionary tale regarding the ease with which one might realize "the potential for bootstrap-based inference to (...) provide improvements in moderate-sized samples" (NR, p. 927).

Because the influence function of an asymptotically linear semiparametric estimator is invariant with respect to the nonparametric estimator upon which it is based (e.g., Newey (1994a, Proposition 1)), looking beyond the influence function is important if the sensitivity of the distributional properties of an estimator or test statistic with respect to user chosen objects such as kernels or bandwidths is a concern. This can be accomplished in various ways, the traditional approach being to work under assumptions that imply asymptotic linearity and then develop asymptotic expansions (of the Edgeworth or Nagar variety) intended to

[1] Another "automatic" measure of dispersion is the variance estimator of Newey (1994b). When applied to the density-weighted average derivative estimator studied in this paper, the variance estimator of Newey (1994b) coincides with Powell, Stock, and Stoker's (1989) variance estimator whose salient properties are characterized in Lemma 1 below.

elucidate the role of "higher-order" terms (e.g., Linton (1995)). Similarly to the Edgeworth expansions employed by Nishiyama and Robinson (2000, 2001, 2005), CCJ's asymptotic distribution theory for PSS's estimator (and its studentized version) is obtained by retaining terms that are asymptotically negligible when the estimator is asymptotically linear. Unlike the traditional approach, the "small bandwidth" approach taken by CCJ 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. Although similar in spirit to the Edgeworth expansion approach to improved asymptotic approximations, the small bandwidth approach of CCJ is conceptually distinct from the approach taken by Nishiyama and Robinson (2000, 2001, 2005) and it is therefore of interest to explore whether the small bandwidth approach gives rise to methodological prescriptions that differ from those obtained using the traditional approach.

The first main result, Theorem 1 below, studies the validity of bootstrap-based approximations to the distribution of PSS's estimator as well as its studentized version in the case where PSS's variance estimator is used for studentization purposes. It is shown that a necessary condition for bootstrap consistency is that the bandwidth vanishes slowly enough for asymptotic linearity to hold. Unlike NR, Theorem 1 therefore suggests that in samples of moderate size even the bootstrap approximations to the distributions of PSS's estimator and test statistic(s) may fail to adequately capture the extent to which these distributions are affected by the choice of the bandwidth, a prediction which is borne out in a small scale Monte Carlo experiment reported in Section 4.

The second main result, Theorem 2, establishes consistency of the bootstrap approximation to the distribution of PSS's estimator studentized by means of a variance estimator proposed by CCJ. As a consequence, Theorem 2 suggests that the fragility with respect to bandwidth choice uncovered by Theorem 1 is a property which should be attributed to PSS's variance estimator rather than the bootstrap distribution estimator. Another prediction of Theorem 2, namely that the bootstrap approximation to the distribution of an appropriately studentized estimator performs well across a wide range of bandwidths, is borne out in the Monte Carlo experiment of Section 4. Indeed, the range of bandwidths across which the bootstrap is found to perform well is wider than the range across which the standard normal approximation is found to perform well, indicating that there is an important sense in which bootstrap-based inference is capable of providing improvements in moderate-sized samples.

The variance estimator used for studentization purposes in Theorem 2 is one for which the studentized estimator is asymptotically standard normal across the entire range of bandwidth sequences considered in CCJ's approach. The final main result, Theorem 3, studies the bootstrap approximation to the distribution of PSS's estimator studentized by means of an alternative variance estimator also proposed by CCJ and finds, perhaps surprisingly, that although the associated studentized estimator is asymptotically standard normal across the entire range of bandwidth sequences considered in CCJ's approach, consistency of the bootstrap requires that the bandwidth vanishes slowly enough for asymptotic linearity to hold.

In addition to NR, whose relation to the present work was discussed in some detail above, the list of papers related to this paper includes Abadie and Imbens (2008) and Gonçalves and Vogelsang (2010). Abadie and Imbens (2008) study a nearest-neighbor matching estimator of a popular estimand in the program evaluation literature (the effect of treatment on the treated) and demonstrate by example that the nonparametric bootstrap variance estimator can be inconsistent in that case. Although the nature of the nonparametric estimator employed by Abadie and Imbens (2008) differs from the kernel estimator studied herein, their inconsistency result would appear to be similar to the equivalence between (i) and (ii) in Theorem 1(a) below. Comparing the results of this paper with those obtained by Abadie and Imbens (2008), one apparent attraction of kernel estimators (relative to nearest-neighbor estimators) is their tractability which allows to develop fairly detailed characterizations of the large-sample behavior of bootstrap procedures, including an array of (constructive) results on how to achieve bootstrap consistency even under departures from asymptotic linearity. Gonçalves and Vogelsang (2010) are concerned with autocorrelation robust inference in stationary regression models and establish consistency of the bootstrap under the fixed-b asymptotics of Kiefer and Vogelsang (2005). Although the fixed-b approach of Kiefer and Vogelsang (2005) is very similar in spirit to the "small bandwidth" approach of CCJ, the fact that some of the results of this paper are invalidity results about the bootstrap is indicative of an important difference between the nature of the functionals being studied in Kiefer and Vogelsang (2005) and CCJ, respectively.

The remainder of the paper is organized as follows. Section 2 introduces the model, presents the statistics under consideration, and summarizes some results available in the literature. Section 3 studies the bootstrap and obtains the main results of the paper. Section 4 summarizes the results of a simulation study. Section 5 concludes. The Appendix contains proofs of the theoretical results.

2. Model and Existing Results

Let $\mathcal{Z}_n = \{z_i = (y_i, x_i')' : i = 1, \ldots, n\}$ be a random sample of the random vector $z = (y, x')'$, where $y \in \mathbb{R}$ is a dependent variable and $x \in \mathbb{R}^d$ is a continuous explanatory variable with a density $f(\cdot)$. The density-weighted average derivative is given by

$$\theta = \mathbb{E}\left[f(x)\frac{\partial}{\partial x}g(x)\right], \qquad g(x) = \mathbb{E}[y|x].$$

It follows from (regularity conditions and) integration by parts that $\theta = -2\mathbb{E}[y\,\partial f(x)/\partial x]$. Noting this, PSS proposed the kernel-based estimator

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

where $\hat{f}_{n,i}(\cdot)$ is a "leave-one-out" estimator of $f(\cdot)$, with $K : \mathbb{R}^d \to \mathbb{R}$ a kernel function and h_n a positive (bandwidth) sequence.

To analyze inference procedures based on $\hat{\theta}_n$, some assumptions on the distribution of z and the properties of the user-chosen ingredients K and h_n are needed. Regarding the model and kernel function, the following assumptions will be made.

Assumption M. (a) $\mathbb{E}[y^4] < \infty$, $\mathbb{E}[\sigma^2(x) f(x)] > 0$ 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 $v(x) = \mathbb{E}[y^2|x]$.

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

Assumption K. (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 $P \geq 2$, $\int_{\mathbb{R}^d}|K(u)|(1 + \|u\|^P)\mathrm{d}u + \int_{\mathbb{R}^d}\|\dot{K}(u)\|(1 + \|u\|^2)\mathrm{d}u < \infty$, 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 following conditions on the bandwidth sequence h_n will play a crucial role in the sequel. (Here, and elsewhere in the paper, limits are taken as $n \to \infty$ unless otherwise noted.)

Condition B. (*Bias*) $\min\left(nh_n^{d+2}, 1\right) nh_n^{2\min(P,Q)} \to 0$.

