Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Bootstrap Inference in the Presence of Bias

Giuseppe Cavaliere, Sílvia Gonçalves, Morten Ørregaard Nielsen & Edoardo Zanelli

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2024, VOL. 00, NO. 0, 1–11: Theory and Methods
https://doi.org/10.1080/01621459.2023.2284980

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Check for updates

Bootstrap Inference in the Presence of Bias

Giuseppe Cavaliere[a,b], Sílvia Gonçalves[c], Morten Ørregaard Nielsen[d], and Edoardo Zanelli[a]

[a]University of Bologna, Bologna, Italy; [b]Exeter Business School, Exeter, UK; [c]McGill University, Montreal, Canada; [d]Aarhus University, Aarhus, Denmark

ABSTRACT
We consider bootstrap inference for estimators which are (asymptotically) biased. We show that, even when the bias term cannot be consistently estimated, valid inference can be obtained by proper implementations of the bootstrap. Specifically, we show that the prepivoting approach of Beran, originally proposed to deliver higher-order refinements, restores bootstrap validity by transforming the original bootstrap *p*-value into an asymptotically uniform random variable. We propose two different implementations of prepivoting (plug-in and double bootstrap), and provide general high-level conditions that imply validity of bootstrap inference. To illustrate the practical relevance and implementation of our results, we discuss five examples: (i) inference on a target parameter based on model averaging; (ii) ridge-type regularized estimators; (iii) nonparametric regression; (iv) a location model for infinite variance data; and (v) dynamic panel data models. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received August 2022
Accepted November 2023

KEYWORDS
Asymptotic bias; Bootstrap; Incidental parameter bias; Model averaging; Nonparametric regression; Prepivoting

1. Introduction

Suppose that θ is a scalar parameter of interest and let $\hat{\theta}_n$ denote an estimator for which

$$T_n := g(n)(\hat{\theta}_n - \theta) \xrightarrow{d} B + \xi_1, \tag{1.1}$$

where $g(n) \to \infty$ is the rate of convergence of $\hat{\theta}_n$, ξ_1 is a continuous random variable centered at zero, and B is an asymptotic bias (our theory in fact allows for a more general formulation of the bias). A typical example is $g(n) = n^{1/2}$ and $\xi_1 \sim N(0, \sigma^2)$. Unless B can be consistently estimated, which is often difficult or impossible, classic (first-order) asymptotic inference on θ based on quantiles of ξ_1 in (1.1) is not feasible. Furthermore, the bootstrap, which is well known to deliver asymptotic refinements over first-order asymptotic approximations as well as bias corrections (Hall 1992; Horowitz 2001; Cattaneo and Jansson 2018, 2022; Cattaneo, Jansson, and Ma 2019), cannot in general be applied to solve the asymptotic bias problem when a consistent estimator of B does not exist. Examples are given below.

Our goal is to justify bootstrap inference based on T_n in the context of asymptotically biased estimators and where a consistent estimator of B does not exist. Consider the bootstrap statistic $T_n^* := g(n)(\hat{\theta}_n^* - \hat{\theta}_n)$, where $\hat{\theta}_n^*$ is a bootstrap version of $\hat{\theta}_n$, such that

$$T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1, \tag{1.2}$$

where $\hat{B}_n$ is the implicit bootstrap bias, and "$\xrightarrow{d^*}_p$" denotes weak convergence in probability (defined below). When $\hat{B}_n - B = o_p(1)$, the bootstrap is asymptotically valid in the usual sense that the bootstrap distribution of T_n^* is consistent for the asymptotic distribution of T_n, that is, $\sup_{x\in\mathbb{R}} |P^*(T_n^* \le x) - P(T_n \le x)| = o_p(1)$.

We consider situations where $\hat{B}_n - B$ is not asymptotically negligible so the bootstrap fails to replicate the asymptotic bias. For example, this happens when the asymptotic bias term in the bootstrap world includes a random (additive) component, that is

$$\hat{B}_n - B \xrightarrow{d} \xi_2 \text{ (jointly with (1.1))}, \tag{1.3}$$

where ξ_2 is a random variable centered at zero. In this case, the bootstrap distribution is random in the limit and hence cannot mimic the asymptotic distribution given in (1.1). Moreover, the distribution of the bootstrap p-value, $\hat{p}_n := P^*(T_n^* \le T_n)$, is not asymptotically uniform, and the bootstrap cannot in general deliver hypothesis tests (or confidence intervals) with the desired null rejection probability (or coverage probability).

In this article, we show that in this nonstandard case valid inference can successfully be restored by proper implementation of the bootstrap. This is done by focusing on properties of the bootstrap p-value rather than on the bootstrap as a means of estimating limiting distributions, which is infeasible due to the asymptotic bias. In particular, we show that such implementations lead to bootstrap inferences that are valid in the sense that they provide asymptotically uniformly distributed p-values.

Our inference strategy is based on the fact that, for some bootstrap schemes, the large-sample distribution of the bootstrap p-value, say $H(u)$, $u \in [0, 1]$, although not uniform, does not depend on B. That is, we can search for bootstrap algorithms which generate bootstrap p-values that, in large samples, are not affected by unknown bias terms. When this is possible,

CONTACT Giuseppe Cavaliere giuseppe.cavaliere@unibo.it Dipartimento di Scienze Economiche, University of Bologna, Piazza Scaravilli 2, Bologna, Italy.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

we can make use of the prepivoting approach of Beran (1987, 1988), which—as we will show in this article—allows to restore bootstrap validity. Specifically, our proposed modified p-value is defined as

$$\tilde{p}_n := \hat{H}_n(\hat{p}_n),$$

where $\hat{H}_n(u)$ is any consistent estimator of $H(u)$, uniformly over $u \in [0, 1]$. The (asymptotic) probability integral transform $\hat{p}_n \mapsto H(\hat{p}_n)$, continuity of $H(u)$, and consistency of $\hat{H}_n(u)$ then guarantee that $\tilde{p}_n$ is asymptotically uniformly distributed. Interestingly, Beran (1987, 1988) proposed this approach to obtain asymptotic refinements for the bootstrap, but did not consider asymptotically biased estimators as we do here.

We propose two approaches to estimating H. First, if $H = H_\gamma$, where γ is a finite-dimensional parameter vector, and a consistent estimator $\hat{\gamma}_n$ of γ is available, then a "plug-in" approach setting $\hat{H}_n = H_{\hat{\gamma}_n}$ can deliver asymptotically uniform p-values. Second, if estimation of γ is difficult (e.g., when γ does not have a closed form expression), we can use a "double bootstrap" scheme (Efron 1983; Hall 1986), where estimation of H is achieved by resampling from the bootstrap data originated in the first level.

For both methods, we provide general high-level conditions that imply validity of the proposed approach. Our conditions are not specific to a given bootstrap method; rather, they can in principle be applied to any bootstrap scheme satisfying the proposed sufficient conditions for asymptotic validity.

Our approach is related to recent work by Shao and Politis (2013) and Cavaliere and Georgiev (2020). In particular, a common feature is that the distribution function of the bootstrap statistic, conditional on the original data, is random in the limit. Cavaliere and Georgiev (2020) emphasize that randomness of the limiting bootstrap measure does not prevent the bootstrap from delivering an asymptotically uniform p-value (bootstrap "unconditional" validity), and provide results to assess such asymptotic uniformity. Our context is different, since the presence of an asymptotic bias term renders the distribution of the bootstrap p-value nonuniform, even asymptotically. In this respect, our work is related to Shao and Politis (2013), who show that t-statistics based on subsampling or block bootstrap methods with bandwidth proportional to sample size may deliver non-uniformly distributed p-values that, however, can be estimated.

To illustrate the practical relevance of our results and to show how to implement them in applied problems, we consider three examples involving estimators that feature an asymptotic bias term. In the first two examples (model averaging and ridge regression), B is not consistently estimable due to the presence of local-to-zero parameters and the standard bootstrap fails. In the third example (nonparametric regression), the bootstrap fails because B depends on the second-order derivative of the conditional mean function, whose estimation requires the use of a different (suboptimal) bandwidth. In these examples, ξ_1 is normal, but $g(n)$ and B are example-specific. Two additional examples are presented in the supplement. The fourth example is a simple location model without the assumption of finite variance, where ξ_1 is not normal and estimators converge at an unknown rate. The fifth example considers inference for dynamic panel data models, where B is the incidental parameter bias.

The remainder of the article is organized as follows. In Section 2 we introduce our three leading examples. Section 3 contains our general results, which we apply to the three examples in Section 4. Section 5 concludes. The supplemental material contains two appendices. Appendix A specializes the general theory to the case of asymptotically Gaussian statistics, and Appendix B contains details and proofs for the three leading examples, as well as two additional examples.

Notation

Throughout this article, the notation $\sim$ indicates equality in distribution. For instance, $Z \sim N(0, 1)$ means that Z is distributed as a standard normal random variable. We write "$x := y$" and "$y =: x$" to mean that x is defined by y. The standard Gaussian cumulative distribution function (cdf) is denoted by Φ; $U_{[0,1]}$ is the uniform distribution on $[0, 1]$, and $\mathbb{I}_{\{\cdot\}}$ is the indicator function. If F is a cdf, F^{-1} denotes the generalized inverse, that is, the quantile function, $F^{-1}(u) := \inf\{v \in \mathbb{R} : F(v) \geq u\}$, $u \in \mathbb{R}$. Unless specified otherwise, all limits are for $n \to \infty$. For matrices a, b, c with n rows, we let $S_{ab} := a'b/n$ and $S_{ab.c} := S_{ab} - S_{ac}S_{cc}^{-1}S_{cb}$, assuming that S_{cc} has full rank.

For a (single level or first-level) bootstrap sequence, say Y_n^*, we use $Y_n^* \xrightarrow{p^*}_p 0$, or equivalently $Y_n^* \xrightarrow{p^*} 0$, in probability, to mean that, for any $\epsilon > 0$, $P^*(|Y_n^*| > \epsilon) \to_p 0$, where P^* denotes the probability measure conditional on the original data D_n. An equivalent notation is $Y_n^* = o_{p^*}(1)$ (where we omit the qualification "in probability" for brevity). Similarly, for a double (or second-level) bootstrap sequence, say Y_n^{**}, we write $Y_n^{**} = o_{p^{**}}(1)$ to mean that for all $\epsilon > 0$, $P^{**}(|Y_n^{**}| > \epsilon) \xrightarrow{p^*}_p 0$, where P^{**} is the probability measure conditional on the first-level bootstrap data D_n^* and on D_n.

