Bootstrap Inference in the Presence of Bias*

Giuseppe Cavaliere† — University of Bologna, Italy and Exeter Business School, UK

Sílvia Gonçalves — McGill University, Canada

Morten Ørregaard Nielsen — Aarhus University, Denmark

July 28, 2022

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 (1987, 1988), 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 applications: (i) a simple location model for i.i.d. data, possibly with infinite variance; (ii) regression models with omitted controls; (iii) inference on a target parameter based on model averaging; (iv) ridge-type regularized estimators; and (v) dynamic panel data models.

Keywords: Asymptotic bias, bootstrap, incidental parameter bias, model averaging, prepivoting, regularization.

1 Introduction

Consider an estimator $\hat{\theta}_n$ of a population parameter θ, which is the object of inference. Classic (first-order) asymptotic inference on θ is based on large-sample results of the form

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

where $g(n) \to \infty$ and, in the standard case of asymptotically Gaussian estimators, $g(n) = n^{1/2}$ and $\xi_1 \sim N(0, \sigma^2)$. However, it is frequently the case that $\hat{\theta}_n$, rather than being centered around

*We thank Federico Bandi, Matias Cattaneo, Christian Gourieroux, Michael Jansson, Anders Kock, Damian Kozbur, David Preinerstorfer, Mikkel Sølvsten, Michael Wolf, and participants at the 2nd 'High Voltage Econometrics' workshop (U Zurich, 2021), the 3rd Italian Congress of Econometrics and Empirical Economics, the 19th School of Time Series and Econometrics, Brazilian Statistical Association, 3rd Italian Meeting on Probability and Mathematical Statistics (U Bologna), as well as seminar participants at CREST, Queen Mary University, University of the Balearic Islands, and York University for useful comments and feedback. Cavaliere thanks the the Italian Ministry of University and Research (PRIN 2017 Grant 2017TA7TYC) for financial support. Nielsen thanks the Danish National Research Foundation for financial support (DNRF Chair grant number DNRF154).

†Corresponding author. Address: Dipartimento di Scienze Economiche, Piazza Scaravilli 2, Bologna, Italy; email: `giuseppe.cavaliere@unibo.it`.

θ as n grows, is asymptotically biased. Specifically, it may hold that

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

where B is an unknown finite constant. Despite (1.1) implying that $\hat{\theta}_n$ is consistent for θ, inference based on the quantiles of ξ_1 is invalid, unless $B = 0$.

Cases of asymptotically biased estimators arise in all areas of econometrics and statistics. For instance, they can appear in regression models with omitted controls (Li and Müller, 2021), model averaging estimation (Liu, 2015), shrinkage and lasso-type estimators (Knight and Fu, 2001), incidental parameter problems (Hahn and Kuersteiner, 2002), nonparametric inference (Calonico, Cattaneo, and Titiunik, 2014, Calonico, Cattaneo, and Farrell, 2018, Cattaneo and Jansson, 2018), among others. In some cases B can be estimated, but that is not always the case, and asymptotic inference based on $\hat{\theta}_n$ is generally infeasible.

The bootstrap, which is well known to deliver asymptotic refinements over first-order asymptotic approximations as well as bias corrections (see Hall, 1992, and Horowitz, 2001), cannot in general be applied to solve the asymptotic bias problem in cases where a consistent estimator of the asymptotic bias does not exist. Specifically, consider a bootstrap procedure generating a bootstrap analogue of $\hat{\theta}_n$, say $\hat{\theta}^*_n$, from an auxiliary bootstrap sample with bootstrap true value $\hat{\theta}_n$. Ideally, as in e.g. Cattaneo and Jansson (2018), the bootstrap analogue of T_n, say T^*_n, should mimic the asymptotic distribution in (1.1); that is,

$$T^*_n := g(n)(\hat{\theta}^*_n - \hat{\theta}_n) \xrightarrow{d^*}_p B + \xi_1, \tag{1.2}$$

where '$\xrightarrow{d^*}_p$' denotes weak convergence in probability, meaning that $P^*(T^*_n \le u) \to_p G(u - B)$, with $G(u)$ being the cumulative distribution funtion [cdf] of ξ_1 and P^* the probability measure induced by the bootstrap (i.e., conditional on the original data). However, the result in (1.2) cannot usually be achieved. In particular, there are at least two main reasons for failure of the bootstrap.

The first reason is when the asymptotic bias cannot be replicated in the bootstrap world. In this case, the large sample distribution of T^*_n converges (in probability) to the distribution function of ξ_1, rather than $B+\xi_1$. Therefore, the bootstrap fails to mimic the bias. For example, this happens when the bootstrap sample is based on $m = o(n)$ observations, as for classic 'm out of n' bootstrap schemes (Arcones and Giné, 1989) or subsampling algorithms (Politis, Romano, and Wolf, 1999).

The second reason is when the asymptotic bias term in the bootstrap world includes a random (additive) component, say ξ_2. If this is the case, the bootstrap distribution is random in the limit and hence cannot mimic the asymptotic distribution given in (1.1). Specifically, rather than (1.2), it may hold that

$$P^*(T^*_n \le u) \to G(u - B - \xi_2) \text{ (weakly)}.$$

In general, this case occurs because in the bootstrap world, population parameters (which are non-random in the original world) are replaced by bootstrap analogues which are random in the bootstrap world, even in the limit. As a consequence, the limit bootstrap measure is random, thus invalidating the bootstrap as a means for estimating the limit distribution (1.1).

In both scenarios, the bootstrap does not mimic the asymptotic distribution of the original statistic. In particular, in both cases the distribution of the bootstrap p-value is not asymptotically uniform, and the bootstrap cannot deliver hypothesis tests (or confidence intervals) with the desired null rejection probability (or coverage probability).

In this paper we show that, in this non-standard case where inference involves an asymptotically biased estimator that cannot be bias-corrected, valid inference can successfully be restored by proper implementation of the bootstrap. This is done by focusing on the properties of the bootstrap p-value, say $\hat{p}_n$, rather than on the bootstrap as a means of estimating limiting distributions, which is infesible due to the asymptotic bias. In particular, we show that such implementations lead to bootstrap inferences which 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, we can make use of the prepivoting approach of Beran (1987, 1988), which — as we will show in this paper — allows to restore bootstrap validity by tranforming the original bootstrap p-value using the unconditional distribution H. Specifically, provided $\hat{p}_n$ has asymptotic cdf H, then under mild conditions, $H(\hat{p}_n)$ is uniformly distributed in large samples. Hence, the mapping $\hat{p}_n \longmapsto H(\hat{p}_n)$ is the key transformation to solve our inference problem. Interestingly, while Beran (1987, 1988) proposed this approach to obtain asymptotic refinements for the bootstrap, as far as we are aware it has never been applied to asymptotically biased estimators before.

Although the asymptotic distribution of the bootstrap p-value, H, does not depend on B, it can still be unknown, as it may depend (often in a non-trivial manner) on a (possibly infinite-dimensional) vector of nuisance parameters, even in the limit. Therefore, we propose different alternative approaches to estimating H.

First, we show that if a consistent estimator of the nuisance parameters exist, then H can be estimated using a simple plug-in approach. That is, if $H = H_\gamma$, where γ is the vector of nuisance parameters, and a consistent estimator $\hat{\gamma}_n$ of γ is available, then the mapping $\hat{p}_n \longmapsto H_{\hat{\gamma}_n}(\hat{p}_n)$ will deliver asymptotically uniform p-values.

Second, we show that if estimation of γ is difficult (e.g., γ does not have a closed form expression), estimation of H can be done by using a 'double bootstrap' scheme, where the underlying idea is that inference on functionals of the bootstrap data can be built up by bootstrapping the bootstrap sample itself. The double bootstrap (Efron, 1983; Hall, 1986; Beran, 1987, 1988) has been employed in the statistics literature to improve the coverage of confidence sets and of bias correction methods; see Chang and Hall (2015) and the references therein. Our context is different: we show that the double bootstrap allows estimation of the (unknown and non-uniform) distribution of the bootstrap p-value 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 restricted to asymptotically Gaussian estimators and statistics. For instance, common assumptions such as finite variance or stationarity are not required. Moreover, our methods 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 non-uniform, 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 'fixed-b' bandwidth (Kiefer and Vogelsang, 2005) may deliver non-uniformly distributed p-values which, however, can be estimated.

To illustrate the the practical relevance of our results and to show how to implement them in applied problems, we initially discuss the main ideas by focusing on inference based on a model averaging estimator obtained as a weighted average of least squares estimates from two simple regression models; see, e.g., Hansen (2007). We show that even in this simple setting, model averaging induces a bias component when non-trivial weight is given to a misspecified model. The bias component does not vanish asymptotically and it may even diverge. While standard bootstrap implementations fail in this case, we show how our proposed bootstrap approach leads to valid inference.

In Section 4 we consider five well-known problems involving estimators that feature an asymptotic bias term.

First, we consider a simple location model without the assumption of finite variance. This case is non-standard, as the limit theory is no longer Gaussian and estimators converge at an unknown rate. We show that although classic bootstraps for infinite variance data fail, our approach delivers valid inference.

Second, we consider a regression model with omitted controls, as discussed in, e.g., Li and Müller (2021). We show that, although the omission of significant regressors induces an asymptotic bias, our method – when applied to the biased estimator – delivers valid inferences.

Third, we revisit the model averaging example in a much more general setting. In the context of Hjort and Claeskens (2003) and Liu (2015), who assume that the non-targeted parameters are local-to-zero, we show that a simple bootstrap algorithm which uses the full model to generate the data, allows for post-model averaging inference unaffected by the asymptotic bias. This complements Hounyo and Lahiri (2021), who show that the bootstrap provides a consistent estimator of the asymptotic variance.

Fourth, we consider estimation of a vector of regression parameters through regularization; in particular, by using a ridge estimator. The ridge estimator can be asymptotically biased when the regularization parameter does not offset the magnitude of the regression coefficients; see Knight and Fu (2001). Standard bootstrap implementations are invalid in this case (Chatterjee and Lahiri, 2010, 2011). Nevertheless, we show that despite the presence of bias, we can construct asymptotically valid (double) bootstrap tests of significance based on the ridge estimator.

Fifth, we consider inference based on panel data estimators subject to incidental parameter bias. In the context of a general nonlinear panel data model, the cross sectional pairs bootstrap (Dhaene and Jochmans, 2015) cannot replicate the bias and consequently is invalid. We show that a prepivoting approach based on a plug-in estimator of the bias is valid (see also Higgins and Jochmans, 2022).

Structure of the paper

The paper is organized as follows. In Section 2 we preview our main results using a simple model averaging estimator. Section 3 contains our general results, which are specialized to the case of asymptotically Gaussian statistics in Section 3.4. Section 4 contains the five applications of our results and Section 5 concludes. All proofs are collected in the appendix.

Notation

Throughout this paper, 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 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 right-continuous generalized inverse, i.e., $F^{-1}(u) := \sup\{v \in \mathbb{R} : F(v) \leq u\}$, $u \in \mathbb{R}$. Unless specified otherwise, all limits are for $n \to \infty$. The space of càdlàg functions $\mathbb{R} \to \mathbb{R}$ (equipped with its Skorokhod J_1-topology; see Kallenberg, 1997, Appendix A2) is denoted by $D(\mathbb{R})$. 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) bootstrap sequence, say Y_n^*, we use $Y_n^* \xrightarrow{p^*}_p 0$, or, equivalently, or $Y_n^* \xrightarrow{p^*} 0$, in probability, to mean that, for any $\epsilon > 0$, $P^*(|Y_n^*| > \epsilon) \to_p 0$, where P^* denotes the probability measure conditionally on the original data D_n. An equivalent notation is $Y_n^* = o_{p^*}(1)$ (where we omit the qualification "in probability" for brevity). We also write $Y_n^* = O_{p^*}(1)$ to mean that $P^*(|Y_n^*| > M) \to_p 0$ for some large enough M. Similarly, for a double 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, and we write $Y_n^{**} = O_{p^{**}}(1)$ to mean that there exists $M < \infty$ such that $P^{**}(|Y_n^{**}| > M) \xrightarrow{p^*}_p 0$.