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

Condition AN. (*Asymptotic Normality*) $n^2 h_n^d \to \infty$.

PSS studied the large sample properties of $\hat{\theta}_n$ and showed that if Assumptions M and K hold and if Conditions B and AL are satisfied, then $\hat{\theta}_n$ is asymptotically linear with (efficient) influence function $L(z) = 2[\partial e(x)/\partial x - y\, \partial f(x)/\partial x - \theta]$; 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. PSS's derivation of this result exploits the fact that the estimator $\hat{\theta}_n$ admits the (n-varying) U-statistic representation $\hat{\theta}_n = \hat{\theta}_n(h_n)$ with

$$\hat{\theta}_n(h) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} U(z_i, z_j; h), \qquad U(z_i, z_j; h) = -h^{-(d+1)} \dot{K}\left(\frac{x_i - x_j}{h}\right)(y_i - y_j),$$

which leads to the Hoeffding decomposition $\hat{\theta}_n - \theta = \mathcal{B}_n + \bar{L}_n + \bar{W}_n$, where

$$\mathcal{B}_n = \theta(h_n) - \theta, \quad \bar{L}_n = n^{-1} \sum_{i=1}^{n} L(z_i; h_n), \quad \bar{W}_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} W(z_i, z_j; h_n),$$

with

$$\theta(h) = \mathbb{E}[U(z_i, z_j; h)], \qquad L(z_i; h) = 2\left[\mathbb{E}[U(z_i, z_j; h)\,|z_i] - \theta(h)\right],$$
$$W(z_i, z_j; h) = U(z_i, z_j; h) - \frac{1}{2}\left(L(z_i; h) + L(z_j; h)\right) - \theta(h).$$

The purpose of Conditions B and AL is to ensure that the terms $\mathcal{B}_n$ and $\bar{W}_n$ in the Hoeffding decomposition are asymptotically negligible. Specifically, because $\mathcal{B}_n = O(h_n^{\min(P,Q)})$ under Assumptions M and K, Condition B ensures that the bias of $\hat{\theta}_n$ is asymptotically negligible. Condition AL, on the other hand, ensures that the "quadratic" term $\bar{W}_n$ in the Hoeffding decomposition is asymptotically negligible because $\sqrt{n}\bar{W}_n = O_p(1/\sqrt{nh_n^{d+2}})$ under Assumptions M and K. In other words, and as the notation suggests, Condition AL is crucial for asymptotic linearity of $\hat{\theta}_n$.

While asymptotic linearity is a desirable feature from the point of view of asymptotic efficiency, a potential concern about distributional approximations for $\hat{\theta}_n$ based on assumptions which imply asymptotic linearity is that such approximations ignore the variability in the "remainder" term $\bar{W}_n$. Thus, classical first-order, asymptotically linear, large sample theory may not accurately capture the finite sample behavior of $\hat{\theta}_n$ in general. It therefore seems desirable to employ inference procedures that are "robust" in the sense that they remain asymptotically valid at least under certain departures from asymptotic linearity.

In an attempt to construct such inference procedures, CCJ generalized (1) and showed that if Assumptions M and K hold and if Conditions B and AN are satisfied, then

$$V_n^{-1/2}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, I_d), \tag{2}$$

where

$$V_n = n^{-1}\Sigma + \binom{n}{2}^{-1} h_n^{-(d+2)}\Delta, \qquad \Delta = 2\mathbb{E}\left[\sigma^2(x) f(x)\right] \int_{\mathbb{R}^d} \dot{K}(u)\dot{K}(u)' \,\mathrm{d}u.$$

Similarly to the asymptotic linearity result of PSS, the derivation of (2) is based on the Hoeffding decomposition of $\hat{\theta}_n$. Instead of requiring asymptotic linearity of the estimator, this result provides an alternative first-order asymptotic theory under weaker assumptions, which simultaneously accounts for both the "linear" and "quadratic" terms in the expansion of $\hat{\theta}_n$. A key difference between (1) and (2) is the presence of the term $\binom{n}{2}^{-1} h_n^{-(d+2)}\Delta$ in V_n, which captures the variability of $\bar{W}_n$. In particular, result (2) 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.[2]

The result (2) suggests that asymptotic standard normality of studentized estimators might be achievable also when Condition AL is replaced by Condition AN. As an estimator of the variance of $\hat{\theta}_n$, PSS considered $\hat{V}_{0,n} = n^{-1}\hat{\Sigma}_n$, where $\hat{\Sigma}_n = \hat{\Sigma}_n(h_n)$,

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

CCJ showed that this estimator admits the stochastic expansion

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

[2] Condition AN permits failure not only of asymptotic linearity, but also of $\sqrt{n}$-consistency (when $nh_n^{d+2} \to 0$). Indeed, $\hat{\theta}_n$ can be inconsistent (when $\underline{\lim}_{n\to\infty} n^2 h_n^{d+2} < \infty$) under Condition AN.

implying in particular that it is consistent only when Condition AL is satisfied. Recognizing this lack of "robustness" of $\hat{V}_{0,n}$ with respect to h_n, CCJ proposed and studied the two alternative estimators

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

where

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

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

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 and K hold and suppose Conditions B and AN are 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* $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, \Omega_0)$, *where*

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

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

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 $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$ 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 $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$ 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 $I_d/2 \leq \Omega_0 \leq I_d$ (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) suggests 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.

Monte Carlo evidence consistent with these conjectures was presented by CCJ. Notably absent from consideration in Lemma 1 and the Monte Carlo work of CCJ are inference procedures based on resampling. In an important contribution, NR studied the behavior of the standard (nonparametric) bootstrap approximation to the distribution of PSS's test statistic and found that under bandwidth conditions slightly stronger than Condition AL bootstrap procedures are not merely valid, but actually capable of achieving asymptotic refinements. This finding leaves open the possibility that bootstrap validity, at least to first-order, might hold also under departures from Condition AL. The first main result presented here (Theorem 1 below) shows that, although the bootstrap approximation to the distribution of $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$ is more accurate than the standard normal approximation across the full range of bandwidth sequences allowed by Condition AN, Condition AL is necessary and sufficient for first-order validity of the standard nonparametric bootstrap approximation to the distribution of PSS's test statistic.

This equivalence can be viewed as a bootstrap analog of Lemma 1(a) and it therefore seems natural to ask whether bootstrap analogs of Lemma 1(c) are available for the inference procedures proposed by CCJ. Theorem 2 establishes a partial bootstrap analog of Lemma 1(c), namely validity of the nonparametric bootstrap approximation to the distribution of $\hat{V}_{1,n}^{-1/2}(\hat{\theta}_n - \theta)$ across the full range of bandwidth sequences allowed by Condition AN. That this result is not merely a consequence of the asymptotic pivotality result reported in Lemma 1(c) is demonstrated by Theorem 3, which shows that notwithstanding the asymptotic pivotality of $\hat{V}_{2,n}^{-1/2}(\hat{\theta}_n - \theta)$, the nonparametric bootstrap approximation to the distribution of the latter statistic is valid only when Condition AL holds.