We use $Y_n^* \xrightarrow{d^*}_p \xi$, or equivalently $Y_n^* \xrightarrow{d^*} \xi$, in probability, to mean that, for all continuity points $u \in \mathbb{R}$ of the cdf of ξ, say $G(u) := P(\xi \leq u)$, it holds that $P^*(Y_n^* \leq u) - G(u) \to_p 0$. Similarly, for a double bootstrap sequence Y_n^{**}, we use $Y_n^{**} \xrightarrow{d^{**}}_{p^*} \xi$, in probability, to mean that $P^{**}(Y_n^{**} \leq u) - G(u) \xrightarrow{p^*}_p 0$ for all continuity points u of G.

2. Examples

In this section we introduce our three leading examples. Example-specific regularity conditions, formally stated results, and additional definitions are given in Appendix B. For each of these examples, we argue that (1.1), (1.2), and (1.3) hold, such that the bootstrap p-values $\hat{p}_n$ are not uniformly distributed rendering standard bootstrap inference invalid. We then return to each example in Section 4, where we discuss how to implement our proposed method and prove its validity.

2.1. Inference after Model Averaging

Setup. We consider inference based on a model averaging estimator obtained as a weighted average of least squares estimators (Hansen 2007). Assume that data are generated according to the linear model

$$y = x\beta + Z\delta + \varepsilon, \tag{2.1}$$

where β is the (scalar) parameter of interest and ε is an n-vector of identically and independently distributed random variables with mean zero and variance σ^2 (henceforth iid$(0,\sigma^2)$), conditional on $W := (x, Z)$.

The researcher fits a set of M models, each of them based on different exclusion restrictions on the q-dimensional vector δ. This setup allows for model averaging both explicitly and implicitly. The former follows, for example, Hansen (2007). The latter includes the common practice of robustness checks in applied research, where the significance of a target coefficient is evaluated through an (often informal) assessment of its significance across a set of regressions based on different sets of controls; see Oster (2019) and the references therein. Specifically, letting R_m denote a $q \times q_m$ selection matrix, the mth model includes x and $Z_m := ZR_m$ as regressors, and the corresponding OLS estimator of β is $\tilde{\beta}_{m,n} := S^{-1}_{xx.Z_m} S_{xy.Z_m}$. Given a set of fixed weights $\omega := (\omega_1, \ldots, \omega_M)'$ such that $\omega_m \in [0,1]$ and $\sum_{m=1}^M \omega_m = 1$, the model averaging estimator is $\tilde{\beta}_n := \sum_{m=1}^M \omega_m \tilde{\beta}_{m,n}$. Then $T_n := n^{1/2}(\tilde{\beta}_n - \beta)$ satisfies $T_n - B_n \to_d \xi_1 \sim N(0, v^2)$, where $v^2 > 0$ and

$$B_n := Q_n n^{1/2}\delta, \quad Q_n := \sum_{m=1}^M \omega_m S^{-1}_{xx.Z_m} S_{xZ.Z_m}.$$

Thus, the magnitude of the asymptotic bias B_n depends on $n^{1/2}\delta$. If δ is local to zero in the sense that $\delta = cn^{-1/2}$ for some vector $c \in \mathbb{R}^q$ (as in, e.g., Hjort and Claeskens 2003; Liu 2015; Hounyo and Lahiri 2023), then $B_n \to_p B := Qc$ with $Q := \text{plim}\, Q_n$, so that (1.1) is satisfied with nonzero B in general. Because B depends on c, which is not consistently estimable, we cannot obtain valid inference from a Gaussian distribution based on sample analogues of B and v^2.

Fixed regressor bootstrap. We generate the bootstrap sample as $y^* = x\hat{\beta}_n + Z\hat{\delta}_n + \varepsilon^*$, where $\varepsilon^*|D_n \sim N(0, \hat{\sigma}^2_n I_n)$, $(\hat{\beta}_n, \hat{\delta}'_n, \hat{\sigma}^2_n)$ is the OLS estimator from the full model, and $D_n = \{y, W\}$. Similar results can be established for the nonparametric bootstrap where ε^* is resampled from the full model residuals. The bootstrap model averaging estimator is given by $\tilde{\beta}^*_n := \sum_{m=1}^M \omega_m \tilde{\beta}^*_{m,n}$, where $\tilde{\beta}^*_{m,n} := S^{-1}_{xx.Z_m} S_{xy^*.Z_m}$. Letting $T^*_n := n^{1/2}(\tilde{\beta}^*_n - \hat{\beta}_n)$, we can show that (1.2) holds with $\hat{B}_n := Q_n n^{1/2}\hat{\delta}_n$ such that, as in (1.3),

$$\hat{B}_n - B_n = Q_n n^{1/2}(\hat{\delta}_n - \delta) \xrightarrow{d} \xi_2 \sim N(0, v_{22}), \quad v_{22} > 0,$$

given in particular the asymptotic normality of $n^{1/2}(\hat{\delta}_n - \delta)$. Because the bias term in the bootstrap world is random in the limit, the conditional distribution of T^*_n is also random in the limit, and in particular does not mimic the asymptotic distribution of the original statistic T_n.

Pairs bootstrap. Consider now a pairs (random design) bootstrap sample $\{y^*_t, x^*_t, z^*_t; t = 1, \ldots, n\}$, based on resampling with replacement from the tuples $\{y_t, x_t, z_t; t = 1, \ldots, n\}$. As is standard, it is useful to recall that the bootstrap data have the representation

$$y^* = x^*\hat{\beta}_n + Z^*\hat{\delta}_n + \varepsilon^*,$$

where $\varepsilon^* = (\varepsilon^*_1, \ldots, \varepsilon^*_n)'$ and ε^*_t is an iid draw from $\hat{\varepsilon}_t = y_t - x_t\hat{\beta}_n - z'_t\hat{\delta}_n$. The pairs bootstrap model averaging estimator is

$$\tilde{\beta}^*_n := \sum_{m=1}^M \omega_m \tilde{\beta}^*_{m,n} \text{ with } \tilde{\beta}^*_{m,n} := S^{-1}_{x^*x^*.Z^*_m} S_{x^*y^*.Z^*_m}$$

and $Z^*_m = Z^*R_m$. The pairs bootstrap statistic is then

$$T^*_n := n^{1/2}(\tilde{\beta}^*_n - \hat{\beta}_n) = B^*_n + n^{1/2} S^{-1}_{x^*x^*} S_{x^*\varepsilon^*},$$

where

$$B^*_n := \sum_{m=1}^M \omega_m S^{-1}_{x^*x^*.Z^*_m} S_{x^*Z^*.Z^*_m} n^{1/2}\hat{\delta}_n.$$

Therefore, and in contrast with the fixed regressor bootstrap (FRB), the term B^*_n is stochastic under the bootstrap probability measure and replaces the bias term $\hat{B}_n$. This difference is not innocuous because it implies that $T^*_n - \hat{B}_n$ no longer replicates the asymptotic distribution of $T_n - B_n$ and (1.2) does not hold. However, this does not prevent our method from working, but it will require a different set of conditions which we will give in Section 3.5.

2.2. Ridge Regression

Setup. We consider estimation of a vector of regression parameters through regularization; in particular, by using a ridge estimator. The model is $y_t = \theta' x_t + \varepsilon_t$, $t = 1, \ldots, n$, where x_t is a $p \times 1$ non-stochastic vector and $\varepsilon_t \sim$ iid$(0, \sigma^2)$. Interest is on testing $\mathsf{H}_0 : g'\theta = r$, based on ridge estimation of θ. Specifically, the ridge estimator has closed form expression $\tilde{\theta}_n = \tilde{S}^{-1}_{xx} S_{xy}$, where $\tilde{S}_{xx} := S_{xx} + n^{-1}c_n I_p$ and c_n is a tuning parameter that controls the degree of shrinkage toward zero. Clearly, $c_n = 0$ corresponds to the OLS estimator, $\hat{\theta}_n$. We are interested in the case where the regressors have limited explanatory power, that is, where $\theta = \delta n^{-1/2}$ is local to zero, which can in fact be taken as a motivation for shrinkage toward zero and hence for ridge estimation. To test H_0, we consider the test statistic $T_n := n^{1/2}(g'\tilde{\theta}_n - r)$. If $n^{-1}c_n \to c_0 \geq 0$ (as in, e.g., Fu and Knight 2000) then, under the null, it holds that $T_n - B_n \to_d \xi_1 \sim N(0, v^2)$, where

$$B_n := -c_n n^{-1/2} g' \tilde{S}^{-1}_{xx}\theta = -c_n n^{-1} g' \tilde{S}^{-1}_{xx}\delta \to B := -c_0 g' \tilde{\Sigma}^{-1}_{xx}\delta$$

with $\tilde{\Sigma}_{xx} := \Sigma_{xx} + c_0 I_p$ and $\Sigma_{xx} := \lim S_{xx}$. Hence, for $c_0 > 0$, $\tilde{\theta}_n$ is asymptotically biased and the bias term cannot be consistently estimated. Consequently, (1.1) is satisfied, and inference based on the quantiles of the $N(0, v^2)$ distribution is invalid unless $c_0 = 0$.