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 $P^*(Y_n^* \leq u) - G(u) \to_p 0$; or, setting $X_n(u) := P^*(Y_n^* \leq u) - G(u)$ (a function of D_n), that $X_n(u) \to_p 0$. Similarly, for a double bootstrap sequence Y_n^{**}, we use $Y_n^{**} \xrightarrow{d^{**}}_{p^*} \xi$, in probability, to mean that $X_n^*(u) := P^{**}(Y_n^{**} \leq u) - G(u)$ satisfies $X_n^*(u) \xrightarrow{p^*}_p 0$ for all continuity points u of G.

2 Model averaging and preview of main results

To motivate our framework, consider the following simple, illustrative example. Suppose we observe a sample $D_n := \{y_t, x_t, z_t; t = 1, \ldots, n\}$ generated by the linear regression model

$$y_t = \theta x_t + \delta z_t + \varepsilon_t, \quad \varepsilon_t \sim \text{ i.i.d.} N(0,1), \tag{2.1}$$

where, for simplicity, we assume a Gaussian distribution with a known variance $\sigma^2 = 1$. The regressors are univariate, non-stochastic, and linearly independent for all n. The goal is inference on θ, such as testing the one-sided null hypothesis $\mathsf{H}_0 : \theta = \theta_0$ versus $\mathsf{H}_1 : \theta < \theta_0$, using model averaging.

The researcher estimates θ by averaging two estimators; $\hat{\theta}_{long}$ is the OLS estimator from the regression of y_t on $(x_t, z_t)'$, and $\hat{\theta}_{short}$ is the OLS estimator from the regression of y_t on x_t, i.e. omitting z_t. That is,

$$\hat{\theta}_n := \omega\hat{\theta}_{long} + (1-\omega)\hat{\theta}_{short}, \tag{2.2}$$

where ω denotes a fixed combination weight. Specifically, $\hat{\theta}_{long} = S_{xx.z}^{-1}S_{xy.z}$ and $\hat{\theta}_{short} = S_{xx}^{-1}S_{xy}$. The estimator $\hat{\theta}_n$ is biased unless $\omega = 1$ or $\delta = 0$. Consider now the test statistic

$$T_n := \sqrt{n}(\hat{\theta}_n - \theta_0) = \omega\sqrt{n}(\hat{\theta}_{long} - \theta_0) + (1-\omega)\sqrt{n}(\hat{\theta}_{short} - \theta_0).$$

The researcher wishes to conduct inference on θ using the statistic T_n.

Since $\hat{\theta}_{long} = \theta_0 + S_{xx.z}^{-1}S_{x\varepsilon.z}$ and $\hat{\theta}_{short} = \theta_0 + S_{xx}^{-1}S_{xz}\delta + S_{xx}^{-1}S_{x\varepsilon}$, it holds that

$$T_n = \underbrace{(1-\omega)S_{xx}^{-1}S_{xz}n^{1/2}\delta}_{=:B_n} + \underbrace{(1-\omega)S_{xx}^{-1}S_{x\varepsilon}n^{1/2} + \omega S_{xx.z}^{-1}S_{x\varepsilon.z}n^{1/2}}_{\sim N(0,v_{1,n}^2)} \sim B_n + \xi_{1,n}$$

with $\xi_{1,n} \sim N(0, v_{1,n}^2)$; $v_{1,n}^2$ is derived in Appendix B.1. Equivalently, $T_n - B_n \sim \xi_{1,n}$. Hence, because of the presence of B_n, inference based on quantiles of the $N(0, v_{1,n}^2)$ distribution is invalid. Note that, under the local-to-zero assumption $\delta = cn^{-1/2}$, the term B_n is $O(1)$, while if δ is fixed B_n diverges. The results in this section hold under both scenarios.

The fact that the distribution of T_n is not centered at zero also complicates bootstrap inference, as we now show. Consider a bootstrap sample $D_n^* := \{y_t^*; t = 1, \ldots, n\}$ generated using estimates from the long model,

$$y_t^* = \hat{\theta}_{long}x_t + \hat{\delta}_{long}z_t + \varepsilon_t^*, \quad \varepsilon_t^* \sim \text{ i.i.d.} N(0, 1).$$

Then, the bootstrap analogue of the model averaging estimator estimator (2.2) is given by

$$\hat{\theta}_n^* := \omega\hat{\theta}_{long}^* + (1-\omega)\hat{\theta}_{short}^*,$$

with $\hat{\theta}_{long}^*$ and $\hat{\theta}_{short}^*$ denoting the OLS estimators of θ from the long and the short regressions based on D_n^*, respectively. The bootstrap analogue of T_n is

$$T_n^* := \sqrt{n}(\hat{\theta}^* - \hat{\theta}_{long}) = \omega\sqrt{n}(\hat{\theta}_{long}^* - \hat{\theta}_{long}) + (\omega - 1)\sqrt{n}(\hat{\theta}_{short}^* - \hat{\theta}_{long}),$$

where $\hat{\theta}_{long}$ is the bootstrap true value since we generate D_n^* using $\hat{\theta}_{long}$. It is straightforward to see that, conditionally on D_n,

$$T_n^* = \underbrace{(1-\omega)S_{xx}^{-1}S_{xz}n^{1/2}\hat{\delta}_{long}}_{=:\hat{B}_n} + \underbrace{(1-\omega)S_{xx}^{-1}S_{x\varepsilon^*}n^{1/2} + \omega S_{xx.z}^{-1}S_{x\varepsilon^*.z}n^{1/2}}_{\sim N(0,v_{1,n}^2)} \sim N(\hat{B}_n, v_{1,n}^2).$$

The presence of the (random) term $\hat{B}_n$ implies that

$$T_n^*|D_n \sim N(\hat{B}_n, v_{1,n}^2)|\hat{B}_n,$$

or, equivalently, $(T_n^* - \hat{B}_n)|D_n \sim \xi_{1,n}$.

Consider a bootstrap test that rejects H_0 at nominal level α whenever $\hat{p}_n \leq \alpha$, where $\hat{p}_n$ is the standard bootstrap p-value

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

To control the asymptotic size of the test, a crucial condition is that $\hat{p}_n$ converges in distribution to $U_{[0,1]}$. As it turns out, this condition is not satisfied in this context. The reason is that $\hat{B}_n - B_n$ is non-zero and, in general, does not vanish (unless $\omega = 1$ or $S_{xz} = 0$), because

$$\hat{B}_n - B_n = (1-\omega)S_{xx}^{-1}S_{xz}(n^{1/2}(\hat{\delta}_{long} - \delta)) \sim \xi_{2,n},$$

with $\xi_{2,n} \sim N\left(0, v_{2,n}^2\right)$; $v_{2,n}^2$ is given in Appendix B.1. Hence,

$$\hat{p}_n = P^*(T_n^* \leq T_n) = P^*\left(\frac{T_n^* - \hat{B}_n}{v_{1,n}} \leq \frac{T_n - \hat{B}_n}{v_{1,n}}\right) = \Phi\left(\frac{T_n - \hat{B}_n}{v_{1,n}}\right),$$

because $T_n^* - \hat{B}_n|D_n \sim N(0, v_{1,n}^2)$. Under our assumptions,

$$T_n - \hat{B}_n = (T_n - B_n) - (\hat{B}_n - B_n) \sim \xi_{1,n} - \xi_{2,n} \sim N\left(0, v_{d,n}^2\right),$$

where $v_{d,n}^2 > 0$ is given in Appendix B.1 and $N\left(0, v_{d,n}^2\right)$ equals $v_{d,n}\Phi^{-1}(U_{[0,1]})$ in distribution. Hence, setting $m_n := v_{d,n}/v_{1,n}$, we have that

$$\hat{p}_n = \Phi\left(\frac{v_{d,n}\Phi^{-1}(U_{[0,1]})}{v_{1,n}}\right) \sim \Phi(m_n\Phi^{-1}(U_{[0,1]})) \neq U_{[0,1]}, \tag{2.3}$$

unless $m_n = 1$. The failure of the standard bootstrap p-value $\hat{p}_n$ is due to the fact that the bias term B_n cannot be replicated by the bootstrap bias $\hat{B}_n$ except when $\hat{B}_n - B_n$ has a degenerate distribution, which occurs only when $\omega = 1$ (i.e., when $\hat{\theta}_n = \hat{\theta}_{long}$). Notice also that failure is not a finite-sample problem: under standard conditions on the regressors, m_n will not tend to one as $n \to \infty$.

Remark 2.1 *It is shown in Appendix B.1 that* $m_n = (1-(1-\omega^2)\hat{\rho}_{xz}^2)^{-1/2}$, *where* $\hat{\rho}_{xz} = S_{xz}(S_{xx}S_{zz})^{-1/2}$ *is the sample correlation between x and z. Clearly, m_n equals one if (i) $\omega = 1$, i.e. all weight is assigned to the large model, or (ii) $\hat{\rho}_{xz} = 0$, i.e. the two regressors are orthogonal.*

Bootstrap failure in the model averaging context is not new; see Hjort and Claeskens (2003), who pointed out the bootstrap invalidity when ω is random, and Section 4.3 below. As our example illustrates, the bootstrap fails even when weights are fixed rather than random. However, in our context, it is indeed possible to solve the bootstrap invalidity problem, as we explain next.

The idea is as follows. Because $\hat{p}_n \sim \Phi(m_n\Phi^{-1}(U_{[0,1]}))$, its distribution function is

$$\begin{aligned} H_n(u) := P(\hat{p}_n \leq u) &= P(\Phi(m_n\Phi^{-1}(U_{[0,1]})) \leq u) \\ &= P(U_{[0,1]} \leq \Phi(m_n^{-1}\Phi^{-1}(u))) = \Phi(m_n^{-1}\Phi^{-1}(u)), \end{aligned}$$

implying that

$$H_n(\hat{p}_n) = \Phi(m_n^{-1}\Phi^{-1}(\hat{p}_n)) \sim U_{[0,1]}.$$

Thus, a test that rejects H_0 whenever $H_n(\hat{p}_n) \leq \alpha$ is *exact*. The mapping of $\hat{p}_n$ into $H_n(\hat{p}_n)$ amounts to the prepivoting approach of Beran (1987), who introduced this idea as a way of obtaining asymptotic refinements for the classic bootstrap. Our application of prepivoting is new: we use it to transform an invalid bootstrap p-value into a valid modified bootstrap p-value.

Prepivoting in this example can be implemented analytically because we observe the finite-sample distribution of $\hat{p}_n$. This follows from the Gaussianity assumption and the fact that the only parameter that enters the distribution of the test statistic is m_n, which is known in this example. In more realistic applications, this is no longer the case, but we can approximate the distribution of $\hat{p}_n$ with a (double) bootstrap estimator or a 'plug-in' estimator of its limiting distribution.