3. The Bootstrap

3.1. Setup. This paper studies two variants of the m-out-of-n replacement bootstrap with $m = m(n) \to \infty$, namely the standard nonparametric bootstrap ($m(n) = n$) and (replacement) subsampling ($m(n)/n \to 0$).[3] To describe the bootstrap procedure(s), let $\mathcal{Z}_n^* = \{z_i^* = (y_i^*, x_i^{*\prime})' : i = 1, \ldots, m(n)\}$ be a random sample with replacement from the observed sample $\mathcal{Z}_n$. The bootstrap analogue of the estimator $\hat{\theta}_n$ is given by $\hat{\theta}^*_{m(n)} = \hat{\theta}^*_{m(n)}(h_{m(n)})$ with

$$\hat{\theta}^*_m(h) = \binom{m}{2}^{-1} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} U(z_i^*, z_j^*; h), \qquad U(z_i^*, z_j^*; h) = -h^{-(d+1)} \dot{K}\left(\frac{x_i^* - x_j^*}{h}\right)(y_i^* - y_j^*),$$

while the bootstrap analogues of the estimators $\hat{\Sigma}_n$ and $\hat{\Delta}_n$ are $\hat{\Sigma}^*_{m(n)} = \hat{\Sigma}^*_{m(n)}(h_{m(n)})$ and $\hat{\Delta}^*_{m(n)} = \hat{\Delta}^*_{m(n)}(h_{m(n)})$, respectively, where

$$\hat{\Sigma}^*_m(h) = \frac{1}{m} \sum_{i=1}^{m} \hat{L}^*_{m,i}(h) \hat{L}^*_{m,i}(h)', \qquad \hat{L}^*_{m,i}(h) = 2\left[\frac{1}{m-1} \sum_{j=1, j\neq i}^{m} U(z_i^*, z_j^*; h) - \hat{\theta}^*_m(h)\right],$$

and

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

$$\hat{W}^*_{m,ij}(h) = U(z_i^*, z_j^*; h) - \frac{1}{2}\left(\hat{L}^*_{m,i}(h) + \hat{L}^*_{m,j}(h)\right) - \hat{\theta}^*_m(h).$$

3.2. Preliminary Lemma. The main results of this paper follow from Lemma 1 and the following lemma, which will be used to characterize the large sample properties of bootstrap analogues of the test statistics $\hat{V}_{k,n}^{-1/2}(\hat{\theta}_n - \theta)$, $k \in \{0, 1, 2\}$. Let superscript $*$ on $\mathbb{P}$, $\mathbb{E}$, 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)).

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

(a) $V^{*-1}_{m(n)} \mathbb{V}^*[\hat{\theta}^*_{m(n)}] \to_p I_d$, *where*

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

[3] This paper employs the terminology introduced in Horowitz (2001). See also Politis, Romano, and Wolf (1999).

(b) $\Sigma_{m(n)}^{*-1}\hat{\Sigma}_{m(n)}^{*} \to_p I_d$ *and* $\Delta^{-1}\hat{\Delta}_{m(n)}^{*} \to_p I_d$*, where*

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

(c) $V_{m(n)}^{*-1/2}(\hat{\theta}_{m(n)}^{*} - \theta_{m(n)}^{*}) \rightsquigarrow_p \mathcal{N}(0, I_d)$.

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

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

with

$$\theta^*(h) = \mathbb{E}^*[U(z_i^*, z_j^*; h)], \qquad L^*(z_i^*; h) = 2[\mathbb{E}^*[U(z_i^*, z_j^*; h)|z_i^*] - \theta^*(h)],$$

$$W^*(z_i^*, z_j^*; h) = U(z_i^*, z_j^*; h) - \frac{1}{2}\left(L^*(z_i^*; h) + L^*(z_j^*; h)\right) - \theta^*(h).$$

Part (a) of Lemma 2 is obtained by noting that

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

where

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

and

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

The fact that $\mathbb{V}^*[W^*(z_i^*, z_j^*; h)] \approx h_m^{-(d+2)}\Delta$ implies that the bootstrap consistently estimates the variability of the "quadratic" term in the Hoeffding decomposition. On the other hand, the fact that $\mathbb{V}^*[\hat{\theta}_n^*] > n^{-1}\mathbb{V}^*[L^*(z_i^*; h_n)] \approx n^{-1}\hat{\Sigma}_n(h_n) = \hat{V}_{0,n}$ implies that the bootstrap variance estimator exhibits an upward bias even greater than that of $\hat{V}_{0,n}$, so the bootstrap variance estimator is inconsistent whenever PSS's estimator is. In their example of bootstrap failure for a nearest-neighbor matching estimator, Abadie and Imbens (2008) found that the (average) bootstrap variance can overestimate as well as underestimate the asymptotic variance of interest. No such ambiguity occurs here, as Lemma 2(a) shows that in the present case the bootstrap variance systematically exceeds the asymptotic variance (when Condition AL fails).

The proof of Lemma 2(b) shows that

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

implying that the asymptotic behavior of $\hat{\Sigma}_m^*$ differs from that of $\hat{\Sigma}_n(h_m)$ whenever Condition AL fails.

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, Theorem 3.5)), Lemma 2(c) is equivalent to the statement that

$$\sup_{t \in \mathbb{R}^d} \left| \mathbb{P}^* \left[V_{m(n)}^{*-1/2} (\hat{\theta}_{m(n)}^* - \theta_{m(n)}^*) \leq t \right] - \Phi_d(t) \right| \to_p 0. \tag{3}$$

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

$$\sup_{\lambda \in \Lambda_d} \sup_{t \in \mathbb{R}^d} \left| \mathbb{P}^* \left[\frac{\lambda'(\hat{\theta}_{m(n)}^* - \theta_{m(n)}^*)}{\sqrt{\lambda' V_{m(n)}^* \lambda}} \leq t \right] - \Phi_1(t) \right| \to_p 0, \tag{4}$$

where $\Lambda_d = \{\lambda \in \mathbb{R}^d : \lambda'\lambda = 1\}$ denotes the unit sphere in $\mathbb{R}^d$.[4] The proof of Lemma 2(c) uses the theorem of Heyde and Brown (1970) to verify (4).

3.3. Bootstrapping PSS. Theorem 1 below is concerned with the ability of the bootstrap to approximate the distributional properties of PSS's test statistic. To anticipate the main findings, notice that Lemma 1 gives

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

[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

$$\sup_{t \in \mathbb{R}^d} \left| \mathbb{P}^* \left[\frac{\lambda'(\hat{\theta}_{m(n)}^* - \theta_{m(n)}^*)}{\sqrt{\lambda' V_{m(n)}^* \lambda}} \leq t \right] - \Phi_1(t) \right| \to_p 0, \qquad \forall \lambda \in \Lambda_d,$$

implies weak convergence in probability of $V_{m(n)}^{*-1/2} (\hat{\theta}_{m(n)}^* - \theta_{m(n)}^*)$.

while, in contrast, in the case of the standard nonparametric bootstrap (when $m(n) = n$) Lemma 2 gives

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

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

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

suggesting that consistency of subsampling 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 is an immediate consequence of Lemmas 1–2 and the continuous mapping theorem for weak convergence in probability (e.g., Xiong and Li (2008, Theorem 3.1)), makes the preceding heuristics precise.