Bootstrap. Consider a pairs (random design) bootstrap sample $\{y^*_t, x^*_t; t = 1, \ldots, n\}$ built by iid resampling from the tuples $\{y_t, x_t; t = 1, \ldots, n\}$. The bootstrap analogue of the ridge estimator is $\tilde{\theta}^*_n := \tilde{S}^{-1}_{x^*x^*} S_{x^*y^*}$, where $\tilde{S}_{x^*x^*} := S_{x^*x^*} + n^{-1}c_n I_p$. The bootstrap statistic is $T^*_n := n^{1/2} g'(\tilde{\theta}^*_n - \hat{\theta}_n)$, which is centered using $\hat{\theta}_n$ to guarantee that ε^*_t and x^*_t are uncorrelated in the bootstrap world. Because we have used a pairs bootstrap, we now have $T^*_n - B^*_n \xrightarrow{d^*}_p \xi_1$ for $B^*_n := -c_n n^{-1/2} g' \tilde{S}^{-1}_{x^*x^*}\hat{\theta}_n$. However, $B^*_n - \hat{B}_n = o_{p^*}(1)$ with $\hat{B}_n := -c_n n^{-1/2} g' \tilde{S}^{-1}_{xx}\hat{\theta}_n$, such that $T^*_n - \hat{B}_n$ still satisfies (1.2). Then (1.3) holds with

$$\hat{B}_n - B_n = -c_n n^{-1} g' \tilde{S}^{-1}_{xx} n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} \xi_2 \sim N(0, v_{22}), \quad v_{22} > 0,$$

so the bootstrap fails to approximate the asymptotic distribution of T_n (see also Chatterjee and Lahiri 2010, 2011).

2.3. Nonparametric Regression

Setup. Consider the model

$$y_t = \beta(x_t) + \varepsilon_t, \quad t = 1, \dots, n, \tag{2.2}$$

where $\beta(\cdot)$ is a smooth function and $\varepsilon_t \sim \text{iid}(0, \sigma^2)$. For simplicity, we consider a fixed-design model; that is, $x_t = t/n$. The goal is inference on $\beta(x)$ for a fixed $x \in (0,1)$. We apply the standard Nadaraya-Watson (fixed-design) estimator $\hat{\beta}_h(x) := (nh)^{-1}\sum_{t=1}^n K((x_t - x)/h)y_t$, where $h = cn^{-1/5}$ for some $c > 0$ is the MSE-optimal bandwidth and K is the kernel function. We do not consider the more general local polynomial regression case, although we conjecture that very similar results will hold. We leave that case for future research. The statistic $T_n := (nh)^{1/2}(\hat{\beta}_h(x) - \beta(x))$ satisfies $T_n - B_n \to_d \xi_1 \sim N(0, v^2)$, where $v^2 := \sigma^2 \int K(u)^2 du > 0$ and

$$B_n := (nh)^{1/2}\left(\frac{1}{nh}\sum_{t=1}^n k_t \beta(x_t) - \beta(x)\right) \tag{2.3}$$

with $k_t := K((x_t - x)/h)$. The bias B_n satisfies

$$B_n = (nh)^{1/2}(h^2\beta''(x)\kappa_2/2 + o(h^2)) \to B := c^{5/2}\beta''(x)\kappa_2/2, \tag{2.4}$$

where $\kappa_2 := \int u^2 K(u)du$ and $\beta''(x)$ denotes the second-order derivative of $\beta(x)$. Thus, (1.1) is satisfied. Estimating B or B_n is challenging because it involves estimating $\beta''(x)$, and although theoretically valid estimators exist, they perform poorly in finite samples. This issue is pointed out by Calonico, Cattaneo, and Titunik (2014) and Calonico, Cattaneo, and Farrell (2018), who propose more accurate bias correction techniques specifically for regression discontinuity designs and nonparametric curve estimation.

Bootstrap. The (parametric) bootstrap sample is generated as $y_t^* = \hat{\beta}_h(x_t) + \varepsilon_t^*$, $t = 1, \dots, n$, where $\varepsilon_t^* | D_n \sim \text{iid}N(0, \hat{\sigma}_n^2)$ with $D_n = \{y_t; t = 1, \dots, n\}$ and $\hat{\sigma}_n^2$ denotes a consistent estimator of σ^2; for example the residual variance. Let $\hat{\beta}_h^*(x) := (nh)^{-1}\sum_{t=1}^n k_t y_t^*$ and $T_n^* := (nh)^{1/2}(\hat{\beta}_h^*(x) - \hat{\beta}_h(x))$. Then (1.2) is satisfied with

$$\hat{B}_n := (nh)^{1/2}\left(\frac{1}{nh}\sum_{t=1}^n k_t \hat{\beta}_h(x_t) - \hat{\beta}_h(x)\right).$$

Because $h = cn^{-1/5}$, (1.3) holds with

$$\hat{B}_n - B_n = (nh)^{1/2}\left(\frac{1}{nh}\sum_{t=1}^n k_t(\hat{\beta}_h(x_t) - \beta(x_t)) - (\hat{\beta}_h(x) - \beta(x))\right)$$
$$\xrightarrow{d} \xi_2 \sim N(0, v_{22}),$$

where $v_{22} > 0$, so the bootstrap is invalid. Two possible solutions to this problem are to generate the bootstrap sample as $y_t^* = \hat{\beta}_g(x_t) + \varepsilon_t^*$, where g is an oversmoothing bandwidth satisfying $ng^5 \to \infty$ (e.g., Härdle and Marron 1991) or to center the bootstrap statistic at its expected value and add a consistent estimator of B (e.g., Härdle and Bowman 1988; Eubank and Speckman 1993). Both approaches require selecting two bandwidths, which is not straightforward. An alternative approach, suggested by Hall and Horowitz (2013), focuses on an asymptotic theory-based confidence interval and applies the bootstrap to calibrate its coverage probability. However, this requires an additional averaging step across a grid of x (their step 6) to asymptotically eliminate ξ_2, and it results in an asymptotically conservative interval. Finally, a non-bootstrap-based solution is undersmoothing using a bandwidth h satisfying $nh^5 \to 0$, although of course that is not MSE-optimal and may result in trivial power against certain local alternatives; see Section 4.3.

3. General Results

3.1. Framework and Invalidity of the Standard Bootstrap

The general framework is as follows. We have a statistic T_n defined as a general function of a sample D_n, for which we would like to compute a valid bootstrap p-value. Usually, T_n is a test statistic or a (possibly normalized) parameter estimator; for example, $T_n = g(n)(\hat{\theta}_n - \theta_0)$. Let D_n^* denote the bootstrap sample, which depends on the original data and on some auxiliary bootstrap variates (which we assume defined jointly with D_n on a possibly extended probability space). Let T_n^* denote the bootstrap version of T_n computed on D_n^*; for example, $T_n^* = g(n)(\hat{\theta}_n^* - \hat{\theta}_n)$. Let $\hat{L}_n(u) := P^*(T_n^* \le u)$, $u \in \mathbb{R}$, denote its distribution function, conditional on the original data. The bootstrap p-value is defined as

$$\hat{p}_n := P^*(T_n^* \le T_n) = \hat{L}_n(T_n).$$

First-order asymptotic validity of $\hat{p}_n$ requires that $\hat{p}_n$ converges in distribution to a standard uniform distribution; that is, that $\hat{p}_n \to_d U_{[0,1]}$. In this section we focus on a class of statistics T_n and T_n^* for which this condition is not necessarily satisfied. The main reason is the presence of an additive "bias" term B_n that contaminates the distribution of T_n and cannot be replicated by the bootstrap distribution of T_n^*.

Assumption 1. $T_n - B_n \to_d \xi_1$, where ξ_1 is centered at zero and the cdf $G(u) = P(\xi_1 \le u)$ is continuous and strictly increasing over its support.

When B_n converges to a nonzero constant B, Assumption 1 can be written $T_n \to_d B + \xi_1$ as in (1.1). If T_n is a normalized version of a (scalar) parameter estimator, that is, $T_n = g(n)(\hat{\theta}_n - \theta_0)$, then we can think of B as the asymptotic bias of $\hat{\theta}_n$ because ξ_1 is centered at zero. Although we allow for the possibility that B_n does not have a limit (and it may even diverge), we will still refer to B_n as a "bias term". More generally, in Assumption 1 we cover any statistic T_n that is not necessarily Gaussian (even asymptotically) and whose limiting distribution is G only after we subtract the sequence B_n. The limiting distribution G may depend on a parameter such that $T_n - B_n$ is not an asymptotic pivot.

Inference based on the asymptotic distribution of T_n requires estimating B_n and any parameter in G. Alternatively, we can use the bootstrap to bypass parameter estimation and directly compute a bootstrap p-value that relies on T_n^* and T_n alone; that is, we consider $\hat{p}_n := P^*(T_n^* \le T_n)$. A set of high-level conditions on T_n^* and T_n that allow us to derive the asymptotic properties of this p-value are described next.

Assumption 2. For some D_n-measurable random variable $\hat{B}_n$, it holds that: (i) $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1$, where ξ_1 is described in Assumption 1; (ii)

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix},$$

where ξ_2 is centered at zero and $F(u) = P(\xi_1 - \xi_2 \le u)$ is a continuous cdf.

Assumption 2(i) states that $T_n^* - \hat{B}_n$ converges in distribution to a random variable ξ_1 having the same distribution function G as $T_n - B_n$.[1] Thus, $\hat{B}_n$ can be thought of as an implicit bootstrap bias that affects the statistic T_n^*, in the same way that B_n affects the original statistic T_n. Assumption 2(ii) complements Assumption 1 by requiring the joint convergence of $T_n - B_n$ and $\hat{B}_n - B_n$ towards ξ_1 and ξ_2, respectively; see also (1.1)–(1.3).

Given Assumption 2(i), we could use the bootstrap distribution of $T_n^* - \hat{B}_n$ to approximate the distribution of $T_n - B_n$. Since B_n is typically unknown, this result is not very useful for inference unless $\hat{B}_n$ is consistent for B_n. In this case, Assumption 2 together with Assumption 1 imply that $\hat{p}_n$ is asymptotically distributed as $U_{[0,1]}$. This follows by noting that if $\hat{B}_n - B_n = o_p(1)$, then $\xi_2 = 0$ a.s., implying that $F(u) = G(u)$. Consequently,

$$\begin{aligned} \hat{p}_n &:= P^*(T_n^* \le T_n) = P^*(T_n^* - \hat{B}_n \le T_n - \hat{B}_n) \\ &= G(T_n - \hat{B}_n) + o_p(1) \text{ (by Assumption 2(i))} \\ &\xrightarrow{d} G(\xi_1 - \xi_2) \text{ (by Assumption 2(ii) and continuity of } G) \\ &\sim U_{[0,1]}, \end{aligned}$$

where the last distributional equality holds by $F = G$ and the probability integral transform. However, this result does not hold if $\hat{B}_n - B_n$ does not converge to zero in probability. Specifically, if $\hat{B}_n - B_n \to_d \xi_2$ (jointly with $T_n - B_n \to_d \xi_1$), then

$$T_n - \hat{B}_n = (T_n - B_n) - (\hat{B}_n - B_n) \xrightarrow{d} \xi_1 - \xi_2 \sim F^{-1}(U_{[0,1]})$$

under Assumptions 1 and 2(ii). When ξ_2 is nondegenerate, $F \neq G$, implying that $\hat{p}_n = G(T_n - \hat{B}_n) + o_p(1)$ is not asymptotically distributed as a standard uniform random variable. This result is summarized in the following theorem.