To explain the double bootstrap approach, suppose we do not know $H_n(u)$ and therefore we cannot compute $H_n(\hat{p}_n)$ analytically. A bootstrap prepivoting approach consists of using the bootstrap to obtain a bootstrap estimator of $H_n(u)$, say $\hat{H}_n(u)$, and then using this approximation to compute $\hat{H}_n(\hat{p}_n)$. Given that $H_n(u) = P(\hat{p}_n \leq u)$, we can define the estimator

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

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, which can be implemented as follows. Let $D_n^{**} := \{y_t^{**}; t = 1, \ldots, n\}$ and

$$y_t^{**} = \hat{\theta}_{long}^* x_t + \hat{\delta}_{long}^* z_t + \varepsilon_t^{**}, \quad \varepsilon_t^{**} \sim \text{ i.i.d.} N(0, 1),$$

where $\hat{\theta}_{long}^*$ and $\hat{\delta}_{long}^*$ are obtained from D_n^* (and not only D_n), as defined above. Importantly, the ε_t^{**}'s are independent of both D_n and D_n^*. The 'second-level' bootstrap estimator of θ_0 is

$$\hat{\theta}_n^{**} := \omega \hat{\theta}_{long}^{**} + (1 - \omega) \hat{\theta}_{short}^{**},$$

where $\hat{\theta}_{long}^{**}$ and $\hat{\theta}_{short}^{**}$ are the OLS estimators of θ from the long and the short regressions based on D_n^{**}, respectively. This implies that

$$T_n^{**} := \sqrt{n}(\hat{\theta}_n^{**} - \hat{\theta}_{long}^*) = \omega \sqrt{n}(\hat{\theta}_{long}^{**} - \hat{\theta}_{long}^*) + (\omega - 1)\sqrt{n}(\hat{\theta}_{short}^{**} - \hat{\theta}_{long}^*).$$

As for the standard bootstrap, it is straightforward to see that, conditionally on D_n^* and D_n,

$$T_n^{**} = \underbrace{(1 - \omega) S_{xx}^{-1} S_{xz} n^{1/2} \hat{\delta}_{long}^*}_{=: \hat{B}_n^*} + \underbrace{(1 - \omega) S_{xx}^{-1} S_{x\varepsilon^{**}} n^{1/2} + \omega S_{xx.z}^{-1} S_{x\varepsilon^{**}.z} n^{1/2}}_{\sim N(0, v_{1,n}^2)} \sim N(\hat{B}_n^*, v_{1,n}^2).$$

Here, the presence of the random term $\hat{B}_n^*$ implies that the second-level bootstrap statistic is conditionally biased,

$$T_n^{**} | \{D_n^*, D_n\} \sim N(\hat{B}_n^*, v_{1,n}^2) | \hat{B}_n^*,$$

or, equivalently,

$$T_n^{**} - \hat{B}_n^* | \{D_n^*, D_n\} \sim \xi_{1,n}.$$

Hence, and crucially, $T_n^{**} - \hat{B}_n^*$, $T_n^* - \hat{B}_n$, and $T_n - B$ all share the same distribution function, given by the cdf of $\xi_{1,n}$.

With this notation, the second-level bootstrap p-value is defined as

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

where P^{**} is the bootstrap probability measure induced by resampling from D_n^* (making $\hat{p}_n^*$ a function of D_n^*). We have that

$$\hat{p}_n^* = P^{**}\left(T_n^{**} \le T_n^*\right) = P^{**}\left(\frac{T_n^{**} - \hat{B}_n^*}{v_{1,n}} \le \frac{T_n^* - \hat{B}_n^*}{v_{1,n}}\right) = \Phi\left(\frac{T_n^* - \hat{B}_n^*}{v_{1,n}}\right) \tag{2.4}$$

because $T_n^{**} - \hat{B}_n^* | \{D_n^*, D_n\} \sim N\left(0, v_{1,n}^2\right)$. Using the fact that

$$T_n^* - \hat{B}_n^* = (T_n^* - \hat{B}_n) - (\hat{B}_n^* - \hat{B}_n) \sim \xi_{1,n} - \xi_{2,n} \sim N(0, v_{d,n}^2) \sim v_{d,n}\Phi^{-1}(U_{[0,1]}),$$

we have that

$$\hat{p}_n^* \sim \Phi(m_n \Phi^{-1}(U_{[0,1]})),$$

where m_n is as defined previously. Therefore

$$\hat{H}_n(u) = P^*(\hat{p}_n^* \le u) = P^*(\Phi(m_n\Phi^{-1}(U_{[0,1]})) \le u) = \Phi(m_n^{-1}\Phi^{-1}(u)) = H_n(u).$$

This shows that the bootstrap distribution function of $\hat{p}_n^*$ coincides with the distribution function of $\hat{p}_n$. Thus, in this example, the double bootstrap modified p-value is exactly equal to the analytically modified p-value:

$$\tilde{p}_n := \hat{H}_n(\hat{p}_n) = \Phi(m_n^{-1}\Phi^{-1}(\hat{p}_n)) = H_n(\hat{p}_n) \sim U_{[0,1]}.$$

Remark 2.2 *Notice that the results given in this section are exact and do not require $n \to \infty$. This follows from the assumption that the distribution of the errors is $N(0,1)$ and known. Should this not be the case, then the arguments hold as $n \to \infty$ under mild conditions; see Section 4.3.*

Remark 2.3 *The pairs bootstrap is another well-known alternative to the fixed regressor bootstrap discussed above. It can be shown that the pairs bootstrap fails to estimate both the mean and variance of the distribution of T_n, but nonetheless the double bootstrap and our methodology deliver valid p-values in this setup as well. See Section 4.3 for details.*

In Table 1 we present the results of a small Monte Carlo simulation experiment to ilustrate the above results numerically. We generate the data from the regression model (2.1) with sample sizes $n = 10, 20, 40$. The regressors are multivariate normally distributed with unit variances and correlation 0.7, and the errors are either standard normal, t_3, or χ_1^2 distributed. The true values are $\theta = 1$ and $\delta = 1$ (although the results are invariant to the true values because we use the unrestricted estimates to construct the bootstrap samples). The estimator is (2.2) with $\omega = 1/2$. We consider two bootstrap schemes. The first is the parametric bootstrap scheme, where $\varepsilon_t^* \sim$ i.i.d.$N(0,1)$, which is denoted as "par." in the table. The second is the non-parametric bootstrap scheme, where ε_t^* is re-sampled independently from the centered residuals $\{\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\}_{t=1,\ldots,n}$, which is denoted as "non-par." Results are based on 10,000 Monte Carlo simulations and $B = 999$ bootstrap replications.

The simulation outcomes in Table 1 clearly ilustrate our theoretical results. The standard bootstrap p-value, $\hat{p}_n$, is much larger than the nominal level of the test. The plug-in modified

Table 1: Simulated rejection frequencies (%) of bootstrap tests

		5% nominal level						10% nominal level					
		par.			non-par.			par.			non-par.		
dist.	n	$\hat{p}_n$	$\tilde{p}_{n,p}$	$\tilde{p}_{n,d}$	$\hat{p}_n$	$\tilde{p}_{n,p}$	$\tilde{p}_{n,d}$	$\hat{p}_n$	$\tilde{p}_{n,p}$	$\tilde{p}_{n,d}$	$\hat{p}_n$	$\tilde{p}_{n,p}$	$\tilde{p}_{n,d}$
N	10	10.1	5.6	5.2	15.2	10.8	5.8	15.9	10.7	10.0	20.2	15.5	10.2
	20	10.3	5.1	5.1	12.0	7.2	4.8	16.1	10.7	10.6	17.7	12.3	9.9
	40	9.9	5.0	5.0	10.5	5.8	4.9	15.6	10.1	10.1	15.9	11.0	9.6
t_3	10	7.6	4.1	4.1	15.2	10.0	5.6	11.7	7.5	7.5	20.4	15.2	9.9
	20	8.1	4.3	4.3	12.2	6.9	4.9	13.3	8.3	8.5	18.1	12.3	9.8
	40	7.6	4.0	4.0	10.8	5.9	5.2	12.9	8.0	8.0	17.1	11.1	9.9
χ^2_1	10	7.6	4.1	4.1	16.2	11.2	6.4	13.0	8.5	8.5	21.9	16.2	10.8
	20	8.1	4.3	4.3	12.9	7.6	5.3	13.3	8.7	8.7	19.1	13.2	10.3
	40	7.6	5.3	5.2	10.9	5.8	4.8	12.9	9.7	9.7	17.3	12.2	9.8

Notes: $\hat{p}_n$ denotes the standard bootstrap; $\tilde{p}_{n,p}$ and $\tilde{p}_{n,d}$ denote the modified bootstrap using the plug-in and the double bootstrap methods, respectively. The parametric bootstrap scheme, where $\varepsilon_t^* \sim$ i.i.d.$N(0,1)$, is denoted as "par." and the non-parametric bootstrap scheme, where ε_t^* is re-sampled independently from the centered residuals $\{\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\}_{t=1,\dots,n}$, is denoted as "non-par." The ε_t's are i.i.d. draws from (standardized) N, t_3, and χ^2_1 distributions. Results are based on 10,000 Monte Carlo simulations and $B = 999$.

p-value, $\tilde{p}_{n,p}$, is close to the nominal level for the parametric bootstrap scheme, but is still oversized for the non-parametric scheme with the smaller sample sizes. Finally, the double bootstrap modified p-value, $\tilde{p}_{n,d}$, is nearly perfectly sized throughout the table.

3 General results

3.1 Framework and invalidity of the standard bootstrap

The general framework is as follows. We have a statistic $T_n = T(D_n)$, 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. 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 $T_n^* = T(D_n^*)$ be a bootstrap analogue of T_n and let $\hat{L}_n(u) := P^*(T_n^* \leq u)$, $u \in \mathbb{R}$, denote its distribution function, conditionally on the original data. The bootstrap p-value is defined as

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

First-order asymptotic validity of $\hat{p}_n$ requires that $\hat{p}_n$ converges in distribution to a standard uniform distribution, i.e. 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 which 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 $G_\gamma(u) = P(\xi_1 \leq u)$ is a continuous cdf.*

When B_n converges to a nonzero constant B, Assumption 1 can be rewritten as

$$T_n \xrightarrow{d} B + \xi_1,$$

where B represents a shift of the center of the distribution G_γ. If ξ_1 is centered at zero and T_n is a normalized version of a (scalar) parameter estimator, i.e. $T_n = g(n)(\hat{\theta}_n - \theta)$, then we can think of B as the asymptotic bias of $\hat{\theta}_n$. Although we do not require ξ_1 to have zero mean and 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'.

The example of Section 2 is such that $T_n - B_n \sim \xi_{1,n}$, where $\xi_{1,n}$ is Gaussian for any n. More generally, in Assumption 1 we cover any statistic T_n that is not necessarily Gaussian (even asymptotically) and whose limiting distribution is G_γ only after we subtract the sequence B_n. We index the limiting distribution G_γ by a parameter γ to allow for the possibility that $T_n - B_n$ is not an asymptotic pivot.

Inference based on the asymptotic distribution of T_n requires estimating B_n and γ. Alternatively, we can use the bootstrap to bypass estimation of B_n and γ and directly compute a bootstrap p-value that relies on T_n^* and T_n alone, i.e. we consider $\hat{p}_n := P^*(T_n^* \leq T_n)$. 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 *(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 such that $F_\phi(u) = P(\xi_1 - \xi_2 \leq u)$ is continuous.

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_γ as $T_n - B_n$.[1] 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. In both cases, $\hat{B}_n$ is a function of the original data, i.e. it is D_n-measurable.