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

(a) *The following are equivalent:*

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

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

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

(c) *If $m(n) \to \infty$ and $m(n)/n \to 0$ and if nh_n^{d+2} is convergent in $\bar{\mathbb{R}}_+$, then*

$$\hat{V}_{0,m(n)}^{*-1/2}(\hat{\theta}_{m(n)}^* - \theta_{m(n)}^*) \rightsquigarrow_p \mathcal{N}(0, \Omega_0).$$

In an obvious way, Theorem 1(a)-(b) can be viewed as a standard nonparametric bootstrap analogue of Lemma 1(a)-(b). In particular, Theorem 1(a) shows that Condition AL is necessary and sufficient for consistency of the bootstrap. This result shows 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. The implication (i) $\Rightarrow$ (iv) in Theorem 1(a) is essentially due to NR.[5] On the other hand, the result that Condition AL is necessary for bootstrap consistency would appear to be new. In Section 4, the finite sample relevance of this sensitivity with respect to bandwidth choice suggested by Theorem 1(a) will be explored in a Monte Carlo experiment.

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 $3I_d/4 \leq \Omega_0^* \leq I_d$ 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 $\hat{V}_{0,n}^{-1/2}(\hat{\theta}_n - \theta)$ is seen to be superior to the standard normal approximation because $\Omega_0 \leq \Omega_0^* \leq I_d$. As a consequence, there is a sense in which the bootstrap offers "refinements" even when Condition AL fails.

Theorem 1(c) shows that a sufficient condition for consistency of subsampling 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 $m(n)$ is even for every n, it follows from Theorem 1(c) that

$$\hat{V}_{0,m(n)}^{*-1/2}(\hat{\theta}^*_{m(n)} - \theta^*_{m(n)}) \rightsquigarrow_p \mathcal{N}(0, I_d),$$

whereas, by Lemma 1(b),

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

This example is intentionally extreme, but the qualitative message that consistency of subsampling 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

[5] The results of NR are obtained under slightly stronger assumptions than those of Lemma 1 and require $nh_n^{d+3}/(\log n)^9 \to \infty$.

a sequence $m(n)$ such that ($m(n) \to \infty$, $m(n)/n \to 0$, and)

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

In other words, employing critical values obtained by means of subsampling does not automatically "robustify" an inference procedure based on PSS's statistic.

3.4. Bootstrapping CCJ. Because both $\hat{V}^{-1/2}_{1,n}(\hat{\theta}_n - \theta)$ and $\hat{V}^{-1/2}_{2,n}(\hat{\theta}_n - \theta)$ are 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

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

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, $m(n) \to \infty$, and if $\overline{\lim}_{n\to\infty} m(n)/n < \infty$, then $\hat{V}^{*-1/2}_{1,m(n)}(\hat{\theta}^*_{m(n)} - \theta^*_{m(n)}) \rightsquigarrow_p \mathcal{N}(0, I_d)$.*

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. The theory presented here does not allow to determine whether the bootstrap approximation enjoys any advantages over the standard normal approximation, but Monte Carlo evidence reported in Section 4 suggests that bootstrap-based inference does have attractive small sample properties.

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

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

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

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

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

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

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 $m(n) \to \infty$ and $m(n)/n \to 0$, then $\hat{V}_{2,m(n)}^{*-1/2}(\hat{\theta}_{m(n)}^* - \theta_{m(n)}^*) \rightsquigarrow_p \mathcal{N}(0, I_d)$.

Theorem 3 and the arguments on which it is based is of interest for at least two reasons. First, while there is no shortage of examples of bootstrap failure in the literature, it seems surprising that the bootstrap fails to approximate the distribution of the asymptotically pivotal statistic $\hat{V}_{2,n}^{-1/2}(\hat{\theta}_n - \theta)$ whenever Condition AL is violated.[6] Second, 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 1–2 that

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

it can be shown that if the assumptions of Lemma 1 hold, then

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

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 subsampling might fail.

3.5. Summary of Results. The main results of this paper are summarized in Table 1. This table describes the limiting distributions of the test statistics proposed by PSS and CCJ, as well as the limiting distributions (in probability) of their bootstrap analogues. (CCJ$_k$

[6] The severity of the bootstrap failure is characterized in Theorem 3(a) and measured by the variance matrix Ω_2^*, which satisfies $I_d \le \Omega_2^* \le 3I_d/2$, 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.

with $k \in \{1,2\}$ refers to the test statistics in Lemma 1(c).) Each panel corresponds to one test statistic, and includes 3 rows corresponding to each approximation used (large sample distribution, standard bootstrap, and replacement subsampling, respectively). Each column analyzes a subset of possible bandwidth sequences, which leads to different approximations in general.

As shown in the table, the "robust" studentized test statistic using $\hat{V}_{1,n}$, denoted CCJ$_1$, is the only statistic that remains valid in all cases. For the studentized test statistic of PSS (first panel), both the standard bootstrap and replacement subsampling are invalid in general, while for the "robust" studentized test statistic using $\hat{V}_{2,n}$, denoted CCJ$_2$, only replacement subsampling is valid. As discussed above, the extent of the failure of the bootstrap and the "direction" of its "bias" are described in the extreme case of $\kappa = 0$. Table 1 also highlights that when nh_n^{d+2} is not convergent in $\bar{\mathbb{R}}_+$, weak convergence (in probability) of any asymptotically non-pivotal test statistic (under the bootstrap distribution) is not guaranteed in general.

4. Simulations

In an attempt to explore whether the theory-based predictions presented above are borne out in samples of moderate size, this section reports the main results from a Monte Carlo experiment. 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 vector of regressors $x_i \in \mathbb{R}^d$, and $\mathbf{1}(\cdot)$ representing the indicator function. The dimension of the covariates is set to $d = 2$ and both components of β are set equal to unity. The vector of regressors is generated using independent random variables with the second component set to $x_{2i} \sim \mathcal{N}(0,1)$. Two data generating processes are considered for the first component of x_i: Model 1 imposes $x_{1i} \sim \mathcal{N}(0,1)$ and Model 2 imposes $x_{1i} \sim (\chi_4 - 4)/\sqrt{8}$, with χ_p a chi-squared random variable with p degrees of freedom. For simplicity only results for the first component of $\theta = (\theta_1, \theta_2)'$ are reported. The population parameters of interest are $\theta_1 = 1/8\pi$ and $\theta_1 \approx 0.03906$ for Model 1 and Model 2, respectively. Note that Model 1 corresponds to the one analyzed in Nishiyama and Robinson (2000, 2005), while both models were also considered in CCJ and Cattaneo, Crump, and Jansson (2010).

The number of simulations is set to $S = 3,000$, the sample size for each simulation is set to $n = 1,000$, and the number of bootstrap replications for each simulation is set to $B = 2,000$. (See Andrews and Buchinsky (2000) for a discussion of the latter choice.) The Monte Carlo experiment is very computationally demanding: each design, with a grid of 15 bandwidths, requires approximately 6 days to complete, when using a `C` code (with wrapper in `R`) parallelized in 150 CPUs (2.33 Ghz). The computer code is available upon request.