Theorem 3.1. Suppose Assumptions 1 and 2 hold. Then $\hat{p}_n \to_d G(F^{-1}(U_{[0,1]}))$.

Proof. First notice that $\hat{p}_n$ and $G(T_n - \hat{B}_n)$ have the same asymptotic distribution because Assumption 2(i) and continuity of G imply that, by Polya's theorem,

$$|\hat{p}_n - G(T_n - \hat{B}_n)| \le \sup_{u \in \mathbb{R}} |P^*(T_n^* - \hat{B}_n \le u) - G(u)| \xrightarrow{p} 0.$$

Next, by Assumption 2(ii), $T_n - \hat{B}_n \to_d \xi_1 - \xi_2$, such that

$$G(T_n - \hat{B}_n) \xrightarrow{d} G(\xi_1 - \xi_2)$$

by continuity of G and the continuous mapping theorem. Since $\xi_1 - \xi_2$ has continuous cdf F, it holds that $\xi_1 - \xi_2 \sim F^{-1}(U_{[0,1]})$, which completes the proof. □

Remark 3.1. The value of $\hat{B}_n$ in Assumption 2(i) depends on the chosen bootstrap algorithm. It is possible that $\hat{B}_n \to_p 0$ for some bootstrap algorithms; examples are given in Remark B.2 and Appendix B.5. If this is the case, then $\xi_2 = -B$ a.s., which implies that

$$F(u) := P(\xi_1 - \xi_2 \le u) = P(\xi_1 \le u - B) = G(u - B),$$

and hence Assumption 2(ii) is not satisfied. In this case the bootstrap p-value satisfies

$$\hat{p}_n \xrightarrow{d} G(G^{-1}(U_{[0,1]}) + B).$$

Note that this distribution is uniform only if $B = 0$. Hence, the p-value depends on B, even in the limit.

Remark 3.2. Under Assumptions 1 and 2, standard bootstrap (percentile) confidence sets are also in general invalid. Consider, for example, the case where $T_n = g(n)(\hat{\theta}_n - \theta_0)$ and T_n^* is its bootstrap analogue with (conditional) distribution function $\hat{L}_n(u)$. A right-sided confidence set for θ_0 at nominal confidence level $1 - \alpha \in (0, 1)$ can be obtained as (e.g., Horowitz 2001, p. 3171) $CI_n^{1-\alpha} := [\hat{\theta}_n - g(n)^{-1}\hat{q}_n(1 - \alpha), +\infty)$, where $\hat{q}_n(1 - \alpha) := \hat{L}_n^{-1}(1 - \alpha)$. Then

$$\begin{aligned} P(\theta_0 \in CI_n^{1-\alpha}) &= P(\hat{\theta}_n - g(n)^{-1}\hat{q}_n(1 - \alpha) \le \theta_0) \\ &= P(T_n \le \hat{q}_n(1 - \alpha)) \\ &= P(\hat{L}_n(T_n) \le 1 - \alpha) \\ &= P(\hat{p}_n \le 1 - \alpha) \nrightarrow 1 - \alpha \end{aligned}$$

because, by Theorem 3.1, $\hat{p}_n$ is not asymptotically uniformly distributed.

Remark 3.3. It is worth noting that, under Assumptions 1 and 2, the bootstrap (conditional) distribution is random in the limit whenever ξ_2 is nondegenerate. Specifically, assume for simplicity that $B_n \to_p B$. Recall that $\hat{L}_n(u) := P^*(T_n^* \le u)$, $u \in \mathbb{R}$, and let $\hat{G}_n(u) := P^*(T_n^* - \hat{B}_n \le u)$. It then holds that

$$\hat{L}_n(u) = \hat{G}_n(u - \hat{B}_n) = G(u - B - (\hat{B}_n - B)) + \hat{a}_n(u),$$

where $\hat{a}_n(u) \le \sup_{u \in \mathbb{R}} |\hat{G}_n(u) - G(u)| = o_p(1)$ by Assumption 2(i), continuity of G, and Polya's theorem. Because $\hat{B}_n - B \to_d \xi_2$, it follows that when ξ_2 is nondegenerate, $\hat{L}_n(u) \to_w G(u - B - \xi_2)$, where $\to_w$ denotes weak convergence of cdf's as (random) elements of a function space (see Cavaliere and Georgiev 2020). The presence of ξ_2 in $G(u - B - \xi_2)$ makes this a random cdf.[2] Therefore, the bootstrap is unable to mimic the asymptotic distribution of T_n, which is $G(u - B)$ by Assumption 1.

[1] Note that we write $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1$ to mean that $T_n^* - \hat{B}_n$ has (conditionally on D_n) the same asymptotic distribution function as the random variable ξ_1. We could alternatively write that $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1^*$ and $T_n - B_n \xrightarrow{d} \xi_1$ where ξ_1^* and ξ_1 are two independent copies of the same distribution, that is, $P(\xi_1 \le u) = P(\xi_1^* \le u)$. We do not make this distinction because we care only about distributional results, but it should be kept in mind.

[2] The same result follows in terms of weak convergence in distribution of $T_n^*|D_n$. Specifically, because $T_n^* = (T_n^* - \hat{B}_n) + (\hat{B}_n - B_n) + B_n$, where $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1^*$ and (jointly) $\hat{B}_n - B_n \xrightarrow{d} \xi_2$ with $\xi_1^* \sim \xi_1$ independent of ξ_2, we have that $T_n^*|D_n \xrightarrow{w} (B + \xi_1^* + \xi_2)|\xi_2$.

Next, we describe two possible solutions to the invalidity of the standard bootstrap p-value $\hat{p}_n$. One relies on the prepivoting approach of Beran (1987, 1988); see Section 3.2. The basic idea is that we modify $\hat{p}_n$ by applying the mapping $\hat{p}_n \mapsto H(\hat{p}_n)$, where $H(u)$ is the asymptotic cdf of $\hat{p}_n$, which makes the modified p-value $H(\hat{p}_n)$ asymptotically standard uniform. Contrary to Beran (1987, 1988), who proposed prepivoting as a way of providing asymptotic refinements for the bootstrap, here we show how to use prepivoting to solve the invalidity of the standard bootstrap p-value $\hat{p}_n$. This result is new in the bootstrap literature. The second approach relies on computing a standard bootstrap p-value based on the modified statistic given by $T_n - \hat{B}_n$; see Section 3.4. Thus, we modify the test statistic rather than modifying the way we compute the bootstrap p-value.

3.2. Prepivoting

Theorem 3.1 implies that

$$\begin{aligned} P(\hat{p}_n \le u) &\to P(G(F^{-1}(U_{[0,1]})) \le u) \\ &= P(U_{[0,1]} \le F(G^{-1}(u))) = F(G^{-1}(u)) =: H(u) \end{aligned}$$

uniformly over $u \in [0,1]$ by Polya's theorem, given the continuity of G and F. Although H is not the uniform distribution, unless $G = F$, it is continuous because G is strictly increasing. Thus, the following corollary to Theorem 3.1 holds by the probability integral transform.

Corollary 3.1. Under the conditions of Theorem 3.1, $H(\hat{p}_n) \to_d U_{[0,1]}$.

Therefore, the mapping of $\hat{p}_n$ into $H(\hat{p}_n)$ transforms $\hat{p}_n$ into a new p-value, $H(\hat{p}_n)$, whose asymptotic distribution is the standard uniform distribution on $[0, 1]$. Inference based on $H(\hat{p}_n)$ is generally infeasible, because we do not observe $H(u)$. However, if we can replace $H(u)$ with a uniformly consistent estimator $\hat{H}_n(u)$ then this approach will deliver a feasible modified p-value $\tilde{p}_n := \hat{H}_n(\hat{p}_n)$. Since the limit distribution of $\tilde{p}_n$ is the standard uniform distribution, $\tilde{p}_n$ is an asymptotically valid p-value. The mapping of $\hat{p}_n$ into $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$ by the estimated distribution of the former corresponds to what Beran (1987) calls "prepivoting." In the following sections, we describe two methods of obtaining a consistent estimator of $H(u)$.

Remark 3.4. The prepivoting approach can also be used to solve the invalidity of confidence sets based on the standard bootstrap; see Remark 3.2. In particular, replace the nominal level $1-\alpha$ by $\hat{H}_n^{-1}(1-\alpha)$ and consider $\widetilde{CI}_n^{1-\alpha} := [\hat{\theta}_n - g(n)^{-1}\hat{q}_n(\hat{H}_n^{-1}(1-\alpha)), +\infty)$. Then

$$\begin{aligned} P(\theta_0 \in \widetilde{CI}_n^{1-\alpha}) &= P(\hat{p}_n \le \hat{H}_n^{-1}(1-\alpha)) \\ &= P(\hat{H}_n(\hat{p}_n) \le 1-\alpha) \to 1-\alpha, \end{aligned}$$

where the last convergence is implied by Corollary 3.1 and consistency of $\hat{H}_n$.