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 the latter 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 $F_\phi(u) = G_\gamma(u)$. Consequently,

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

[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, i.e. $P(\xi_1 \leq u) = P(\xi_1^* \leq u)$. We do not make this distinction because we care only about distributional results, but it should be kept in mind.

where the last distributional equality holds by $F_\phi = G_\gamma$ 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

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

under Assumptions 1 and 2(ii). When ξ_2 is nondegenerate, $F_\phi \neq G_\gamma$, implying that $\hat{p}_n = G_\gamma(T_n - \hat{B}_n) + o_p(1)$ is not asymptotically distributed as a standard uniform random variable. This result is summarized in the following theorem, which generalizes the result given in (2.3) for the example in Section 2.

THEOREM 3.1 *Suppose Assumptions 1 and 2 hold. Then* $\hat{p}_n \to_d G_\gamma(F_\phi^{-1}(U_{[0,1]}))$.

REMARK 3.1 *The value of $\hat{B}_n$ in Assumption 2(i) depends on the chosen bootstrap algorithm. For instance, it is possible that $\hat{B}_n \to_p 0$. Examples are given in Sections 4.1 and 4.5. If this is the case, then Assumption 2(ii) is still satisfied; specifically, it holds with $\xi_2 = -B$ a.s., which implies that*

$$F_\phi(u) := P(\xi_1 - \xi_2 \leq u) = P(\xi_1 \leq u - B) = G_\gamma(u - B),$$

and from Theorem 3.1 it follows that the bootstrap p-value satisfies

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

Notice that this distribution is not uniform unless $B = 0$. 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 in particular the case where $T_n = g(n)(\hat{\theta}_n - \theta)$ and T_n^* is its bootstrap analogue with (conditional) distribution function $\hat{L}_n(u)$. Interest is in constructing a right-sided confidence set for the unknown parameter θ. Using a standard percentile method, a confidence set at the nominal confidence level $1-\alpha \in (0,1)$ obtained by test inversion is of the form (see 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,

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

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) measure is random in the limit whenever ξ_2 is non-degenerate. Specifically, assume for simplicity that $B_n \to_p B$. Recall that $\hat{L}_n(u) := P^*(T_n^* \leq u)$, $u \in \mathbb{R}$, and let $\hat{G}_{\gamma,n}(u) := P^*(T_n^* - \hat{B}_n \leq u)$. It then holds that*

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

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

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). The basic idea is that we modify $\hat{p}_n$ by applying an empirical monotone mapping $\tilde{p}_n = H_n(\hat{p}_n)$ which makes the modified p-value $\tilde{p}_n$ asymptotically standard uniform. Contrary to Beran (1987, 1988), who proposed this approach as a way of providing asymptotic refinements for the bootstrap, here we show how to use this approach in order 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 $T_n - \hat{B}_n$. Thus, we modify the test statistic rather than modifying the way we compute the bootstrap p-value.

3.2 Prepivoting

Let $H_n(u) := P(\hat{p}_n \leq u)$ denote the distribution function of the bootstrap p-value $\hat{p}_n$, where $u \in [0, 1]$. Theorem 3.1 implies that

$$\begin{aligned} H_n(u) &\to P(G_\gamma(F_\phi^{-1}(U_{[0,1]})) \leq u) = P(F_\phi^{-1}(U_{[0,1]}) \leq G_\gamma^{-1}(u)) \\ &= P(U_{[0,1]} \leq F_\phi(G_\gamma^{-1}(u))) = F_\phi(G_\gamma^{-1}(u)) =: H(u), \end{aligned}$$

uniformly over $u \in [0, 1]$, given the continuity of G_γ and F_ϕ. Although H is not the uniform distribution, unless $G_\gamma = F_\phi$, the following corollary to Theorem 3.1 holds.

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

Therefore, the mapping of $\hat{p}_n$ into $H_n(\hat{p}_n)$ transforms $\hat{p}_n$ into a new p-value, $H_n(\hat{p}_n)$, whose asymptotic distribution is the standard uniform distribution on $[0, 1]$. Inference based on $H_n(\hat{p}_n)$ is generally infeasible, because we do not observe $H_n(u)$. However, if we can replace $H_n(u)$ 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 $H_n(u)$.

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, it suffices to replace the nominal confidence level $1 - \alpha$ by $\hat{H}_n^{-1}(1 - \alpha)$, i.e. to consider the set*

$$\widetilde{CI}_n^{1-\alpha} := [\hat{\theta}_n - g(n)^{-1}\hat{q}_n(\hat{H}_n^{-1}(1 - \alpha)), +\infty).$$

[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 $T_n^*|D_n \xrightarrow{w} (B + \xi_1^* + \xi_2)|\xi_2$.

Then, as in Remark 3.2, we obtain

$$P(\theta \in \widetilde{CI}_n^{1-\alpha}) = P(\hat{p}_n \leq \hat{H}_n^{-1}(1-\alpha)) + o(1) = P(\hat{H}_n(\hat{p}_n) \leq 1-\alpha) + o(1) \to 1-\alpha,$$

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

Remark 3.5 *The result in Corollary 3.1 can also be used if interest is in right-tailed or two-tailed tests. Consider first a right-tailed p-value, say $\hat{p}_{n,r} := P^*(T_n^* > T_n) = 1 - \hat{L}_n(T_n) = 1 - \hat{p}_n$ (notice that, because the conditional cdf of T_n^* is continuous in the limit, this p-value is asymptotically equivalent to $P^*(T_n^* \geq T_n)$). It holds that $H_{n,r}(u) := P(\hat{p}_{n,r} \leq u) = P(\hat{p}_n \geq 1-u) = 1 - H_n(1-u) + o_p(1)$ uniformly in u. Therefore, the modified right-tailed p-value, $\tilde{p}_{n,r} := H_{n,r}(\hat{p}_{n,r})$, satisfies*

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

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

3.2.1 Plug-in approach

In view of Theorem 3.1, a simple approach to estimating $H_n(u) = P(\hat{p}_n \leq u)$ is to use

$$\hat{H}_n(u) = F_{\hat{\phi}_n}(G_{\hat{\gamma}_n}^{-1}(u)),$$

where $\hat{\gamma}_n$ and $\hat{\phi}_n$ denote consistent estimators of γ and ϕ, respectively. This leads to a plug-in modified p-value defined as

$$\tilde{p}_n = F_{\hat{\phi}_n}(G_{\hat{\gamma}_n}^{-1}(\hat{p}_n)).$$

By consistency of $\hat{\gamma}_n$ and $\hat{\phi}_n$ and continuity of F_ϕ and G_γ, it follows immediately that

$$\tilde{p}_n = F_\phi(G_\gamma^{-1}(\hat{p}_n)) + o_p(1) \xrightarrow{d} F_\phi(G_\gamma^{-1}(G_\gamma(F_\phi^{-1}(U_{[0,1]})))) = U_{[0,1]}.$$

This result is summarized next.

Corollary 3.2 *Under Assumptions 1 and 2, if $(\hat{\gamma}_n, \hat{\phi}_n) \to_p (\gamma, \phi)$ then $\tilde{p}_n = F_{\hat{\phi}_n}(G_{\hat{\gamma}_n}^{-1}(\hat{p}_n)) \to_d U_{[0,1]}$.*

The plug-in approach relies on consistent estimators of the asymptotic distributions G_γ and F_ϕ, but does not require estimating the 'bias term' B_n. When estimating γ and ϕ is simple, this approach is attractive since it does not require any double resampling. However, computation of γ and ϕ is case specific and may be cumbersome in practice. An automatic approach is to use the bootstrap to estimate $H_n(u)$, as we describe next.

3.2.2 Double bootstrap

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

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

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); let T_n^{**} denote the bootstrap version of T_n^* computed on D_n^{**}, i.e. $T_n^{**} = T(D_n^{**})$. With this notation, the second-level bootstrap p-value is defined as

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

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 *For ξ_1 and ξ_2 as defined in Assumptions 1 and 2, (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_γ 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, i.e. when $\xi_2 = 0$ a.s., implying $F_\phi = G_\gamma$. When $F_\phi \neq G_\gamma$, $\hat{B}_n^*$ is not a consistent estimator of $\hat{B}_n$. However, under Assumption 3,

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

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 $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$.*

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 asymptotic uniformly distributed random variable. This result holds under Assumptions 1, 2, and 3, independently of whether $G_\gamma = F_\phi$ or not. When $G_\gamma = F_\phi$ 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 asymptotic 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 $G_\gamma \neq F_\phi$ 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 paper.

3.3 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 their 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, we can show that $\tilde{p}_n$ 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 $\tilde{p}_n := P^*(\hat{p}_n^* \leq \hat{p}_n)$, where

$$\hat{p}_n^* := P^{**}(T_n^{**} \leq T_n^*) = P^{**}(T_n^{**} - \hat{B}_n^* \leq T_n^* - \hat{B}_n^*) = G_\gamma(T_n^* - \hat{B}_n^*) + o_{p^*}(1),$$

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

$$\hat{p}_n := P^*(T_n^* \leq T_n) = P^*(T_n^* - \hat{B}_n \leq T_n - \hat{B}_n) = G_\gamma(T_n - \hat{B}_n) + o_p(1).$$

It follows that

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

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

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

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, $\hat{p}_n = P^*(T_n^* \leq T_n)$. Note that we do not require $\xi_2 = 0$, i.e. $\hat{B}_n - B_n$ and $\hat{B}_n^* - \hat{B}_n$ do not need to converge to zero.

Corollary 3.3 is useful when $\hat{B}_n$ and $\hat{B}_n^*$ are easy to compute, e.g. when they are available analytically as functions of D_n and D_n^*, respectively, as it avoids implementing a double bootstrap. When deriving $\hat{B}_n$ and $\hat{B}_n^*$ explicitly is cumbersome or impossible, 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 $\hat{B}_n^*$.

3.4 Special case: T_n is asymptotically Gaussian

In this section, we specialize Assumptions 1, 2, and 3 to the case where $T_n = \sqrt{n}(\hat{\theta}_n - \theta)$ is a normalized parameter estimator whose limiting distribution is normal. For simplicity, we focus on θ being a scalar but the results generalize to the multivariate context. We consider the following special case of Assumption 1.

Assumption 1′ *It holds that* $T_n - B_n \to_d N(0, v^2)$, *where* $v^2 > 0$.

Assumption 1′ covers statistics T_n based on asymptotically biased estimators: when $B_n \to_p B$, we have $T_n \to_d N\left(B, v^2\right)$, in which case B is the asymptotic bias of $\hat{\theta}_n$. More generally, we can interpret B_n as a bias term that approximates $E(\sqrt{n}(\hat{\theta}_n - \theta))$ although B_n does not need to have a limit. Note that Assumption 1′ obtains from Assumption 1 when we let $\xi_1 \sim N(0, v^2)$ and $G_\gamma(u) = \Phi(u/v)$, in which case $\gamma = v$.

Let D_n^* denote a bootstrap sample from D_n and let $\hat{\theta}_n^*$ be a bootstrap version of $\hat{\theta}_n$. The bootstrap analogue of T_n is $T_n^* = \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$.

ASSUMPTION 2′ *It holds that (i)* $T_n^* - \hat{B}_n \xrightarrow{d^*}_p N(0, v^2)$*, and (ii)*

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} \xrightarrow{d} N(0, V), \quad V := (v_{ij}), \quad i, j = 1, 2,$$

where $v_d^2 := v_{11} + v_{22} - 2v_{12} > 0$ *with* $v_{11} := v^2 > 0$.