The simulation study presents evidence on the performance of the standard nonparametric bootstrap across an appropriate grid of possible bandwidth choices. Three test statistics are considered for the bootstrap procedure:

$$\mathrm{PSS}^* = \frac{\lambda'(\hat{\theta}_n^* - \theta_n^*)}{\sqrt{\lambda'\hat{V}_{0,n}^*\lambda}}, \qquad \mathrm{NR}^* = \frac{\lambda'(\hat{\theta}_n^* - \theta_n^* - \hat{\mathcal{B}}_n)}{\sqrt{\lambda'\hat{V}_{0,n}^*\lambda}}, \qquad \mathrm{CCJ}^* = \frac{\lambda'(\hat{\theta}_n^* - \theta_n^*)}{\sqrt{\lambda'\hat{V}_{1,n}^*\lambda}},$$

with $\lambda = (1, 0)'$, and where $\hat{\mathcal{B}}_n$ is a bias-correction estimate. The first test statistic (PSS*) corresponds to the bootstrap analogue of the classical, asymptotically linear, test statistic proposed by PSS. The second test statistic (NR*) corresponds to the bias-corrected statistic proposed by NR. The third test statistic (CCJ*) corresponds to the bootstrap analogue of the robust, asymptotically normal, test statistic proposed by CCJ. For implementation, a standard Gaussian product kernel is used for $P = 2$, and a Gaussian density-based multiplicative kernel is used for $P = 4$. The bias-correction estimate $\hat{\mathcal{B}}_n$ is constructed using a plug-in estimator for the population bias with an initial bandwidth choice of $b_n = 1.2h_n$, as discussed in Nishiyama and Robinson (2000, 2005)..

The results are summarized in Figure 1 ($P = 2$) and Figure 2 ($P = 4$). These figures plot the empirical coverage for the three competing 95% confidence intervals as a function of the choice of bandwidth. To facilitate the analysis two additional horizontal lines at 0.90 and at the nominal coverage rate 0.95 are included for reference. In each figure, the first and second rows correspond to Models 1 and 2, respectively. Also, for each figure, the first column depicts the results for the competing confidence intervals using the standard nonparametric bootstrap to approximate the quantiles of interest, while the second column does the same but using the large sample distribution quantiles (e.g., $\Phi_1^{-1}(0.975) \approx 1.96$). Finally, each plot also includes three population bandwidth selectors available in the literature for density-weighted average derivatives as vertical lines. Specifically, h_{PS}, h_{NR} and h_{CCJ} denote the population "optimal" bandwidth choices described in Powell and Stoker (1996), NR and Cattaneo, Crump, and Jansson (2010), respectively. The bandwidths differ in general, although $h_{PS} = h_{NR}$ when $d = 2$ and $P = 2$. (For a detailed discussion and comparison of these bandwidth selectors, see Cattaneo, Crump, and Jansson (2010).)

The main results are consistent across all designs considered. First, it is seen that bootstrapping PSS induces a "bias" in the distributional approximation for small bandwidths, as predicted in Theorem 1. Second, bootstrapping CCJ (which uses $\hat{V}_{1,n}$) provides a close-to-correct approximation for a range of small bandwidth choices, as predicted by Theorem 2. Third, by comparing these results across columns (bootstrapping vs. Gaussian approximations), it is seen that the "bias" in the distributional approximation of PSS for small bandwidths is smaller (leading to shorter confidence intervals) than the corresponding "bias" introduced from using the Gaussian approximation (longer confidence intervals), as predicted by Theorem 1.

In addition, it is found that the range of bandwidths with close-to-correct coverage has been enlarged for both PSS and CCJ when using the bootstrap approximation instead of the Gaussian approximation. The bias correction proposed by Nishiyama and Robinson (2000, 2005) does not seem to work well when $P = 2$ (Figure 1), but works somewhat better when $P = 4$ (Figure 2).[7]

Based on the theoretical results developed in this paper, and the simulation evidence presented, it appears that confidence intervals based on the bootstrap distribution of CCJ perform the best, as they are valid under quite weak conditions. In terms of bandwidth selection, the Monte Carlo experiment shows that h_{CCJ} falls clearly inside the "robust" range of bandwidths in all cases. Interestingly, and because bootstrapping CCJ seems to enlarge the "robust" range of bandwidths, the bandwidth selectors h_{PS} and h_{NR} also appear to be "valid" when coupled with the bootstrapped confidence intervals based on CCJ*.

5. CONCLUSION

Employing the "small bandwidth" asymptotic framework of CCJ, this paper has developed theory-based predictions of finite sample behavior of a variety of bootstrap-based inference procedures associated with the kernel-based density-weighted averaged derivative estimator proposed by PSS. In important respects, the predictions and methodological prescriptions emerging from the analysis presented here differ from those obtained using Edgeworth expansions by NR. The results of a small-scale Monte Carlo experiment were found to be consistent with the theory developed here, indicating in particular that while the properties of inference procedures employing the variance estimator of PSS are very sensitive to bandwidth choice, this sensitivity can be ameliorated by using a "robust" variance estimator proposed in CCJ.

[7] It seems plausible that these conclusions are sensitive to the choice of initial bandwidth b_n for the construction of the estimator $\hat{\mathcal{B}}_n$, but we have made no attempt to improve on the initial bandwidth choice advocated by Nishiyama and Robinson (2000, 2005).

6. Appendix

For any $\lambda \in \mathbb{R}^d$, let $\tilde{U}_{ij,n}(\lambda) = \lambda'[U(z_i, z_j; h_n) - \theta(h_n)]$ and define the n-varying U-statistics

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

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

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

as well as their bootstrap analogues

$$T^*_{1,m}(\lambda) = \binom{m}{2}^{-1} \sum_{1\leq i<j\leq m} \tilde{U}^*_{ij,m}(\lambda), \qquad T^*_{2,m}(\lambda) = \binom{m}{2}^{-1} \sum_{1\leq i<j\leq m} \tilde{U}^*_{ij,m}(\lambda)^2,$$

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

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

where $\tilde{U}^*_{ij,m}(\lambda) = \lambda'[U(z^*_i, z^*_j; h_m) - \theta^*(h_m)]$. (Here, and elsewhere in the Appendix, the dependence of $m(n)$ on n has been suppressed.)

The proof of Lemma 2 uses four technical lemmas, proofs of which are available upon request. The first lemma is a simple algebraic result relating $\hat{\Sigma}_n$ and $\hat{\Delta}_n$ (and their bootstrap analogues) to $T_{1,n}$, $T_{2,n}$, $T_{3,n}$, and $T_{4,n}$ (and their bootstrap analogues).

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) $\lambda'\hat{\Sigma}^*_m(h_m)\lambda = 4[1+o(1)]\,m^{-1}T^*_{2,m}(\lambda) + 4[1+o(1)]T^*_{3,m}(\lambda) - 4T^*_{1,m}(\lambda)^2$,
(d) $h_m^{-(d+2)}\lambda'\hat{\Delta}^*_m(h_m)\lambda = [1+o(1)]T^*_{2,m}(\lambda) - T^*_{1,m}(\lambda)^2 - 2[1+o(1)]T^*_{3,m}(\lambda) + 2[1+o(1)]T^*_{4,m}(\lambda)$.

The next lemma, which follows by standard properties of (n-varying) U-statistics (e.g., NR and CCJ), gives some asymptotic properties of $T_{1,n}$, $T_{2,n}$, $T_{3,n}$, and $T_{4,n}$ (and their bootstrap analogues). Let $\eta_n = 1/\min\left(1, nh_n^{d+2}\right)$.