Remark 3.5. Corollary 3.1 can also be applied to right-tailed or two-tailed tests. The right-tailed p-value, say $\hat{p}_{n,r} := P^*(T_n^* > T_n) = 1 - \hat{L}_n(T_n) = 1 - \hat{p}_n$, has cdf $P(\hat{p}_{n,r} \le u) = P(\hat{p}_n \ge 1-u) = 1-P(\hat{p}_n < 1-u) = 1-H(1-u)+o(1)$ uniformly in u. Note that, because the conditional cdf of T_n^* is continuous in the limit, the p-value $\hat{p}_{n,r}$ is asymptotically equivalent to $P^*(T_n^* \ge T_n)$. Thus, by Corollary 3.1, the modified right-tailed p-value, $\tilde{p}_{n,r} := 1 - \hat{H}_n(\hat{p}_{n,r})$, satisfies

$$\tilde{p}_{n,r} = 1 - H(1-\hat{p}_{n,r}) + o_p(1) = 1 - H(\hat{p}_n) + o_p(1) \xrightarrow{d} U_{[0,1]}.$$

Similarly, for two-tailed tests the equal-tailed bootstrap p-value, $\tilde{p}_{n,\text{et}} := 2\min\{\tilde{p}_n, \tilde{p}_{n,r}\} = 2\min\{\tilde{p}_n, 1-\tilde{p}_n\}$, satisfies $\tilde{p}_{n,\text{et}} \to_d U_{[0,1]}$ by Corollary 3.1 and the continuous mapping theorem.

3.2.1. Plug-in Approach

Suppose $H(u) = H_\gamma(u)$ depends on a finite-dimensional parameter, γ. In view of Theorem 3.1, a simple approach to estimating $H(u)$ is to use

$$\hat{H}_n(u) = H_{\hat{\gamma}_n}(u),$$

where $\hat{\gamma}_n$ denotes a consistent estimator of γ. This leads to a plug-in modified p-value defined as

$$\tilde{p}_n = H_{\hat{\gamma}_n}(\hat{p}_n).$$

By consistency of $\hat{\gamma}_n$ and under the assumption that H_γ is continuous in γ, it follows immediately that

$$\tilde{p}_n = H(\hat{p}_n) + o_p(1) \xrightarrow{d} F(G^{-1}(G(F^{-1}(U_{[0,1]})))) = U_{[0,1]}.$$

This result is summarized next.

Corollary 3.2. Let Assumptions 1 and 2 hold, and suppose $H_\gamma(u)$ is continuous in (γ, u). If $\hat{\gamma}_n \to_p \gamma$ then $\tilde{p}_n = H_{\hat{\gamma}_n}(\hat{p}_n) \to_d U_{[0,1]}$.

The plug-in approach relies on a consistent estimator of the asymptotic distribution H, but does not require estimating the "bias term" B_n. When estimating γ is simple, this approach is attractive since it does not require any double resampling. Examples are given in Section 4. However, computation of γ is case-specific and may be cumbersome in practice. An automatic approach is to use the bootstrap to estimate $H(u)$, as we describe next.

3.2.2. Double Bootstrap

Following Beran (1987, 1988), we can estimate $H(u)$ with the bootstrap. That is, we let

$$\hat{H}_n(u) = P^*(\hat{p}_n^* \le u),$$

where $\hat{p}_n^*$ is the bootstrap analogue of $\hat{p}_n$. Since $\hat{p}_n$ is itself a bootstrap p-value, computing $\hat{p}_n^*$ requires a double bootstrap. In particular, let D_n^{**} denote a further bootstrap sample of size n based on D_n^* and some additional bootstrap variates (defined jointly with D_n and D_n^* on a possibly extended probability space), and let T_n^{**} denote the bootstrap version of T_n^* computed on D_n^{**}. With this notation, the second-level bootstrap p-value is defined as

$$\hat{p}_n^* := P^{**}(T_n^{**} \le T_n^*),$$

where P^{**} denotes the bootstrap probability measure conditional on D_n^* and D_n (making $\hat{p}_n^*$ a function of D_n^* and D_n). This leads to a double bootstrap modified p-value, as given by

$$\tilde{p}_n := \hat{H}_n(\hat{p}_n) = P^*(\hat{p}_n^* \leq \hat{p}_n).$$

In order to show that $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$, we add the following assumption.

Assumption 3. Let ξ_1 and ξ_2 be as defined in Assumptions 1 and 2. For some (D_n^*, D_n)-measurable random variable $\hat{B}_n^*$, it holds that: (i) $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} \xi_1$, in probability, and (ii) $T_n^* - \hat{B}_n^* \xrightarrow{d^*}_p \xi_1 - \xi_2$.

Assumption 3 complements Assumptions 1 and 2 by imposing high-level conditions on the second-level bootstrap statistics. Specifically, Assumption 3(i) assumes that T_n^{**} has asymptotic distribution G only after we subtract $\hat{B}_n^*$. This term is the second-level bootstrap analogue of $\hat{B}_n$. It depends only on the first-level bootstrap data D_n^* and is not random under P^{**}. The second part of Assumption 3 follows from Assumption 2 in the special case that $\hat{B}_n^* - \hat{B}_n = o_{p^*}(1)$, in probability; that is, when $\xi_2 = 0$ a.s., implying $F = G$. When $F \neq G$, $\hat{B}_n^*$ is not a consistent estimator of $\hat{B}_n$. However, under Assumption 3,

$$T_n^* - \hat{B}_n^* = (T_n^* - \hat{B}_n) - (\hat{B}_n^* - \hat{B}_n) \xrightarrow{d^*}_p \xi_1 - \xi_2 = F^{-1}(U_{[0,1]})$$

implying that $T_n^* - \hat{B}_n^*$ mimics the distribution of $T_n - \hat{B}_n$. This suffices for proving the asymptotic validity of the double bootstrap modified p-value, $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$, as proved next.

Theorem 3.2. Under Assumptions 1, 2, and 3, it holds that $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$.

Proof. To prove this result, recall that $\hat{H}_n(u) = P^*(\hat{p}_n^* \leq u)$ and $P(\hat{p}_n \leq u) \to H(u) = F(G^{-1}(u))$ uniformly in $u \in \mathbb{R}$, since H is a continuous distribution function by Assumptions 1 and 2. Thus, we have that

$$\begin{aligned}\hat{p}_n^* &= P^{**}(T_n^{**} \leq T_n^*) = P^{**}(T_n^{**} - \hat{B}_n^* \leq T_n^* - \hat{B}_n^*)\\ &= G(T_n^* - \hat{B}_n^*) + o_{p^*}(1), \quad \text{by Assumption 3(i)},\\ &= G(F^{-1}(U_{[0,1]})) + o_{p^*}(1), \quad \text{by Assumption 3(ii)},\end{aligned}$$

where $G(F^{-1}(U_{[0,1]}))$ is a random variable whose distribution function is H. Hence,

$$\sup_{u\in\mathbb{R}} |\hat{H}_n(u) - H(u)| = o_p(1).$$

Since $H(\hat{p}_n) \to_d U_{[0,1]}$, we can conclude that $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$. □

Theorem 3.2 shows that prepivoting the standard bootstrap p-value $\hat{p}_n$ by applying the mapping $\hat{H}_n$ transforms it into an asymptotically uniformly distributed random variable. This result holds under Assumptions 1, 2, and 3, independently of whether $G = F$ or not. When $G = F$ then $\hat{p}_n \to_d U_{[0,1]}$ (as implied by Theorem 3.1). In this case, the prepivoting approach is not necessary to obtain a first-order asymptotically valid test, although it might help further reducing the size distortion of the test. This corresponds to the setting of Beran (1987, 1988), where prepivoting was proposed as a way of reducing the level distortions of confidence intervals. When $G \neq F$ then $\hat{p}_n$ is not asymptotically uniform and a standard bootstrap test based on $\hat{p}_n$ is asymptotically invalid, as shown in Theorem 3.1. In this case, prepivoting transforms an asymptotically invalid bootstrap p-value into one that is asymptotically valid. This setting was not considered by Beran (1987, 1988) and is new to our article.

3.3. Power of Tests

In this section we explicitly consider a testing situation. Suppose we are interested in testing $\mathsf{H}_0 : \theta = \bar{\theta}$ against $\mathsf{H}_1 : \theta < \bar{\theta}$. Specifically, defining $T_n(\theta) := g(n)(\hat{\theta}_n - \theta)$, we consider the test statistic $T_n(\bar{\theta})$. The corresponding bootstrap p-value is $\hat{p}_n(\bar{\theta})$ with $\hat{p}_n(\theta) := P^*(T_n^* \leq T_n(\theta))$. When the null hypothesis is true, that is, when $\bar{\theta} = \theta_0$ with θ_0 denoting the true value, we find $T_n(\bar{\theta}) = T_n(\theta_0) = T_n$ and $\hat{p}_n(\bar{\theta}) = \hat{p}_n(\theta_0) = \hat{p}_n$, where T_n and $\hat{p}_n$ are as defined previously. If Assumptions 1 and 2 hold under the null, Theorem 3.1 and Corollary 3.1 imply that tests based on $H(\hat{p}_n(\bar{\theta}))$ have correct asymptotic size, where H continues to denote the asymptotic cdf of $\hat{p}_n$.

To analyze power, we consider $\theta_0 = \bar{\theta} + a_n$ for some deterministic sequence a_n. Then $a_n = 0$ under the null hypothesis, whereas $a_n = a < 0$ corresponds to a fixed alternative and $a_n = a/g(n)$ for $a < 0$ corresponds to a local alternative. Thus, we define $\pi_n := g(n)(\theta_0 - \bar{\theta}) = g(n)a_n$ so that $T_n(\bar{\theta}) = T_n + \pi_n$.

Theorem 3.3. Suppose Assumptions 1 and 2 hold. (i) If $\pi_n \to \pi$ then $H(\hat{p}_n(\bar{\theta})) \to_d F(F^{-1}(U_{[0,1]}) + \pi)$. (ii) If $\pi_n \to -\infty$ then $P(H(\hat{p}_n(\bar{\theta})) \leq \alpha) \to 1$ for any nominal level $\alpha > 0$.