Assumption 2′(i) requires the bootstrap statistic $T_n^* - \hat{B}_n$ to mimic the asymptotic distribution of $T_n - B_n$, as in Assumption 2(i). However, and contrary to Assumption 2(i), here this limiting distribution is the zero mean Gaussian distribution (i.e. $G_\gamma(u) = \Phi(u/v)$), which means that we can interpret $\hat{B}_n$ as a bootstrap bias correction term, i.e. $\hat{B}_n = E^*(\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n))$. Assumption 2′(ii) assumes that $\hat{B}_n - B_n$ is also asymptotically distributed as a zero mean Gaussian random variable (jointly with $T_n - B_n$).[3] An implication of this assumption is that

$$T_n - \hat{B}_n = (T_n - B_n) - (\hat{B}_n - B_n) \xrightarrow{d} N(0, v_d^2), \tag{3.1}$$

where $v_d^2 := v_{11} + v_{22} - 2v_{12}$. We do not require V to be positive definite; for instance, $v_{22} = 0$ whenever $\hat{B}_n - B_n = o_p(1)$, and in fact V can be rank deficient even when $v_{22} > 0$, one example being that studied in Section 2. However, we do impose the restriction that $v_d^2 > 0$. This ensures that the limiting distribution function of $T_n - \hat{B}_n$, given by $F_\phi(u) = \Phi(u/v_d)$, is well-defined and continuous, as assumed in Assumption 2(ii). Note that we can let $\phi = V$ in this case, or simply set $\phi = v_d$.

Let $\hat{p}_n$ denote the standard bootstrap p-value as defined in Section 3. We then obtain the following result.

COROLLARY 3.4 *Under Assumptions 1′ and 2′,* $\hat{p}_n \to_d \Phi(m\Phi^{-1}(U_{[0,1]}))$*, where* $m^2 := v_d^2/v^2$.

Corollary 3.4 follows immediately from Theorem 3.1 when we let $G_\gamma(u) = \Phi(u/v)$ and $F_\phi(u) = \Phi(u/v_d)$. It shows that the asymptotic distribution of $\hat{p}_n$ is uniform only when $m = 1$, or equivalently when $v_d^2 = v^2$. In this case, the difference $\hat{B}_n - B_n$ is $o_p(1)$. When $v_d^2 \neq v^2$, $\hat{B}_n - B_n$ is random even in the limit, implying that the limiting bootstrap distribution function of T_n^* is conditionally random. Although random limit bootstrap measures do not necessarily invalidate bootstrap inference, as discussed by Cavaliere and Georgiev (2020), this is not the case here. However, we can solve the problem of bootstrap invalidity by applying the prepivoting approach or by modifying the test statistic from T_n to $T_n - \hat{B}_n$.

To describe the prepivoting approach, note that the limiting distribution of $\hat{p}_n$ is given by

$$H(u) := P(\hat{p}_n \leq u) = \Phi(m^{-1}\Phi^{-1}(u)).$$

Hence, a plug-in approach amounts to estimating $m^2 := v_d^2/v^2$, where v^2 and v_d^2 are defined in Assumption 2′. Suppose that $\hat{v}_n^2$ and $\hat{v}_{d,n}^2$ are consistent estimators of v^2 and v_d^2 (i.e., assume that

[3] In terms of Assumption 2, Assumption 2′ corresponds to the case where the vector $\xi = (\xi_1, \xi_2)'$ is a multivariate normal distribution with covariance matrix V.

$(\hat{v}_n^2, \hat{v}_{d,n}^2) \to_p (v^2, v_d^2))$ and let $\hat{m}_n^2 := \hat{v}_{d,n}^2/\hat{v}_n^2$. Then, by Corollary 3.2, it immediately follows that

$$\tilde{p}_n = \Phi(\hat{m}_n^{-1}\Phi^{-1}(\hat{p}_n)) \xrightarrow{d} U_{[0,1]}$$

under Assumptions $1'$ and $2'$. For brevity, we do not formalize this result here.

To describe the double bootstrap modified p-value, $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$, when applied to the special case where T_n satisfies Assumption $1'$, we now introduce Assumption $3'$.

Assumption $3'$ *Let $T_n^{**} = \sqrt{n}(\hat{\theta}_n^{**} - \hat{\theta}_n^*)$ and suppose that (i) $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} N(0, v^2)$, in probability, and (ii) $T_n^* - \hat{B}_n^* \xrightarrow{d^*}_p N(0, v_d^2)$, where v_d^2 is as defined in Assumption $2'$(ii).*

Under Assumption $3'$(i), the double bootstrap distribution of $T_n^{**} - \hat{B}_n^*$ mimics the distribution of $T_n^* - \hat{B}_n$, where the double bootstrap bias term $\hat{B}_n^* = E^{**}(\sqrt{n}(\hat{\theta}_n^{**} - \hat{\theta}_n^*))$ is asymptotically centered at $\hat{B}_n$ under Assumption $3'$(ii). When $v_d^2 \neq v^2$, the double bootstrap bias is not a consistent estimator of $\hat{B}_n$, but this is not needed for the asymptotic validity of the modified double bootstrap p-value $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$ defined in Section 3.

By application of Theorem 3.2, $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$ under Assumptions $1'$, $2'$, and $3'$. We can also provide a result analogous to Corollary 3.3 under these assumptions. In this case, if closed-form expressions for $\hat{B}_n$ and $\hat{B}_n^*$ are not available, we can approximate these bootstrap expectations by Monte Carlo simulations and then compute $P^*(T_n^* - \hat{B}_n^* \leq T_n - \hat{B}_n)$ as a valid bootstrap p-value. Note, however, that this approach is computationally as intensive as the prepivoting approach based on $\tilde{p}_n$ since it too requires two layers of resampling.

Remark 3.6 *Contrary to Beran (1987, 1988), in our context the first level of prepivoting, e.g. by the double bootstrap, is used to obtain an asymptotically valid bootstrap p-value. Therefore, inference based on $\tilde{p}_n$ does not necessarily provide an asymptotic refinement over inference based on an asymptotic approach that does not require the bootstrap. Nevertheless, the Monte Carlo results in Table 1 seem to suggest an asymptotic refinement for the double bootstrap, at least for the non-parametric bootstrap scheme.*

In the special case where the bias term B_n is of sufficiently small order, the arguments in Beran (1987, 1988) apply, and an asymptotic refinement can be obtained. We also conjecture that, in the general case, an asymptotic refinement could be obtained by further iterating the bootstrap.

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, when the first-level bootstrap does not match the variance of the original statistic.

Assumption 4 *Assumption 2 holds with part (i) replaced by (i) $T_n^* - \hat{B}_n \xrightarrow{d^*}_p \zeta_1$, where $J_\gamma(u) = P(\zeta_1 \leq u)$ is a continuous cdf.*

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 non-random (with respect to P^*, i.e., conditional on the original data) term, $\hat{B}_n$, such that $T_n^* - \hat{B}_n$ has the same asymptotic distribution as $T_n - B_n$.

Remark 3.7 *An important case that is covered by Assumption 4 is when there is a random (with respect to P^*, i.e. depending on the bootstrap data) term, say 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 $B_n^* - \hat{B}_n \xrightarrow{p^*}_p 0$. However, letting B_n^* and ζ_1 be such that $B_n^* \xrightarrow{d^*}_p \zeta_1 - \xi_1$, then Assumption 4 covers this case.*

As illustrated by Remark 3.7, Assumption 4 generalizes Assumption 2 to allow for bootstrap methods where the 'centering term' B_n^* depends on the bootstrap data. An important example that falls into this case is the pairs bootstrap, which we study in more detail in Section 4.3.

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_γ replaced by J_γ, and hence omitted.

Theorem 3.3 *If Assumptions 1 and 4 hold then $\hat{p}_n \to_d J_\gamma(F_\phi^{-1}(U_{[0,1]}))$.*

Clearly, a plug-in approach based on G_γ as described in Section 3.2.1 would be invalid because $G_\gamma \neq J_\gamma$ in general. However, it follows straigthforwardly by the same arguments as applied in Section 3.2.1 that a plug-in approach based on J_γ 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

$$\begin{aligned}\hat{p}_n^* &= P^{**}(T_n^{**} \leq T_n^*) = P^{**}(T_n^{**} - \hat{B}_n^* \leq T_n^* - \hat{B}_n^*)\\ &= J_\gamma(T_n^* - \hat{B}_n^*) + o_p^*(1) \xrightarrow{d^*}_p J_\gamma(\xi_1 - \xi_2) = J_\gamma(F_\phi^{-1}(U_{[[0,1]}))\end{aligned}$$

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, $\tilde{p}_n := \hat{H}_n(\hat{p}_n) = P^*(\hat{p}_n^* \leq \hat{p}_n)$, is asymptotically valid, which is stated next. The proof is essentially identical to that of Theorem 3.2 and hence omitted.

Theorem 3.4 *Under Assumptions 1, 4, and 5, it holds that $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$.*

Remark 3.8 *Consider again the case with a random bootstrap centering term in Remark 3.7, where $B_n^* \xrightarrow{d^*}_p \zeta_1 - \xi_1$ 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 $B_n^{**} \xrightarrow{d^{**}}_{p^*} \zeta_1 - \xi_1$, in probability, then Assumption 5 is satisfied. As in Remark 3.7 this setup allows us to cover the pairs bootstrap.*

Remark 3.9 *In the case of asymptotically Gaussian statistics discussed in Section 3.4, Assumptions 4 and 5 simplify straightforwardly. In Assumption 2′(i) we would assume that* $T_n^* - \hat{B}_n \xrightarrow{d^*}_p N(0, v_s^2)$ *and in Assumption 3′(i) that* $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} N(0, v_s^2)$, *in probability, for some* $v_s^2 > 0$, *while the rest of Assumptions 1′–3′ are unchanged. The results of Section 3.4 continue to apply under these more general conditions, replacing* $G_\gamma(u) = \Phi(u/v)$ *with* $J_\gamma(u) = \Phi(u/v_s)$ *and consequently defining* $m := v_d^2/v_s^2$.

4 Applications

4.1 Inference under heavy tails

Setup. We consider a simple location model with heavy-tailed data; thus demonstrating how our analysis applies to a non-Gaussian asymptotic framework.

Consider a sample of n i.i.d. random variables $\{y_t\}$. Interest is in inference on θ in the location model

$$y_t = \theta + \varepsilon_t, \quad E(\varepsilon_t) = 0,$$

when the ε_t's follow a symmetric, stable random variable $S(\alpha)$ with tail index $\alpha \in (1, 2)$ and the location parameter is local to zero, i.e. $\theta = n^{1/\alpha - 1}c$.[4] Under these assumptions, $E(|\varepsilon_t|^{\alpha+\delta}) = +\infty$ for any $\delta \geq 0$; in particular, ε_t has infinite variance. Notice that θ is local of order $n^{1/\alpha-1}$ rather than the usual $n^{-1/2}$ because of the slower convergence rate of the OLS-type estimator when the variance of ε_t is infinite. We consider a special case of the estimator analyzed in Section 2, viz.

$$\hat{\theta}_n := \omega\hat{\theta}_{long} + (1 - \omega)\hat{\theta}_{short},$$

where $\hat{\theta}_{long} := \bar{y}_n$, $\hat{\theta}_{short} := 0$, and ω denotes a fixed combination weight. It holds that

$$T_n := n^{1-1/\alpha}(\hat{\theta}_n - \theta) = (\omega - 1)c + \omega n^{1-1/\alpha}\bar{\varepsilon}_n \sim B + \omega S(\alpha), \tag{4.1}$$

where $B := (\omega - 1)c$; equivalently, $T_n - B \sim \xi_1 := \omega S(\alpha)$. Hence, Assumption 1 is satisfied with $G_\gamma(u) = P(\omega S(\alpha) \leq u) = \Psi_\alpha(\omega^{-1}u)$, where $\Psi_\alpha(u) := P(S(\alpha) \leq u)$ is continuous. Inference based on quantiles of ξ_1 is invalid because it misses the term B.