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

(a) $T_{1,n}(\lambda) = o_p(\sqrt{\eta_n})$,
(b) $T_{2,n}(\lambda) = \mathbb{E}[\tilde{U}_{ij,n}(\lambda)^2] + o_p(h_n^{-(d+2)})$,
(c) $T_{3,n}(\lambda) = \mathbb{E}[(\mathbb{E}[\tilde{U}_{ij,n}(\lambda)|z_i])^2] + o_p(\eta_n)$,
(d) $T_{4,n}(\lambda) = o_p\left(\eta_n\right)$,
(e) $h_n^{d+2}\mathbb{E}[\tilde{U}_{ij,n}(\lambda)^2] \to \lambda'\Delta\lambda$ and $\mathbb{E}[(\mathbb{E}[\tilde{U}_{ij,n}(\lambda)|z_i])^2] \to \lambda'\Sigma\lambda/4$,
(f) $T^*_{1,m}(\lambda) = o_p(\sqrt{\eta_m})$,
(g) $T^*_{2,m}(\lambda) = \mathbb{E}^*[\tilde{U}^*_{ij,m}(\lambda)^2] + o_p(h_m^{-(d+2)})$,
(h) $T^*_{3,m}(\lambda) = \mathbb{E}^*[(\mathbb{E}[\tilde{U}^*_{ij,m}(\lambda)|\mathcal{Z}_n, z^*_i])^2] + o_p(\eta_m)$,
(i) $T^*_{4,m}(\lambda) = o_p(\eta_m)$,
(j) $h_m^{d+2}\mathbb{E}^*[\tilde{U}^*_{ij,m}(\lambda)^2] \to_p \lambda'\Delta\lambda$ and $\mathbb{E}^*[(\mathbb{E}[\tilde{U}^*_{ij,m}(\lambda)\,|\mathcal{Z}_n, z^*_i])^2] - \lambda'\hat{\Sigma}_n(h_m)\lambda/4 \to_p 0$.

The next lemma, which can be established by expanding sums and using simple bounding arguments, is used to establish a pointwise version of (4).

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

(a) $\mathbb{E}[(\mathbb{E}[\tilde{U}^*_{ij,m}(\lambda)|\mathcal{Z}_n, z^*_i])^4] = O(\eta_m^2 + h_m^2\eta_m^3)$,
(b) $\mathbb{E}[\tilde{U}^*_{ij,m}(\lambda)^4] = O(h_m^{-(3d+4)})$,
(c) $\mathbb{E}[(\mathbb{E}[\tilde{U}^*_{ij,m}(\lambda)^2|\mathcal{Z}_n, z^*_i])^2] = O(m^{-1}h_m^{-(3d+4)} + h_m^{-(2d+4)})$,
(d) $\mathbb{E}[(\mathbb{E}[\tilde{U}^*_{ij,m}(\lambda)\tilde{U}^*_{ik,m}(\lambda)|\mathcal{Z}_n, z^*_j, z^*_k])^2] = O(h_m^{-(d+4)} + m^{-2}h_m^{-(3d+4)})$,
(e) $\mathbb{E}[(\mathbb{E}[\mathbb{E}[\tilde{U}^*_{ij,m}(\lambda)|\mathcal{Z}_n, z^*_i]\tilde{U}^*_{ij,m}(\lambda)|\mathcal{Z}_n, z^*_j])^2] = O(1 + m^{-1}h_m^{-(d+4)} + m^{-3}h_m^{-(3d+4)})$.

Finally, the following lemma about quadratic forms is used to deduce (4) from its pointwise counterpart.

Lemma A-4. There exist constants C and J (only dependent on d) and a collection $l_1, \ldots, l_J \in \Lambda_d$ such that, for every $d \times d$ matrix M,

$$\sup_{\lambda \in \Lambda_d} \left(\lambda' M \lambda\right)^2 \leq C \sum_{j=1}^{J} \left(l_j' M l_j\right)^2.$$

Proof of Lemma 2. By the properties of the (conditional on $\mathcal{Z}_n$) Hoeffding decomposition, $\mathbb{E}[L^*(z_i^*;h)|\mathcal{Z}_n] = 0$ and $\mathbb{E}[W^*(z_i^*, z_j^*;h)|\mathcal{Z}_n, z_i^*] = 0$, so

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

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

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

Also, for any $\lambda \in \mathbb{R}^d$, it can be shown that

$$\lambda'\mathbb{V}^*[W^*(z_i^*, z_j^*;h_m)]\lambda = h_m^{-(d+2)}\left(\frac{n-1}{n}\right)\left[\lambda'\hat{\Delta}_n(h_m)\lambda + o_p\left(1\right)\right] - \frac{3}{2}\left(\frac{n-1}{n}\right)^2 \lambda'\hat{\Sigma}_n(h_m)\lambda.$$

Therefore, using Lemmas A-1 and A-2,

$$\mathbb{V}^*[W^*(z_i^*, z_j^*;h_m)] = h_m^{-(d+2)}\Delta + o_p\left(m\eta_m\right),$$

completing the proof of part (a).

Next, using Lemmas A-1 and A-2,

$$\begin{aligned}
\lambda'\hat{\Sigma}_m^*(h_m)\lambda &= 4[1+o\left(1\right)]m^{-1}T_{2,m}^*(\lambda) + 4[1+o\left(1\right)]T_{3,m}^*(\lambda) - 4T_{1,m}^*(\lambda)^2 \\
&= \lambda'\hat{\Sigma}_n(h_m)\lambda + 4m^{-1}h_m^{-(d+2)}\lambda'\Delta\lambda + o_p(\eta_m) \\
&= \lambda'\Sigma_m^*\lambda + o_p\left(\eta_m\right),
\end{aligned}$$

establishing part (b).

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):

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

For any $\lambda \in \Lambda_d$,

$$\frac{\lambda'\hat{\theta}_m^* - \lambda'\theta_m^*}{\sqrt{\lambda'\mathbb{V}^*[\hat{\theta}_m^*]\lambda}} = \sum_{i=1}^m Y_{i,m}^*\left(\lambda\right),$$

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

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

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

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

Moreover, by Lemma A-4,

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

if (and only if) the following hold for every $\lambda \in \Lambda_d$:

$$\sum_{i=1}^{m} \mathbb{E}^{*}\left[Y_{i,m}^{*}(\lambda)^4 \right] \to_p 0 \tag{5}$$

and

$$\mathbb{E}^{*}\left[\left(\sum_{i=1}^{m} \mathbb{E}\left[Y_{i,m}^{*}(\lambda)^2 \middle| \mathcal{F}_{i-1,n}^{*} \right] - 1 \right)^2 \right] \to_p 0. \tag{6}$$

The proof of part (c) will be completed by fixing $\lambda \in \Lambda_d$ and verifying (5)–(6). First, using $(\lambda' \mathbb{V}^{*}[\hat{\theta}_m^{*}]\lambda)^{-1} = O_p(m\eta_m^{-1})$ and basic inequalities, it can be shown that (5) holds if