Proof. As in the proof of Theorem 3.1 we have, by Assumption 2(i),

$$\begin{aligned}\hat{p}_n(\bar{\theta}) = P^*(T_n^* \leq T_n(\bar{\theta})) &= P^*(T_n^* - \hat{B}_n \leq T_n - \hat{B}_n + \pi_n)\\ &= G(T_n - \hat{B}_n + \pi_n) + o_p(1).\end{aligned}$$

If $\pi_n \to \pi$ then $\hat{p}_n(\bar{\theta}) \to_d G(F^{-1}(U_{[0,1]}) + \pi)$ by Assumption 2(ii), so that

$$H(\hat{p}_n(\bar{\theta})) \xrightarrow{d} H(G(F^{-1}(U_{[0,1]}) + \pi)) = F(F^{-1}(U_{[0,1]}) + \pi)$$

by definition of $H(u)$. If $\pi_n \to -\infty$ then $\hat{p}_n(\bar{\theta}) \to_p 0$ because $T_n - \hat{B}_n = O_p(1)$ by Assumption 2(ii), so that $H(\hat{p}_n(\bar{\theta})) \to_p H(0) = 0$ and $P(H(\hat{p}_n(\bar{\theta})) \leq \alpha) \to 1$ for any $\alpha > 0$. □

It follows from Theorem 3.3(ii) that a left-tailed test that rejects for small values of $H(\hat{p}_n(\bar{\theta}))$ is consistent. Furthermore, it follows from Theorem 3.3(i) that such a test has nontrivial asymptotic local power against $\pi < 0$. Specifically, the asymptotic local power against π is given by $P(H(\hat{p}_n(\bar{\theta})) \leq \alpha) \to F(F^{-1}(\alpha) - \pi)$. Interestingly, this only depends on F and not on G. As above, to implement the modified p-value, $H(\hat{p}_n(\bar{\theta}))$, in practice, we would need a (uniformly) consistent estimator of H, that is, the asymptotic distribution of the bootstrap p-value when the null hypothesis is true. This could be either the plug-in or double bootstrap estimators, as discussed in Sections 3.2.1 and 3.2.2.

Note that Assumption 2 is still assumed to hold in Theorem 3.3. That is, the bootstrap statistic T_n^* is assumed to have the same asymptotic behavior under the null and under the alternative. This is commonly the case when the bootstrap algorithm does not impose the null hypothesis when generating the bootstrap data.

3.4. Bootstrap p-value based on $T_n - \hat{B}_n$

The double bootstrap modified p-value $\tilde{p}_n$ depends only on the statistic T_n and its bootstrap analogues T_n^* and T_n^{**}. It does not involve computing explicitly $\hat{B}_n$ or $\hat{B}_n^*$, but in some applications it can be computationally costly as it requires two levels of resampling. As it turns out, $\tilde{p}_n$ is asymptotically equivalent to a single-level bootstrap p-value that is based on bootstrapping the statistic $T_n - \hat{B}_n$, as we show next.

By definition, the double bootstrap modified p-value is given by $\tilde{p}_n := P^*(\hat{p}_n^* \le \hat{p}_n)$, where

$$\begin{aligned}\hat{p}_n^* &:= P^{**}(T_n^{**} \le T_n^*) = P^{**}(T_n^{**} - \hat{B}_n^* \le T_n^* - \hat{B}_n^*)\\ &= G(T_n^* - \hat{B}_n^*) + o_{p^*}(1),\end{aligned}$$

in probability, given Assumption 3. Similarly, under Assumptions 1 and 2,

$$\begin{aligned}\hat{p}_n &:= P^*(T_n^* \le T_n) = P^*(T_n^* - \hat{B}_n \le T_n - \hat{B}_n)\\ &= G(T_n - \hat{B}_n) + o_p(1).\end{aligned}$$

It follows that

$$\begin{aligned}\tilde{p}_n &:= P^*(\hat{p}_n^* \le \hat{p}_n) = P^*(G(T_n^* - \hat{B}_n^*) \le G(T_n - \hat{B}_n)) + o_p(1)\\ &= P^*(T_n^* - \hat{B}_n^* \le T_n - \hat{B}_n) + o_p(1)\end{aligned}$$

because G is continuous. We summarize this result in the following corollary.

Corollary 3.3. Under Assumptions 1, 2, and 3, $\tilde{p}_n = P^*(T_n^* - \hat{B}_n^* \le T_n - \hat{B}_n) + o_p(1)$.

Theorem 3.2 shows that $\tilde{p}_n \to_d U_{[0,1]}$ and hence is asymptotically valid. In view of this, Corollary 3.3 shows that removing $\hat{B}_n$ from T_n and computing a bootstrap p-value based on the new statistic, $T_n - \hat{B}_n$, also solves the invalidity problem of the standard bootstrap p-value, $\hat{p}_n = P^*(T_n^* \le T_n)$. Note that we do not require $\xi_2 = 0$, that is, $\hat{B}_n - B_n$ and $\hat{B}_n^* - \hat{B}_n$ do not need to converge to zero.

When $\hat{B}_n$ and $\hat{B}_n^*$ are easy to compute, for example, when they are available analytically as functions of D_n and D_n^*, respectively, Corollary 3.3 is useful as it avoids implementing a double bootstrap. When this is not the case, that is, when deriving $\hat{B}_n$ and $\hat{B}_n^*$ explicitly is cumbersome or impossible, we may be able to estimate $\hat{B}_n$ from the bootstrap and $\hat{B}_n^*$ from a double bootstrap. Corollary 3.3 then shows that the double bootstrap modified p-value $\tilde{p}_n$ is a convenient alternative since it depends only on T_n, T_n^*, and T_n^{**}. It is important to note that none of these approaches requires the consistency of $\hat{B}_n$ and $\hat{B}_n^*$.

3.5. A More General Set of High-Level Conditions

We conclude this section by providing an alternative set of high-level conditions that cover bootstrap methods for which $T_n^* - \hat{B}_n$ has a different limiting distribution than $T_n - B_n$. This may happen, for example, for the pairs bootstrap; see Section 2.1 and Remark 3.6.

Assumption 4. Assumption 2 holds with part (i) replaced by (i) $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1$, where ζ_1 is centered at zero and the cdf $J(u) = P(\zeta_1 \le u)$ is continuous and strictly increasing over its support.

Under Assumption 4, $T_n^* - \hat{B}_n$ does not replicate the distribution of $T_n - B_n$. This is to be understood in the sense that there does not exist a P^*-measurable term $\hat{B}_n$ such that $T_n^* - \hat{B}_n$ has the same asymptotic distribution as $T_n - B_n$.

An important generalization provided by Assumption 4 compared with Assumption 2 is to allow for bootstrap methods where the "centering term," say B_n^*, depends on the bootstrap data. That is, to allow cases where there is a random (with respect to P^*, that is, depending on the bootstrap data) term B_n^* such that $T_n^* - B_n^* \xrightarrow{d^*}_p \xi_1$ and hence has the same asymptotic distribution as $T_n - B_n$. Clearly, this violates Assumption 2 unless $B_n^* - \hat{B}_n \xrightarrow{p^*}_p 0$ (as in the ridge regression in Section 2.2). However, letting ζ_1 be such that $B_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1 - \xi_1$, then Assumption 4 covers the former case.

Remark 3.6. A leading example where $T_n^* - B_n^* \xrightarrow{d^*}_p \xi_1$ and hence has the same asymptotic distribution as $T_n - B_n$ is the pairs bootstrap as in Section 2.1 for the model averaging example. We study this case in more detail in Section 4.1.

The asymptotic distribution of the bootstrap p-value under Assumption 4 is given in the following theorem. The proof is identical to that of Theorem 3.1, with G replaced by J, and hence omitted.

Theorem 3.4. If Assumptions 1 and 4 hold then $\hat{p}_n \to_d J(F^{-1}(U_{[0,1]}))$.

Theorem 3.4 implies that now $P(\hat{p}_n \le u) \to P(J(F^{-1}(U_{[0,1]})) \le u) = F(J^{-1}(u)) =: H(u)$. Clearly, a plug-in approach to estimating this $H(u)$ based on G as described in Section 3.2.1 would be invalid because $G \neq J$ in general. However, it follows straightforwardly by the same arguments as applied in Section 3.2.1 that a plug-in approach based on J will deliver an asymptotically valid plug-in modified p-value.

To implement an asymptotically valid double bootstrap modified p-value we consider the following high-level condition.

Assumption 5. Assumption 3 holds with part (i) replaced by (i) $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} \zeta_1$, in probability, where ζ_1 is defined in Assumption 4.

Under Assumption 5, the second-level bootstrap statistic, $T_n^{**} - \hat{B}_n^*$, replicates the distribution of the first-level statistic, $T_n^* - \hat{B}_n$. Thus, the second-level bootstrap p-value is

$$\begin{aligned}\hat{p}_n^* := P^{**}(T_n^{**} \le T_n^*) &= P^{**}(T_n^{**} - \hat{B}_n^* \le T_n^* - \hat{B}_n^*)\\ &= J(T_n^* - \hat{B}_n^*) + o_{p^*}(1)\\ &\xrightarrow{d^*}_p J(\xi_1 - \xi_2) = J(F^{-1}(U_{[0,1]}))\end{aligned}$$

under Assumption 5. Hence, the second-level bootstrap p-value has the same asymptotic distribution as the original bootstrap p-value. It follows that the double bootstrap modified p-value, $\tilde{p}_n := \hat{H}_n(\hat{p}_n) = P^*(\hat{p}_n^* \le \hat{p}_n)$, is asymptotically valid, which is stated next. The proof is essentially identical to that of Theorem 3.2 and hence omitted.

Theorem 3.5. Under Assumptions 1, 4, and 5, it holds that $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$.

Remark 3.7. Consider again the case with a random bootstrap centering term in Remark 3.6, where $B_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1 - \xi_1$ such that $T_n^* - B_n^* \xrightarrow{d^*}_p \xi_1$. Within this setup, we can consider double bootstrap methods such that, for a random (with respect to P^{**}) term B_n^{**} we have $T_n^{**} - B_n^{**} \xrightarrow{d^{**}}_{p^*} \xi_1$, in probability. Thus, the asymptotic distribution of the second-level bootstrap statistic mimics that of the first-level statistic. When B_n^{**} and ζ_1 are such that $B_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} \zeta_1 - \xi_1$, in probability, then Assumption 5 is satisfied. As in Remark 3.6 this setup allows us to cover the pairs bootstrap.