Bootstrap. It is well known that the standard bootstrap fails to be valid under infinite variance (Knight, 1989). The 'm out of n' bootstrap (see Politis et al., 1999, and the references therein) is an attractive option, but fails to mimic the non-centrality parameter B, see Remark 4.1 below. Instead, we consider the parametric bootstrap of Cornea-Madeira and Davidson (2015), which only requires a consistent estimator $\hat{\alpha}_n$ of the tail index α, assumed to lie in a compact set. The bootstrap sample is generated as

$$y_t^* = \hat{\theta}_{long} + \varepsilon_t^*, \quad \varepsilon_t^* \sim \text{ i.i.d.} S(\hat{\alpha}_n),$$

and the bootstrap estimator is $\hat{\theta}_n^* := \omega\hat{\theta}_{long}^* = \omega(\hat{\theta}_{long} + \bar{\varepsilon}_n^*)$ with $\bar{\varepsilon}_n^* := n^{-1}\sum_{t=1}^n \varepsilon_t^*$. The

[4] The results in this section can easily be generalized to the case where the ε_t's are not necessarily symmetric and/or are in the domain of attraction of a stable law with index $\alpha \in (0, 1)$, as in Cornea-Madeira and Davidson (2015).

bootstrap analogue of T_n then satisfies

$$T_n^* := n^{1-1/\alpha}(\hat{\theta}_n^* - \hat{\theta}_{long}) = \omega n^{1-1/\alpha}\bar{\varepsilon}_n^* + \hat{B}_n \text{ with } \hat{B}_n := (\omega-1)n^{1-1/\alpha}\hat{\theta}_{long}.$$

Now, $n^{1-1/\alpha}\bar{\varepsilon}_n^* \xrightarrow{d^*}_p S(\alpha)$ by Proposition 1 in Cornea-Madeira and Davidson (2015) and, therefore,

$$T_n^* - \hat{B}_n \xrightarrow{d^*}_p \xi_1 := \omega S(\alpha).$$

This shows that Assumption 2(i) is satisfied in this example. Notice that the bias term in the bootstrap world satisfies, jointly with (4.1),

$$\hat{B}_n - B = (\omega-1)n^{1-1/\alpha}\bar{\varepsilon}_n \sim (\omega-1)S(\alpha) =: \xi_2.$$

Specifically, because both T_n and $\hat{B}_n$ depend on the data through $\bar{\varepsilon}_n$ only, we have that $(\xi_1, \xi_2) \sim (\omega, \omega-1)S(\alpha)$, implying that $\xi_1 - \xi_2 \sim S(\alpha)$. Hence, Assumption 2(ii) is satisfied with $F_\phi(u) = P(S(\alpha) \le u) = \Psi_\alpha(u)$. Since the cdf of $\xi_1 \sim \omega S(\alpha)$ can be written as $G_\gamma(u) = \Psi_\alpha(\omega^{-1}u)$, it follows by Theorem 3.1 that $\hat{p}_n \to_d G_\gamma(F_\phi^{-1}(U_{[[0,1]})) = \Psi_\alpha(\omega^{-1}\Psi_\alpha^{-1}(U_{[0,1]}))$ and, therefore,

$$H_n(u) := P(\hat{p}_n \le u) \to H(u) := P(\Psi_\alpha(\omega^{-1}\Psi_\alpha^{-1}(U_{[[0,1]})) \le u) = \Psi_\alpha(\omega\Psi_\alpha^{-1}(u)),$$

which differs from u unless $\omega = 1$.

Because ω is known and we can estimate α consistently with $\hat{\alpha}_n$, we can estimate $H(u)$ consistently with $\hat{H}_n(u) := \Psi_{\hat{\alpha}_n}(\omega\Psi_{\hat{\alpha}_n}^{-1}(u))$ and obtain a valid plug-in modified p-value,

$$\tilde{p}_n = \hat{H}_n(\hat{p}_n) = \Psi_{\hat{\alpha}_n}(\omega\Psi_{\hat{\alpha}_n}^{-1}(\hat{p}_n)),$$

by application of Corollary 3.2.

Alternatively, we can estimate $H_n(u)$ using the double bootstrap estimator $\hat{H}_n(u) := P^*(\hat{p}_n^{**} \le u)$, where $\hat{p}_n^{**} := P^{**}(T_n^{**} \le T_n^*)$. Specifically, let the double bootstrap sample $\{y_t^{**}\}$ be generated as

$$y_t^{**} = \hat{\theta}_{long}^* + \varepsilon_t^{**}, \quad \varepsilon_t^{**} \sim \text{ i.i.d.} S(\hat{\alpha}_n),$$

and set $\hat{\theta}_n^{**} := \omega\hat{\theta}_{long}^{**} = \omega\hat{\theta}_{long}^* + \omega\bar{\varepsilon}_n^{**}$, where $\bar{\varepsilon}_n^{**} := n^{-1}\sum_{t=1}^n \varepsilon_t^{**}$. The (second-level) bootstrap analogue of T_n^* then satisfies

$$T_n^{**} := n^{1-1/\alpha}(\hat{\theta}_n^{**} - \hat{\theta}_{long}^*) = \omega n^{1-1/\alpha}\bar{\varepsilon}_n^{**} + \hat{B}_n^* \text{ with } \hat{B}_n^* := (\omega-1)n^{1-1/\alpha}\hat{\theta}_{long}^*.$$

Since ε_t^{**} is generated from $S(\hat{\alpha}_n)$, where $\hat{\alpha}_n$ depends only on D_n, the distribution of ε_t^{**}, conditionally on D_n^* and D_n, is the same as the distribution of ε_t^*, conditionally on D_n. This implies that

$$n^{1-1/\alpha}\bar{\varepsilon}_n^{**} \xrightarrow{d^{**}}_{p*} S(\alpha),$$

in probability, by Proposition 1 of Cornea-Madeira and Davidson (2015). Therefore,

$$T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p*} \xi_1 = \omega S(\alpha),$$

in probability, showing that Assumption 3(i) is satisfied. Since

$$\hat{B}_n^* - \hat{B}_n = (\omega-1)n^{1-1/\alpha}(\hat{\theta}_{long}^* - \hat{\theta}_{long}) = (\omega-1)n^{1-1/\alpha}\bar{\varepsilon}_n^*$$

and $T_n^* - \hat{B}_n = \omega n^{1-1/\alpha}\bar{\varepsilon}_n^*$, Assumption 3(ii) is also satisfied in this example. Thus, $\tilde{p}_n = \hat{H}_n(\hat{p}_n) \to_d U_{[0,1]}$ by Theorem 3.2.

REMARK 4.1 *Consider the 'm out of n' bootstrap data generating process,*

$$y_t^* = \hat{\theta}_{long,n} + \varepsilon_t^*, \quad t = 1, \dots, m,$$

where ε_t^* is an i.i.d. sample from the residuals $\hat{\varepsilon}_t = y_t - \hat{\theta}_{long,n}$, $t = 1, \dots, n$. Note that we now index $\hat{\theta}_{long,n}$ with n to emphasize the fact that $\hat{\theta}_{long,n}$ is based on n observations whereas the bootstrap sample size is m. Then

$$T_m^* := m^{1-1/\alpha}(\hat{\theta}_m^* - \hat{\theta}_{long,n}) = \omega m^{1-1/\alpha}\bar{\varepsilon}_m^* + (\omega - 1)m^{1-1/\alpha}\hat{\theta}_{long,n},$$

where $m^{1-1/\alpha}\bar{\varepsilon}_m^* \xrightarrow{d^*}_p S(\alpha)$ as $m \to \infty$; see Arcones and Giné (1989). Moreover, if $m = o(n)$,

$$\hat{B}_m := (\omega - 1)m^{1-1/\alpha}\hat{\theta}_{long,n} = (\omega - 1)m^{1-1/\alpha}n^{1/\alpha-1}(n^{1-1/\alpha}\hat{\theta}_{long,n}) = O_p((m/n)^{1/\alpha-1}) = o_p(1),$$

which shows that $T_m^* \xrightarrow{d^*}_p \omega S(\alpha)$. Hence, Assumption 2(i) is satisfied with $\xi_1 := \omega S(\alpha)$ and $\hat{B}_n = 0$. Since $B := (\omega - 1)c \neq 0$, Assumption 2(ii) holds with $\xi_2 := -B$ a.s. By Remark 3.1, it then follows that

$$\hat{p}_m := P^*(T_m^* \leq T_n) \xrightarrow{d} G_\gamma(G_\gamma^{-1}(U_{[0,1]}) - B)) = \Psi_\alpha(\Psi_\alpha^{-1}(U_{[0,1]}) - B).$$

This shows that the limiting distribution of $\hat{p}_m$ depends on B. Since B cannot be consistently estimated, the 'm out of n' bootstrap cannot be used to solve the problem.

4.2 Linear regression with omitted control variables

SETUP. In this section we consider inference about a target coefficient in a linear regression with a potentially large number of control variables, some of which may have limited explanatory power. Inference is based on a (misspecified) 'short' regression which omits controls whose coefficients are 'small' as a function of the sample size. Excluding such controls (e.g., using a pretest procedure) leads to estimators with omitted variable bias and invalidates standard applications of the bootstrap (e.g. Leeb and Pötsher, 2005). We apply the approach described in Section 3 to solve this invalidity problem.

More specifically, data are assumed to be generated by the linear model

$$y = x\beta + Q\eta + Z\delta + \varepsilon, \tag{4.2}$$

where $\varepsilon_t|W \sim$ i.i.d.$(0, \sigma^2)$ with $W := (x, Q, Z)$. The $n \times 1$ vector x contains observations on a target regressor, and Q, Z are matrices of control variables. The matrix Q contains the set of controls that are included in the model (baseline controls), whereas Z contains potential additional controls of limited explanatory power, i.e., their associated coefficients are local to zero and of the form

$$\delta = cn^{-1/2}.$$

This setup is similar to Li and Müller (2021), who develop an inference procedure which exploits a bound on the quadratic mean of the effect of the control variables on y. As Li and Müller

(2021), we assume for simplicity that $x'x = n$ and $Z'Z = nI_p$, where $p = \dim(Z)$; moreover, both x and Z have been projected off the baseline controls Q, i.e. $Q'x = 0$ and $Z'Q = 0$.

The goal is to test the null hypothesis $\mathsf{H}_0 : \beta = 0$. To do so, the econometrician estimates β and η by running the 'short' regression of y on (x, Q), i.e. Z is omitted from the regression. We let $\tilde{b}_n := (\tilde{\beta}_n, \tilde{\eta}_n', 0')'$ denote the restricted OLS estimator of $b := (\beta, \eta', \delta')'$ and $\tilde{\varepsilon} := y - W\tilde{b}_n$ the corresponding residuals. We reserve the notation $\hat{b}_n := (\hat{\beta}_n, \hat{\eta}_n', \hat{\delta}_n')'$ for the OLS estimator from the 'long' regression (4.2). The latter estimator is used below to generate bootstrap samples.