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

and

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

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

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

and

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

Next, consider (6). Because

$$\begin{aligned}&(\lambda'\mathbb{V}^{*}[\hat{\theta}_m^{*}]\lambda)\left[\sum_{i=1}^{m}\mathbb{E}\left[\left.Y_{i,m}^{*}\left(\lambda\right)^2\right|\mathcal{F}_{i-1,n}^{*}\right]-1\right]\\ &=\binom{m}{2}^{-2}\sum_{i=1}^{m}\left(\mathbb{E}\left[\left.\left(\sum_{j=1}^{i-1}W_{ij,m}^{*}\left(\lambda\right)\right)^2\right|\mathcal{F}_{i-1,n}^{*}\right]-\sum_{j=1}^{i-1}\mathbb{E}^{*}\left[W_{ij,m}^{*}\left(\lambda\right)^2\right]\right)\\ &\qquad+2m^{-1}\binom{m}{2}^{-1}\sum_{i=1}^{m}\sum_{j=1}^{i-1}\mathbb{E}\left[L_{i,m}^{*}\left(\lambda\right)W_{ij,m}^{*}\left(\lambda\right)|\mathcal{F}_{i-1,n}^{*}\right],\end{aligned}$$

it suffices to show that

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

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

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

By simple calculations and Lemma A-3,

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

$$\begin{aligned}R_{4,m} &= O\left(m^{-2}\eta_m^{-2}\mathbb{E}\left[\left(\mathbb{E}[W_{ij,m}^{*}(\lambda)\,W_{ik,m}^{*}(\lambda)\,|\mathcal{Z}_n, z_j^*, z_k^*]\right)^2\right]\right) \\ &= O\left(m^{-2}\eta_m^{-2}\mathbb{E}\left[\left(\mathbb{E}[\tilde{U}_{ij,m}^{*}(\lambda)\,\tilde{U}_{ik,m}^{*}(\lambda)\,|\mathcal{Z}_n, z_j^*, z_k^*]\right)^2\right]\right) \\ &= O\left(m^{-2}\eta_m^{-2}h_m^{-(d+4)} + m^{-4}\eta_m^{-2}h_m^{-(3d+4)}\right) = O\left(h_m^{d} + m^{-2}h_m^{-d}\right) \to 0,\end{aligned}$$

$$\begin{aligned}R_{5,m} &= O\left(m^{-1}\eta_m^{-2}\mathbb{E}\left[\left(\mathbb{E}\left[L_{i,m}^{*}(\lambda)\,W_{ij,m}^{*}(\lambda)\,|\mathcal{Z}_n, z_j^*\right]\right)^2\right]\right) \\ &= O\left(m^{-1}\eta_m^{-2}\mathbb{E}\left[\left(\mathbb{E}\left[\mathbb{E}[\tilde{U}_{ij,m}^{*}(\lambda)\,|\mathcal{Z}_n, z_i^*]\tilde{U}_{ij,m}^{*}(\lambda)\,|\mathcal{Z}_n, z_j^*\right]\right)^2\right]\right) \\ &= O\left(m^{-1}\eta_m^{-2} + m^{-2}\eta_m^{-2}h_m^{-(d+4)} + m^{-4}\eta_m^{-2}h_m^{-(3d+4)}\right) = O\left(m^{-1} + h_m^{d} + m^{-2}h_m^{d}\right) \to 0,\end{aligned}$$

as was to be shown. ■

REFERENCES

ABADIE, A., AND G. W. IMBENS (2008): "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, 76(6), 1537–1557.

ANDREWS, D. W. K., AND M. BUCHINSKY (2000): "A Three-Step Method for Choosing the Number of Bootstrap Repetitions," *Econometrica*, 68(1), 23–51.

CATTANEO, M. D., R. K. CRUMP, AND M. JANSSON (2009): "Small Bandwidth Asymptotics for Density-Weighted Average Derivatives," working paper.

——— (2010): "Robust Data-Driven Inference for Density-Weighted Average Derivatives," forthcoming in *Journal of the American Statistical Association.*

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics, Volume VI*, ed. by J. J. Heckman, and E. Leamer, pp. 5549–5632. Elsevier Science B.V., New York.

CHEN, X., O. LINTON, AND VAN KEILEGOM (2003): "Estimation of Semiparametric Models when The Criterion Function Is Not Smooth," *Econometrica*, 71(5), 1591–1608.

GINE, E., AND J. ZINN (1990): "Bootstrapping General Empirical Measures," *Annals of Probability*, 18(2), 851–869.

GONÇALVES, S., AND T. J. VOGELSANG (2010): "Block Bootstrap HAC Robust Tests: The Sophistication of the Naive Bootstrap," forthcoming in *Econometric Theory.*

HEYDE, C. C., AND 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 *Handbook of Econometrics, Volume V*, ed. by J. Heckman, and E. Leamer, pp. 3159–3228. Elsevier Science B.V., New York.

KIEFER, N. M., AND T. J. VOGELSANG (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21(6), 1130–1164.

LINTON, O. (1995): "Second Order Approximation in the Partialy Linear Regression Model," *Econometrica*, 63(5), 1079–1112.

NEWEY, W. K. (1994a): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62(6), 1349–1382.

——— (1994b): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10(2), 1349–1382.

Newey, W. K., and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Volume IV*, ed. by R. F. Engle, and D. L. McFadden, pp. 2111–2245. Elsevier Science B.V., New York.

Nishiyama, Y., and P. M. Robinson (2000): "Edgeworth Expansions for Semiparametric Averaged Derivatives," *Econometrica*, 68(4), 931–979.

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

——— (2005): "The Bootstrap and the Edgeworth Correction for Semiparametric Averaged Derivatives," *Econometrica*, 73(3), 197–240.

Politis, D., J. Romano, and M. Wolf (1999): *Subsampling*. Springer, New York.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

Powell, J. L., and T. M. Stoker (1996): "Optimal Bandwidth Choice for Density-Weighted Averages," *Journal of Econometrics*, 75(2), 291–316.

Xiong, S., and G. Li (2008): "Some Results on the Convergence of Conditional Distributions," *Statistics and Probability Letters*, 78(18), 3249–3253.

Table 1: Summary of Main Results

		$\lim_{n\to\infty} nh_n^{d+2} = \kappa \in \bar{\mathbb{R}}_+$			$\underline{\lim}_{n\to\infty} nh_n^{d+2} \neq \overline{\lim}_{n\to\infty} nh_n^{d+2}$
		$\kappa = \infty$	$\kappa \in (0,\infty)$	$\kappa = 0$	
PSS	Large sample distribution	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, \Omega_0)$	$\mathcal{N}(0, I_d/2)$	–
	Standard bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, \Omega_0^*)$	$\mathcal{N}(0, 3I_d/4)$	–
	Replacement subsampling	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, \Omega_0)$	$\mathcal{N}(0, I_d/2)$	–
CCJ_1	Large sample distribution	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$
	Standard bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$
	Replacement subsampling	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$
CCJ_2	Large sample distribution	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$
	Standard bootstrap	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, \Omega_2^*)$	$\mathcal{N}(0, 3I_d/2)$	–
	Replacement subsampling	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$	$\mathcal{N}(0, I_d)$

Notes:
(i) PSS, CCJ_1 and CCJ_2 denote the studentized test statistics using $\hat{V}_{0,n}$, $\hat{V}_{1,n}$ and $\hat{V}_{2,n}$, respectively.
(ii) Ω_0, Ω_0^*, Ω_2^* are defined in Lemma 1(b), Theorem 1(b) and Theorem 3(a), respectively.
(iii) Lemmas 1–2 specify other assumptions and conditions imposed.