4. Examples Continued

In this section we revisit our three leading examples from Section 2, where we argued that standard boostrap inference is invalid due to the presence of bias. In this section we show how to apply our general theory in each example. Again, we refer to Appendix B for detailed derivations.

4.1. Inference after Model Averaging

Fixed regressor bootstrap. Extending the arguments in Section 2.1, we obtain the following result.

Lemma 4.1. Under regularity conditions stated in Appendix B.1, Assumptions 1 and 2 are satisfied with $(\xi_1, \xi_2)' \sim N(0, V)$, where $V := (v_{ij}), i, j = 1, 2$, is positive definite and continuous in ω, σ^2, and $\Sigma_{WW} := \text{plim}\, S_{WW}$.

By Lemma 4.1, the conditions of Theorem 3.1 hold with $G(u) = \Phi(u/v_{11})$ and $F(u) = \Phi(u/v_d)$, where $v_d^2 = v_{11} + v_{22} - 2v_{12} > 0$. Then Theorem 3.1 implies that the standard bootstrap p-value satisfies $\hat{p}_n \to_d \Phi(m\Phi^{-1}(U_{[0,1]}))$ with $m^2 := v_d^2/v^2$. Because ω is known and σ^2, Σ_{WW} are easily estimated, a consistent estimator $\hat{m}_n \to_p m$ is available, and the plug-in approach in Corollary 3.2 can be implemented by considering the modified p-value, $\tilde{p}_n = \Phi(\hat{m}_n^{-1}\Phi^{-1}(\hat{p}_n))$. Inspection of the proofs shows that our modified bootstrap approach is asymptotically valid whether δ is fixed or local-to-zero. In the former case, B_n is $O_p(n^{1/2})$ rather than $O_p(1)$, implying that B_n diverges in probability and $\tilde{\beta}_n$ is not even consistent for β. Despite this, the modified bootstrap p-value is asymptotically valid.

Alternatively, we can implement the double bootstrap as in Section 3.2.2. Specifically, let

$$y^{**} = x\hat{\beta}_n^* + Z\hat{\delta}_n^* + \varepsilon^{**},$$

where $\varepsilon^{**}|\{D_n, D_n^*\} \sim N(0, \hat{\sigma}_n^{*2} I_n)$, $(\hat{\beta}_n^*, \hat{\delta}_n^{*\prime}, \hat{\sigma}_n^{*2})$ is the OLS estimator obtained from the full model estimated on the first-level bootstrap data, and $D_n^* = \{y^*, W\}$. The double bootstrap statistic is $T_n^{**} := n^{1/2}(\tilde{\beta}_n^{**} - \hat{\beta}_n^*)$, where $\tilde{\beta}_n^{**} := \sum_{m=1}^M \omega_m \tilde{\beta}_{m,n}^{**}$ with $\tilde{\beta}_{m,n}^{**} := S_{xx.Z_m}^{-1} S_{xy^{**}.Z_m}$ defined as the double bootstrap OLS estimator from the mth model. The double bootstrap modified p-value is then $\tilde{p}_n = P^*(\hat{p}_n^* \le \hat{p}_n)$ with $\hat{p}_n^* = P^{**}(T_n^{**} \le T_n^*)$.

Lemma 4.2. Under the conditions of Lemma 4.1, Assumption 3 holds with $\hat{B}_n^* := Q_n n^{1/2}\hat{\delta}_n^*$.

Lemma 4.2 shows that Assumption 3 is verified in this example. The asymptotic validity of the double bootstrap modified p-value now follows from Lemmas 4.1 and 4.2 and Theorem 3.2.

Pairs bootstrap. For the pairs bootstrap we verify the high-level conditions in Section 3.5. To simplify the discussion we consider the case with scalar z_t in (2.1) and where we "average" over only one model ($M = 1$), which is the simplest model in which z_t is omitted from the regression. That is, we estimate β by regression of y on x, that is, $\tilde{\beta}_n := S_{xx}^{-1} S_{xy}$. In this special case, $T_n - B_n \to_d N(0, v^2)$ with $v^2 := \sigma^2 \Sigma_{xx}^{-1}$ and $B_n := S_{xx}^{-1} S_{xz} n^{1/2}\delta$.

Lemma 4.3. Under regularity conditions stated in Appendix B.1, it holds that $T_n^* - \hat{B}_n \xrightarrow{d^*}_p N(0, v^2 + \kappa^2)$, where $\hat{B}_n := S_{xx}^{-1} S_{xz} n^{1/2}\hat{\delta}_n$ and $\kappa^2 := d_r(\delta)'\Sigma_r d_r(\delta)$ with $d_r(\delta) := \delta(\Sigma_{xx}^{-1}, -\Sigma_{xx}^{-2}\Sigma_{xz})'$.

Notice that, in contrast to the FRB, the asymptotic variance of T_n^* fails to replicate that of T_n because of the term $\kappa^2 > 0$. This implies that the methodology developed in Theorem 3.1 and its corollaries no longer applies. Instead, we can apply the theory of Section 3.5. In particular, Lemma 4.3 shows that Assumption 4(i) holds in this case with $\zeta_1 \sim N(0, v^2 + \kappa^2)$. Lemma 4.3 also shows that $\hat{B}_n$ is the same for the pairs bootstrap and the FRB, such that Lemma 4.1 shows that Assumptions 1 and 2(ii) are verified. This implies that Theorem 3.4 holds for this example. Using similar arguments, it can be shown that Assumption 5 also holds for this example, which implies that the double bootstrap p-values are asymptotically uniformly distributed.

Under local alternatives of the form $\beta_0 = \bar{\beta} + an^{-1/2}$, where $\bar{\beta}$ is the value under the null (Section 3.3), the asymptotic local power function for the modified p-value is given by $\Phi(\Phi^{-1}(\alpha) - a/v_d)$; see Theorem 3.3. It is not difficult to verify that this is the same power function as that obtained from a test based directly on $\hat{\beta}_n$ from the full model (2.1).

4.2. Ridge Regression

To complete the example in Section 2.2, we can proceed as in the previous example.

Lemma 4.4. Under the null hypothesis and the regularity conditions stated in Appendix B.2, Assumptions 1 and 2 are satisfied with $(\xi_1, \xi_2)' \sim N(0, V)$, where $V := (v_{ij}), i, j = 1, 2$, is positive definite and continuous in c_0, σ^2, and Σ_{xx}.

As in Section 4.1, Lemma 4.4 and Theorem 3.1 imply that the standard bootstrap p-value satisfies $\hat{p}_n \to_d \Phi(m\Phi^{-1}(U_{[0,1]}))$, where we now have $m^2 = (g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g)^{-1}g'\Sigma_{xx}^{-1}g$. Note that this result holds irrespectively of θ being fixed or local to zero. Thus, the bootstrap is invalid unless $c_0 = 0$ which implies $m = 1$. For the plug-in method, a simple consistent estimator of m is given by $\hat{m}_n^2 := (g'\tilde{S}_{xx}^{-1}S_{xx}\tilde{S}_{xx}^{-1}g)^{-1}g'S_{xx}^{-1}g$, and inference based on the plug-in modified p-value $\tilde{p}_n = \Phi(\hat{m}_n^{-1}\Phi^{-1}(\hat{p}_n))$ is then asymptotically valid by Corollary 3.2.

To implement the double bootstrap method, we can draw the double bootstrap sample $\{y_t^{**}, x_t^{**}; t = 1, \ldots, n\}$ as iid from $\{y_t^*, x_t^*; t = 1, \ldots, n\}$. Accordingly, the second-level bootstrap ridge estimator is $\tilde{\theta}_n^{**} := \tilde{S}_{x^{**}x^{**}}^{-1}S_{x^{**}y^{**}}$ with associated test statistic $T_n^{**} := n^{1/2}g'(\tilde{\theta}_n^{**} - \hat{\theta}_n^*)$, which is centered at the first-level bootstrap OLS estimator, $\hat{\theta}_n^*$. It is straightforward to show that, without additional conditions, Assumption 3 holds.

Lemma 4.5. Under the conditions of Lemma 4.4, Assumption 3 holds with $\hat{B}_n^* := -c_n n^{-1/2}g'\tilde{S}_{x^*x^*}^{-1}\hat{\theta}_n^*$.

Validity of the double bootstrap modified p-value $\tilde{p}_n = P^*(\hat{p}_n^* \leq \hat{p}_n)$ now follows by application of Theorem 3.2.

4.3. *Nonparametric Regression*

Again, we complete the example in Section 2.3 by proceeding as in the previous examples.

Lemma 4.6. Under regularity conditions stated in Appendix B.3, Assumptions 1 and 2 are satisfied with $(\xi_1, \xi_2)' \sim N(0, V)$, where $V := (v_{ij}), i, j = 1, 2$, is positive definite and continuous in σ^2 and the kernel function.

As before, Lemma 4.6 and Theorem 3.1 imply that the standard bootstrap p-value satisfies $\hat{p}_n \to_d \Phi(m\Phi^{-1}(U_{[0,1]}))$, where we now have $m^2 := 4 + (\int K^2(u)du)^{-1}(\int(\int K(s-u)K(s)ds)^2du - 4\int K(u)\int K(u-s)K(s)dsdu)$. Thus, in this example, m need not be estimated because it is observed once K is chosen. Therefore, valid inference is feasible with the modified p-value $\tilde{p}_n = H(\hat{p}_n) = \Phi(m^{-1}\Phi^{-1}(\hat{p}_n))$; see Corollary 3.1.