The t-ratio for testing H_0 is given by

$$T_n := \frac{\tilde{\beta}_n}{s(\tilde{\beta}_n)} = \frac{S_{xx.Q}^{-1} S_{xy.Q}}{\tilde{\sigma}_n \sqrt{n^{-1} S_{xx.Q}^{-1}}} = \frac{n^{1/2} S_{xy}}{\tilde{\sigma}_n},$$

where we define

$$s^2(\tilde{\beta}_n) := \tilde{\sigma}_n^2 n^{-1} S_{xx.Q}^{-1} = \tilde{\sigma}_n^2 n^{-1} \text{ with } \tilde{\sigma}_n^2 := n^{-1} \tilde{\varepsilon}' \tilde{\varepsilon},$$

and note that $S_{xx.Q} = 1$ and $S_{xy.Q} = S_{xy}$ because $x'x = n$ and $Q'x = 0$.

We impose the following conditions.

Assumption OC *(i)* $\varepsilon_t | W \sim$ *i.i.d.*$(0, \sigma^2)$, *where* $W = (x, Q, Z)$*; (ii)* $S_{WW} \to_p \Sigma$, *where* $\Sigma = (\Sigma_{ab})_{a,b \in \{x,Q,Z\}}$ *is positive definite; (iii)* $n^{1/2} S_{W\varepsilon} \to_d N(0, \sigma^2 \Sigma)$.

Assumption OC(i) formalizes the i.i.d. assumption on ε, conditional on W. Assumptions OC(ii) and OC(iii) are high-level conditions for which more primitive conditions are well known. The asymptotic distribution of T_n is presented in the following lemma.

Lemma 4.1 *Under Assumption OC and* $\mathsf{H}_0 : \beta = 0$*, it holds that* $T_n - B \to_d N(0, 1)$*, where* $B := \Sigma_{xZ} \sigma^{-1} c$.

Lemma 4.1 implies that Assumption 1′ holds for this example with $v^2 = 1$ and $B_n = \Sigma_{xZ} \sigma^{-1} c =: B$. Hence, unless $\Sigma_{xZ} = 0$ or $c = 0$, inference based on standard Gaussian critical values is invalid even if the explanatory power of the omitted controls is limited since we cannot consistently estimate B because it depends on the local parameter c.

Bootstrap. We generate the bootstrap sample as

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

where $\varepsilon^* | D_n \sim N(0, \hat{\sigma}_n^2 I_n)$ with $D_n := \{y, W\}$, $\hat{\sigma}_n^2 = n^{-1} \hat{\varepsilon}' \hat{\varepsilon}$, and $\hat{\varepsilon} = y - W\hat{b}_n$.[5]

Let $\tilde{b}_n^* := (\tilde{\beta}_n^*, \tilde{\eta}_n^{*\prime}, 0')'$ denote the bootstrap OLS estimators from the 'short' regression of y^* on (x, Q) and let $\tilde{\varepsilon}^* := y^* - x\tilde{\beta}_n^* - Q\tilde{\eta}_n^* := y^* - W\tilde{b}_n^*$ denote the corresponding residuals. The bootstrap analogue of T_n is

$$T_n^* := (\tilde{\beta}_n^* - \hat{\beta}_n)/s(\tilde{\beta}_n^*) = n^{1/2}(\tilde{\beta}_n^* - \hat{\beta}_n)/\tilde{\sigma}_n^*,$$

[5] The same results hold if we let $\varepsilon^* | D_n \sim N\left(0, \tilde{\sigma}_n^2 I_n\right)$. We can also establish similar results for the nonparametric bootstrap where ε^* is generated from the empirical distribution function of $\tilde{\varepsilon}$ or $\hat{\varepsilon}$ under a slighly stronger set of assumptions.

where

$$s^2(\tilde{\beta}_n^*) := \tilde{\sigma}_n^{*2} n^{-1} S_{xx.Q}^{-1} = n^{-1}\tilde{\sigma}_n^{*2}$$

and $\tilde{\sigma}_n^{*2} := n^{-1}\tilde{\varepsilon}^{*\prime}\tilde{\varepsilon}^*$.

LEMMA 4.2 *Under Assumption OC it holds that (i)*

$$T_n^* - \hat{B}_n \xrightarrow{d^*}_p N(0,1), \;\; \textit{where } \hat{B}_n := S_{xZ}\sigma^{-1} n^{1/2}\hat{\delta}_n,$$

and (ii) imposing also $\mathsf{H}_0 : \beta = 0$,

$$\begin{pmatrix} T_n - B \\ \hat{B}_n - B \end{pmatrix} \xrightarrow{d} N(0, V), \;\; \textit{where } V := \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_{xZ}\Sigma_{ZZ.x}^{-1}\Sigma_{Zx} \end{pmatrix}.$$

Lemma 4.2 shows that Assumption $2'$(i) holds with $v^2 = 1$ and $\hat{B}_n := S_{xZ}\sigma^{-1}n^{1/2}\hat{\delta}_n$ and that Assumption $2'$(ii) holds with $v_d^2 = 1 + \Sigma_{xZ}\Sigma_{ZZ.x}^{-1}\Sigma_{Zx} > 0$ and $\Sigma_{ZZ.x} = I - \Sigma_{Zx}\Sigma_{xZ}$.

REMARK 4.2 *We can show that*

$$\hat{B}_n - B = \Sigma_{xZ}\sigma^{-1}n^{1/2}(\hat{\delta}_n - \delta) + o_p(1),$$

where

$$n^{1/2}(\hat{\delta}_n - \delta) := S_{ZZ.x}^{-1} n^{1/2} S_{Z\varepsilon.x} \xrightarrow{d} N(0, \sigma^2\Sigma_{ZZ.x}^{-1}).$$

Hence, the bootstrap 'bias term' $\hat{B}_n$ is not consistent for B. Instead,

$$\hat{B}_n - B \xrightarrow{d} N(0, \Sigma_{xZ}\Sigma_{ZZ.x}^{-1}\Sigma_{Zx}) =: \xi_2$$

(jointly with $T_n - B \xrightarrow{d} N(0,1) =: \xi_1$). Since

$$T_n^* - \hat{B}_n|\hat{B}_n \xrightarrow{d^*}_p N(0,1),$$

this implies that

$$T_n^* \xrightarrow{d^*}_d N(B + \xi_2, 1)|\xi_2 \sim B + \xi_1 + \xi_2|\xi_2,$$

where $\xrightarrow{d^*}_d$ denotes weak convergence in distribution; see Remark 3.3 and Cavaliere and Georgiev (2020). Hence, the bootstrap distribution of T_n^* is random in the limit and does not match the limiting distribution of T_n, which is $N(B,1)$ under our assumptions; see Lemma 4.1. However, because ξ_2 has mean zero, this bootstrap replicates B on average.

By Corollary 3.4,

$$\hat{p}_n = P^*(T_n^* \le T_n) \xrightarrow{d} \Phi(m\Phi^{-1}(U_{[0,1]})) \text{ with } m^2 := 1 + \Sigma_{xZ}\Sigma_{ZZ.x}^{-1}\Sigma_{Zx},$$

which implies that, for any $u \in (0,1)$,

$$H_n(u) := P(\hat{p}_n \le u) \to H(u) := \Phi(\Phi^{-1}(u)/m).$$

Notice that when Z is univariate, m reduces to

$$m = \sqrt{\frac{1}{1-\rho_{xZ}^2}},$$

where $\rho_{xZ} = \Sigma_{xZ}$ corresponds to the population correlation coefficient between x and Z. Hence, $m \in [1, \infty)$ and equals one if and only if x and Z are orthogonal.

To sum up, this bootstrap is invalid in general. However, and crucially, inspection of the asymptotic distribution H of the bootstrap p-value shows that, although it differs from the cdf of a uniform random variable, it does not depend on the nuisance parameter B. This is essential in order to restore bootstrap validity using the machinery in Section 3.

REMARK 4.3 *Consider an alternative (fixed-design) bootstrap which resamples from the 'short' regression, i.e.* $y^* = x\tilde{\beta}_n + Q\tilde{\eta}_n + \varepsilon^*$, *where* $\varepsilon^*|D_n \sim N(0, \tilde{\sigma}_n^2 I_n)$. *The bootstrap analogue of* T_n *is*

$$T_n^* := (\tilde{\beta}_n^* - \tilde{\beta}_n)/s(\tilde{\beta}_n^*) = \tilde{\sigma}_n^{*-1} n^{1/2} S_{x\varepsilon^*} = \tilde{\sigma}_n^{-1} n^{1/2} S_{x\varepsilon^*} + o_{p^*}(1), \text{ in probability,}$$

because $\tilde{\sigma}_n^{*2} - \tilde{\sigma}_n^2 \xrightarrow{p^*}_p 0$. *Since, conditionally on* D_n, $\tilde{\sigma}_n^{-1} n^{1/2} S_{x\varepsilon^*} \sim N(0,1)$ *for any* n, *it follows that*

$$T_n^* \xrightarrow{d^*}_p N(0,1).$$

This shows that Assumption 2′(i) is satisfied with $\hat{B}_n = 0$. *Hence, this bootstrap does not replicate* B, *not even on average. In terms of the bootstrap p-value we find*

$$\hat{p}_n := P^*(T_n^* \leq T_n) = \Phi(T_n) + o_p(1) \xrightarrow{d} \Phi(B + \Phi^{-1}(U_{[0,1]})),$$

which implies that, for any $u \in (0,1)$,

$$P(\hat{p}_n \leq u) \to P(\Phi(B + \Phi^{-1}(U_{[0,1]})) \leq u) = P(U_{[0,1]} \leq \Phi(\Phi^{-1}(u) - B)) = \Phi(\Phi^{-1}(u) - B) \neq u.$$

Contrary to $\hat{p}_n$ *based on resampling from the 'long' regression, the limiting distribution of the bootstrap p-value based on resampling from the 'short' regression depends on* B. *Since* B *cannot be consistently estimated from the data, we cannot restore bootstrap validity using the prepivoting approach.*

RESTORING BOOTSTRAP VALIDITY. One approach is to estimate $H(u) := \Phi(\Phi^{-1}(u)/m)$ consistently using $\hat{H}_n(u) := \Phi(\Phi^{-1}(u)/\hat{m}_n)$, where

$$\hat{m}_n := (1 + S_{xZ} S_{ZZ.x}^{-1} S_{Zx})^{1/2}.$$

By Assumption OC, $\hat{m}_n \to_p m$, implying that a valid plug-in modified p-value is

$$\tilde{p}_n = \hat{H}_n(\hat{p}_n) = \Phi(\Phi^{-1}(\hat{p}_n)/\hat{m}_n)$$

by application of Corollary 3.2.

Alternatively, we can estimate $H_n(u)$ using the double bootstrap estimator $\hat{H}_n(u) := P^*(\hat{p}_n^* \leq u)$, where $\hat{p}_n^* := P^{**}(T_n^{**} \leq T_n^*)$ as described in Section 3.2.2. Specifically, let the second-level bootstrap sample be

$$y^{**} = x\hat{\beta}_n^* + Q\hat{\eta}_n^* + Z\hat{\delta}_n^* + \varepsilon^{**} := W\hat{b}_n^* + \varepsilon^{**},$$

where $\hat{b}_n^* := (\hat{\beta}_n^*, \hat{\eta}_n^{*\prime}, \hat{\delta}_n^{*\prime})'$ is obtained from the 'long' regression of y^* on W, and we let $\varepsilon^{**}|\{D_n^*, D_n\} \sim N(0, \hat{\sigma}_n^{*2} I_n)$, where $\{D_n^*, D_n\} := \{y, W, y^*\}$. We let $\tilde{b}_n^{**} := (\tilde{\beta}_n^{**}, \tilde{\eta}_n^{**\prime}, 0')'$ denote the restricted OLS estimator from regressing y^{**} on (x, Q), and we define $\tilde{\varepsilon}^{**} := y - W\tilde{b}_n^{**}$.