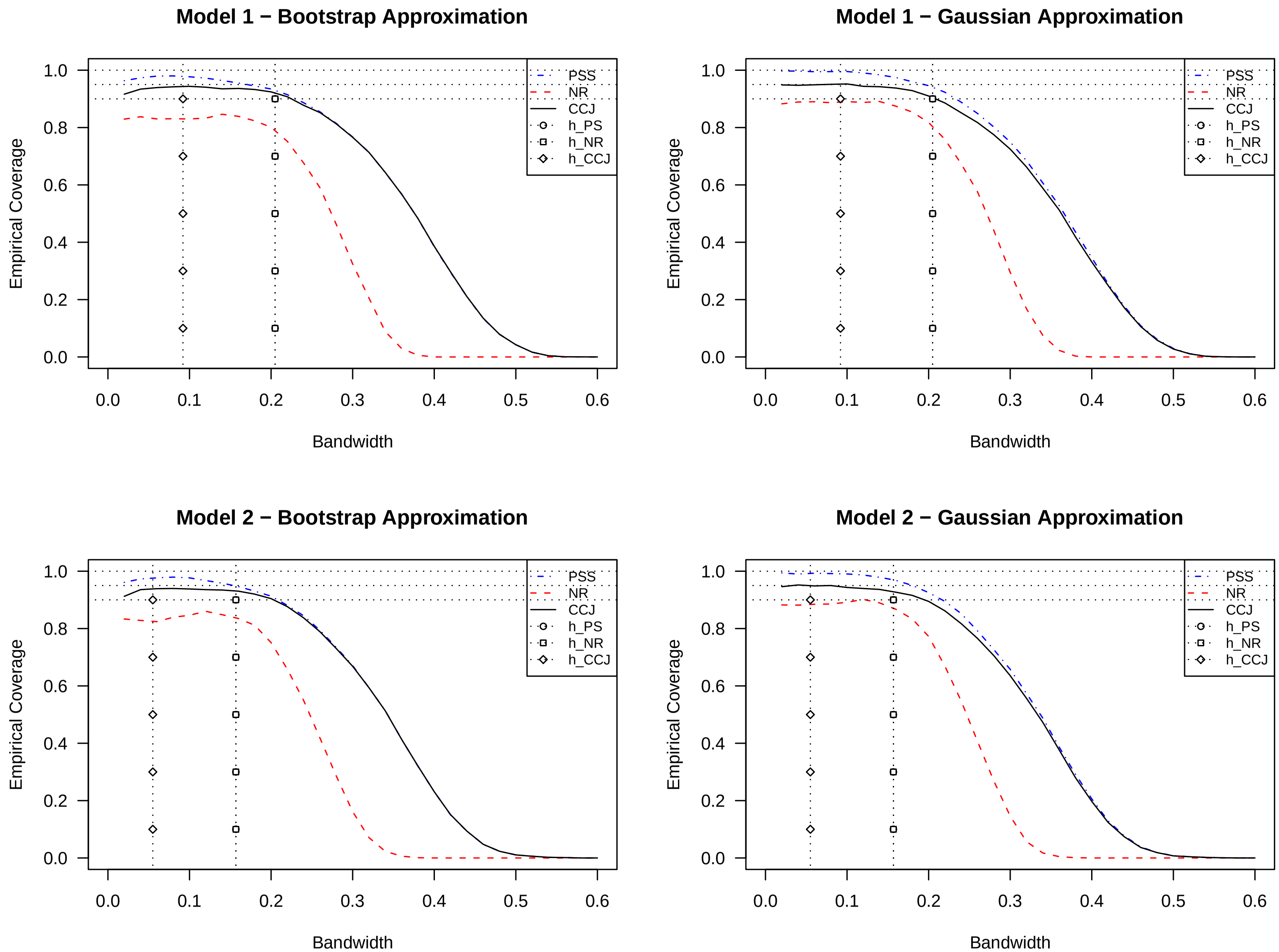

Figure 1: Empirical Coverage Rates for 95% Confidence Intervals: $d = 2$, $P = 2$, $n = 1,000$, $B = 2,000$

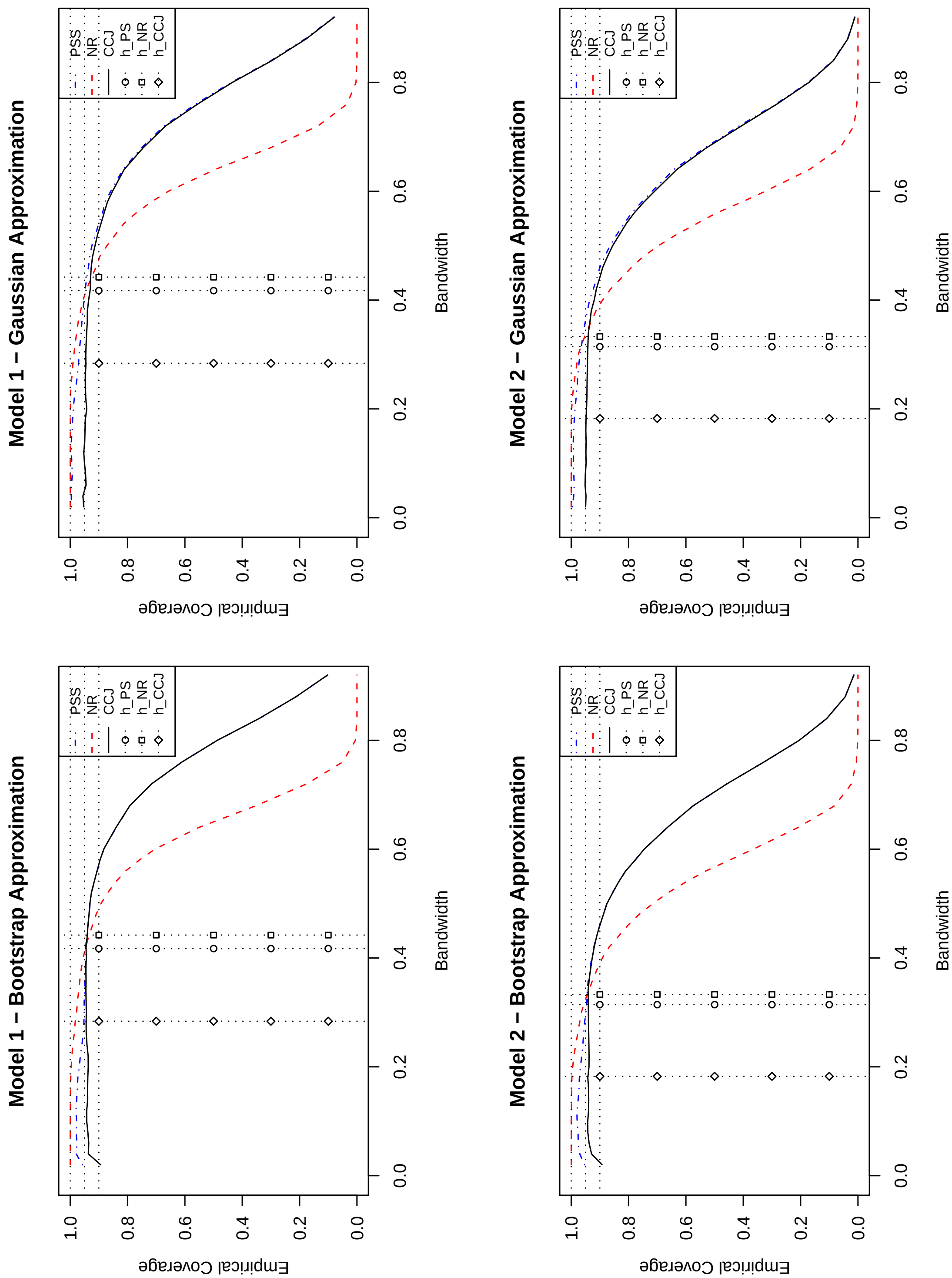

Figure 2: Empirical Coverage Rates for 95% Confidence Intervals: $d = 2$, $P = 4$, $n = 1,000$, $B = 2,000$