We can also apply a double bootstrap modification. Let $y_t^{**} = \hat{\beta}_h^*(x_t) + \varepsilon_t^{**}, t = 1, \ldots, n$, where $\varepsilon_t^{**}|\{D_n, D_n^*\} \sim \text{iid}N(0, \hat{\sigma}_n^{*2})$ with $D_n^* := \{y_t^*; t = 1, \ldots, n\}$ and $\hat{\sigma}_n^{*2}$ denoting the residual variance from the first-level bootstrap data. The double bootstrap analogue of T_n is $T_n^{**} := (nh)^{1/2}(\hat{\beta}_h^{**}(x) - \hat{\beta}_h^*(x))$, where $\hat{\beta}_h^{**}(x) := (nh)^{-1}\sum_{t=1}^n k_t y_t^{**}$. This can be decomposed as $T_n^{**} = \xi_{1,n}^{**} + \hat{B}_n^*$, where $\hat{B}_n^* := (nh)^{1/2}((nh)^{-1}\sum_{t=1}^n k_t\hat{\beta}_h^*(x_t) - \hat{\beta}_h^*(x))$. Unfortunately, although $\xi_{1,n}^{**}$ satisfies Assumption 3(i), $\hat{B}_n^*$ does not satisfy Assumption 3(ii). The reason is that $\hat{B}_n^* - \hat{B}_n = \xi_{2,n}^* + \hat{B}_{2,n} - \hat{B}_n$, where $\xi_{2,n}^*$ satisfies Assumption 3(ii), but $\hat{B}_{2,n} := (nh)^{-1}\sum_{t=1}^n k_t\hat{B}_n(x_t)$ is a smoothed version of $\hat{B}_n$ (evaluated at x_t) and although $\hat{B}_{2,n} - \hat{B}_n$ is mean zero it is not $o_p(1)$. However, $\hat{B}_{2,n} - \hat{B}_n$ is observed, so this is easily corrected by defining $\bar{T}_n^{**} := T_n^{**} - (\hat{B}_{2,n} - \hat{B}_n)$. Then we have the following result.

Lemma 4.7. Under the conditions of Lemma 4.6, Assumption 3 holds with T_n^{**} and $\hat{B}_n^*$ replaced by $\bar{T}_n^{**}$ and $\bar{B}_n^* := \hat{B}_n^* - (\hat{B}_{2,n} - \hat{B}_n)$, respectively.

The validity of the double bootstrap modified p-value $\tilde{p}_n := P^*(\hat{p}_n^* \leq \hat{p}_n)$, where $\hat{p}_n^* := P^{**}(\bar{T}_n^{**} \leq T_n^*)$, follows from Lemma 4.7 and Theorem 3.2. This in turn implies that confidence intervals based on the double bootstrap are asymptotically valid; see also Remark 3.4. We note that Hall and Horowitz (2013) also proposed, without theory, a version of their calibration method based on the double bootstrap. Our double bootstrap-based method for confidence intervals corresponds to their steps 1–5, and where we need a correction they have instead a step 6 in which they average over a grid of x.

Finally, under local alternatives of the form $\beta_0(x) = \bar{\beta} + an^{-2/5}$, where $\bar{\beta}$ is the value under the null (Section 3.3), the asymptotic local power function for the modified p-value is given by $\Phi(\Phi^{-1}(\alpha) - a/v_d)$; see Theorem 3.3. Alternatively, we could consider a "bias-free" test based on undersmoothing; that is, using a bandwidth h satisfying $nh^5 \to 0$ such that $B_n \to 0$ and inference can be based on quantiles of $\xi_1 \sim N(0, v_{11}^2)$. In contrast to our procedure, however, such a test has only trivial power against $\bar{\beta} + an^{-2/5}$ because $(nh)^{1/2}an^{-2/5} \to 0$.

5. Concluding Remarks

In this article, we have shown that in statistical problems involving bias terms that cannot be estimated, the bootstrap can be modified to provide asymptotically valid inference. Intuitively, the main idea is the following: in some important cases, the bootstrap can be used to "debias" a statistic whose bias is non-negligible, but when doing so additional "noise" is injected. This additional noise does not vanish because the bias cannot be consistently estimated, but it can be handled either by a "plug-in" method or by an additional (i.e., double) bootstrap layer. Specifically, our solution is simple and involves (i) focusing on the bootstrap p-value; (ii) estimating its asymptotic distribution; (iii) mapping the original (invalid) p-value into a new (valid) p-value using the prepivoting approach. These steps are easy to implement in practice and we provide sufficient conditions for asymptotic validity of the associated tests and confidence intervals.

Our results can be generalized in several directions. For instance, there is a growing literature where inference on a parameter of interest is combined with some auxiliary information in the form of a bound on the bias of the estimator in question. These bounds appear, for example, in Oster (2019) and Li and Müller (2021). It is of interest to investigate how our analysis can be extended in order to incorporate such bounds. Other possible extensions include non-ergodic problems, large-dimensional models, and multivariate estimators or statistics. All these extensions are left for future research.

Supplementary Materials

The supplemental material contains two appendices. Appendix A describes in detail the conditions and results of the article under the special case of asymptotically Gaussian statistics. Appendix B contains details and proofs for the three examples in the article, as well as two additional examples. Additional references are included at the end of the supplement.

Acknowledgments

We thank Federico Bandi, Matias Cattaneo, Christian Gourieroux, Philip Heiler, Michael Jansson, Anders Kock, Damian Kozbur, Marcelo Moreira, David Preinerstorfer, Mikkel Sølvsten, Luke Taylor, Michael Wolf, and participants at the AiE Conference in Honor of Joon Y. Park, 2022 Conference on Econometrics and Business Analytics (CEBA), 2023 Conference on Robust Econometric Methods in Financial Econometrics, 2022 EC^2 conference, 2nd "High Voltage Econometrics" workshop, 2023 IAAE Conference, 3rd Italian Congress of Econometrics and Empirical Economics, 3rd Italian Meeting on Probability and Mathematical Statistics, 19th School of Time Series and Econometrics, Brazilian Statistical Association, 2023 Société Canadienne de Sciences Économiques, 2022 Virtual Time Series Seminars, as well as seminar participants at Aarhus University, CREST, FGV - Rio, FGV - São Paulo, Ludwig Maximilian University of Munich, Queen Mary University, Singapore Management University, UFRGS, University of the Balearic Islands, University of Oxford, University of Pittsburgh, University of Victoria, York University, for useful comments and feedback.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Cavaliere thanks the Italian Ministry of University and Research (PRIN 2017 Grant 2017TA7TYC) for financial support. Gonçalves thanks the Natural Sciences and Engineering Research Council of Canada for financial support (NSERC grant number RGPIN-2021-02663). Nielsen thanks the Danish National Research Foundation for financial support (DNRF Chair grant number DNRF154).

ORCID

Giuseppe Cavaliere https://orcid.org/0000-0002-2856-0005
Sílvia Gonçalves https://orcid.org/0009-0009-9064-3109
Morten Ørregaard Nielsen https://orcid.org/0000-0002-1337-9844
Edoardo Zanelli https://orcid.org/0009-0001-1575-5719

References

Beran, R. (1987), "Prepivoting to Reduce Level Error in Confidence Sets," *Biometrika*, 74, 457–468. [2,6,7]

——— (1988), "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687–97. [2,6,7]

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018), "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 113, 767–779. [4]

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326. [4]

Cattaneo, M. D., and Jansson, M. (2018), "Kernel-based Semiparametric Estimators: Small Bandwidth Asymptotics and Bootstrap Consistency," *Econometrica*, 86, 955–995. [1]

——— (2022), "Average Density Estimators: Efficiency and Bootstrap Consistency," *Econometric Theory*, 38, 1140–1174. [1]

Cattaneo, M. D., Jansson, M., and Ma, X. (2019), "Two-Step Estimation and Inference with Possibly Many Included Covariates," *Review of Economic Studies*, 86, 1095–1122. [1]

Cavaliere, G., and Georgiev, I. (2020), "Inference Under Random Limit Bootstrap Measures," *Econometrica*, 88, 2547–2574. [2,5]

Chatterjee, A., and Lahiri, S. N. (2010), "Asymptotic Properties of the Residual Bootstrap for Lasso Estimators," *Proceedings of the American Mathematical Society*, 138, 4497–4509. [3]

Chatterjee, A., and Lahiri, S. N. (2011), "Bootstrapping Lasso Estimators," *Journal of the American Statistical Association*, 106, 608–625. [3]

Efron, B. (1983), "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation," *Journal of American Statistical Association*, 78, 316–331. [2]

Eubank, R. L., and Speckman, P. L. (1993), "Confidence Bands in Nonparametric Regression," *Journal of the American Statistical Association*, 88, 1287–1301. [4]

Fu, W., and Knight, K. (2000), "Asymptotics for Lasso-Type Estimators," *Annals of Statistics*, 28, 1356–1378. [3]

Hall, P. (1986), "On the Bootstrap and Confidence Intervals," *Annals of Statistics*, 14, 1431–1452. [2]

——— (1992), *The Bootstrap and Edgeworth Expansion*, Berlin: Springer-Verlag. [1]

Hall, P., Horowitz, J. (2013), "A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions," *Annals of Statistics*, 41, 1892–1921. [4,10]

Hansen, B. E. (2007), "Least Squares Model Averaging," *Econometrica*, 75, 1175–1189. [2,3]

Härdle, W., and Bowman, A. W. (1988), "Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands," *Journal of the American Statistical Association*, 83, 102–110. [4]

Härdle, W., and Marron, J. S. (1991), "Bootstrap Simultaneous Error Bars for Nonparametric Regression," *Annals of Statistics*, 19, 778–796. [4]

Hjort, N., and Claeskens, G. (2003), "Frequentist Model Average Estimators," *Journal of the American Statistical Association*, 98, 879–899. [3]

Horowitz, J. L. (2001), "The Bootstrap," in *Handbook of Econometrics* (Vol. 5), eds. J. J. Heckman, and E. Leamer, pp. 3159–3228, Elsevier: Amsterdam. [1,5]

Hounyo, U., and Lahiri, K. (2023), "Estimating the Variance of a Combined Forecast: Bootstrap-based Approach," *Journal of Econometrics*, 232, 445–468. [3]

Li, C., and Muller, U. (2021), "Linear Regression with Many Controls of Limited Explanatory Power," *Quantitative Economics*, 12, 405–442. [10]

Liu, C.-A. (2015), "Distribution Theory of the Least Squares Averaging Estimator," *Journal of Econometrics*, 186, 142–159. [3]

Oster, E. (2019), "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37, 187–204. [3,10]

Shao, X., and Politis, D. N. (2013), "Fixed b Subsampling and the Block Bootstrap: Improved Confidence Sets based on p-value Calibration," *Journal of the Royal Statistical Society*, Series B, 75, 161–184. [2]