The second-level bootstrap analogue of T_n is defined as

$$T_n^{**} := (\tilde{\beta}_n^{**} - \hat{\beta}_n^*)/s(\tilde{\beta}_n^{**}) = n^{1/2}(\tilde{\beta}_n^{**} - \hat{\beta}_n^*)/\tilde{\sigma}_n^{**},$$

since

$$s^2(\tilde{\beta}_n^{**}) := \tilde{\sigma}_n^{**2} n^{-1} S_{xx.Q}^{-1} = n^{-1}\tilde{\sigma}_n^{**2},$$

where $\tilde{\sigma}_n^{**2} := n^{-1}\tilde{\varepsilon}^{**\prime}\tilde{\varepsilon}^{**}$. We show in Lemma 4.3 below that Assumption $3'$ is verified with $\hat{B}_n^* := S_{xZ}\sigma^{-1}n^{1/2}\hat{\delta}_n^*$, $\hat{B}_n := S_{xZ}\sigma^{-1}n^{1/2}\hat{\delta}_n$, and V as given in Lemma 4.2. The validity of the double bootstrap modified p-value, $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$, then follows from Lemmas 4.1, 4.2 and 4.3.

Lemma 4.3 *Under Assumption OC it holds that (i)* $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} N(0,1)$, *in probability, where* $\hat{B}_n^* := S_{xZ}\sigma^{-1}n^{1/2}\hat{\delta}_n^*$, *and (ii)*

$$\begin{pmatrix} T_n^* - \hat{B}_n \\ \hat{B}_n^* - \hat{B}_n \end{pmatrix} \xrightarrow{d^*}_p N(0, V),$$

where $\hat{B}_n$ *and* V *are as defined in Lemma 4.2.*

4.3 Inference after model averaging

Setup. We consider the case of inference on a target parameter based on averaging across different econometric models. The results in this section generalize those of Section 2. In particular, we consider model averaging based on M models and relax the Gaussianity and known variance assumptions.

Assume that data are generated according to the linear DGP

$$y = x\beta + Z\delta + \varepsilon, \tag{4.3}$$

where β is the (scalar) target parameter and $\varepsilon_t|W \sim$ i.i.d.$(0, \sigma^2)$ with $W := (x, Z)$. The goal is to test the null hypothesis $\mathsf{H}_0 : \beta = 0$.

The econometrician fits a set of M models, each of them based on different exclusion restrictions on the q-dimensional vector δ. This setup allows model averaging both explicitly and implicity. The former follows, e.g., Hansen (2007). The latter includes the common practice of robustness check in applied economics, 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, let R_m denote a $q \times q_m$ selection matrix which selects the potential controls. The m-th model includes x and $Z_m := ZR_m$ as regressors, and the corresponding OLS estimator of β is

$$\tilde{\beta}_{m,n} = S_{xx.Z_m}^{-1} S_{xy.Z_m},$$

where $S_{ab.c}$ is as defined previously. Note that if $R_m = I_q$, then $Z_m = Z$ and $\tilde{\beta}_{m,n} = S_{xx.Z}^{-1} S_{xy.Z} =: \hat{\beta}_n$ is the OLS estimator based on the full set of controls Z. If $R_m = 0$ then $Z_m = 0$, in which case $\tilde{\beta}_{m,n} = (x'x)^{-1}x'y$ is the OLS estimator based on the smallest possible model that only includes x as a regressor.

Given a set of fixed weights $\omega := (\omega_1, \dots, \omega_n)' \in \mathcal{H} := \{x \in [0,1]^M : \sum_{m=1}^M x_m = 1\}$, the model averaging estimator of β based on the M approximating models is the (linear) estimator

$$\tilde{\beta}_n := \sum_{m=1}^{M} \omega_m \tilde{\beta}_{m,n}.$$

REMARK 4.4 *We assume that the weights ω are fixed and independent of n. A popular example in forecasting is to use equal weighting. We could allow for stochastic weights as long as these are constant in the limit. This would be the case, for example, when the weights are based on moments that can be consistently estimated.*

In general, $\tilde{\beta}_n$ may suffer from finite sample and asymptotic bias, since it is based on averaging over a set of possibly misspecified models. This implies that the asymptotic distribution of $T_n := n^{1/2}(\tilde{\beta}_n - \beta)$ may not be centered at zero, as we show next.

We impose the following conditions.

ASSUMPTION MA *(i) $\varepsilon_t|W \sim i.i.d.(0, \sigma^2)$, where $W = (x, Z)$; (ii) $S_{WW} \to_p \Sigma_{WW} := \begin{pmatrix} \Sigma_{xx} & \Sigma_{xZ} \\ \Sigma_{Zx} & \Sigma_{ZZ} \end{pmatrix}$ with $\text{rank}(\Sigma_{WW}) = q + 1$; (iii) $n^{1/2} S_{W\varepsilon} \to_d N(0, \Omega)$ with $\Omega := \sigma^2 \Sigma_{WW}$.*

To describe the asymptotic distribution of T_n, we introduce the following notation. We let $\Sigma_{xZ_m} := \Sigma_{xZ} R_m$, $\Sigma_{Z_m Z_m} := R_m' \Sigma_{ZZ} R_m$, $\Sigma_{xx.Z_m} := \Sigma_{xx} - \Sigma_{xZ} R_m (R_m' \Sigma_{ZZ} R_m)^{-1} R_m' \Sigma_{Zx}$, and $\Sigma_{xZ.Z_m} := \Sigma_{xZ} - \Sigma_{xZ} R_m (R_m' \Sigma_{ZZ} R_m)^{-1} R_m' \Sigma_{ZZ}$. We also let $A_n := \sum_{m=1}^M \omega_m S_{xx.Z_m}^{-1} n^{-1} x' M_{Z_m}$, where $M_{Z_m} := I_n - Z_m (Z_m' Z_m)^{-1} Z_m'$. Finally, $\bar{d}_{M,n}' := \sum_{m=1}^M \omega_m S_{xx.Z_m}^{-1} (1, -S_{xZ_m} S_{Z_m Z_m}^{-1} R_m')$ and $\bar{d}_M' := \sum_{m=1}^M \omega_m \Sigma_{xx.Z_m}^{-1} (1, -\Sigma_{xZ_m} \Sigma_{Z_m Z_m}^{-1} R_m')$. With this notation,

$$\tilde{\beta}_n = A_n y = A_n x \beta + A_n Z \delta + A_n \varepsilon = \beta + A_n Z \delta + A_n \varepsilon,$$

and the following lemma holds.

LEMMA 4.4 *Under Assumption MA and $\mathsf{H}_0 : \beta = 0$, it holds that $T_n := n^{1/2}(\tilde{\beta}_n - \beta)$ satisfies*

$$T_n - B_n =: n^{1/2} A_n \varepsilon = \bar{d}_{M,n}' (n^{1/2} S_{W\varepsilon}) \xrightarrow{d} N(0, v^2), \tag{4.4}$$

where

$$v^2 := \bar{d}_M' \Omega \bar{d}_M \text{ and } B_n := n^{1/2} A_n Z \delta = \sum_{m=1}^{M} \omega_m S_{xx.Z_m}^{-1} S_{xZ.Z_m} n^{1/2} \delta. \tag{4.5}$$

Lemma 4.4 shows that Assumption 1′ holds for this example with v^2 and B_n as indicated in (4.5). The asymptotic variance v^2 is a quadratic form of Ω, whose weights $\bar{d}_M'$ depend on population moments of the regressors and on ω; see (B.4). The variance v^2 can be consistently estimated by replacing Ω and $\bar{d}_M$ by consistent estimators. Alternatively, we can use the bootstrap, as recently shown by Hounyo and Lahiri (2021).

However, inference based on critical values obtained from a normal distribution with mean zero is invalid asymptotically due to the presence of B_n.

REMARK 4.5 *As (4.5) shows, B_n may not vanish in the limit, even if δ is small. To see this, note that the order of magnitude of B_n is the same as that of $n^{1/2}\delta$ because $S_{xx.Z_m}^{-1} S_{xZ.Z_m} \to_p$*

$\Sigma_{xx.Z_m}^{-1}\Sigma_{xZ.Z_m}$ *under our assumptions. Consequently, 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, and Liu, 2015), then it follows that*

$$B_n \to_p B := \sum_{m=1}^{M} \omega_m \Sigma_{xx.Z_m}^{-1}\Sigma_{xZ.Z_m} c,$$

which is not zero in general (unless, e.g., $c = 0$ *or* $\Sigma_{xZ} = 0$*). Hence, under the local-to-zero assumption on* δ*,*

$$T_n := n^{1/2}(\tilde{\beta}_n - \beta) \xrightarrow{d} N(B, v^2),$$

showing that $\tilde{\beta}_n$ *is consistent for* β*, but it is asymptotically biased. Because the asymptotic bias* B *depends on* c*, which is not consistently estimable, we cannot obtain valid critical values from a Gaussian distribution centered at some sample analogues of* B *and* v^2*. In particular, replacing* B *with* $\hat{B}_n = \sum_{m=1}^{M} \omega_m S_{xx.Z_m}^{-1} S_{xZ.Z_m} n^{1/2}\hat{\delta}_n$ *is not asymptotically valid as the asymptotic distribution of* $T_n - \hat{B}_n$ *is not* $N(0, v^2)$*.*

As we will show next, the presence of B_n complicates standard bootstrap inference. However, we will show that the modified p-value bootstrap approach described in Section 3 can be successfully applied in this context. In particular, 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 inconsistency, we can obtain a modified bootstrap p-value that is asymptotically valid.

Bootstrap. We consider a fixed regressor bootstrap (FRB) algorithm based on the full model. More specifically, we generate the bootstrap sample as

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

where $\varepsilon^*|D_n \sim N(0, \hat{\sigma}_n^2 I_n)$, $\hat{\sigma}_n^2 = n^{-1}\hat{\varepsilon}'\hat{\varepsilon}$ is the OLS residual variance from the full model, and $D_n = \{y, W\}$. Similar results can be established for the nonparametric bootstrap where ε^* is obtained from the empirical distribution function of $\hat{\varepsilon}$.

The model averaging estimator in the bootstrap world is given by

$$\tilde{\beta}_n^* := \sum_{m=1}^{M} \omega_m \tilde{\beta}_{m,n}^*,$$

where $\tilde{\beta}_{m,n}^* := S_{xx.Z_m}^{-1} S_{xy^*.Z_m}$ is the bootstrap OLS estimator from the m^{th} model. We can write $\tilde{\beta}_n^* = A_n y^*$, which implies that

$$\tilde{\beta}_n^* = A_n x\hat{\beta}_n + A_n Z\hat{\delta}_n + A_n\varepsilon^*.$$

It holds then that

$$T_n^* := n^{1/2}(\tilde{\beta}_n^* - \hat{\beta}_n) = \hat{B}_n + n^{1/2}A_n\varepsilon^*,$$

where

$$\hat{B}_n := n^{1/2}A_n Z\hat{\delta}_n = \sum_{m=1}^{M} \omega_m S_{xx.Z_m}^{-1} S_{xZ.Z_m} n^{1/2}\hat{\delta}_n.$$

We can show that

$$T_n^* - \hat{B}_n = n^{1/2} A_n \varepsilon^* = \bar{d}'_{M,n}(n^{1/2} S_{W\varepsilon^*}) \xrightarrow{d^*}_p N(0, v^2)$$

since $n^{1/2} S_{W\varepsilon^*}|D_n \sim N(0, \hat{\Omega}_n)$ with $\hat{\Omega}_n := \hat{\sigma}_n^2 S_{WW} \to_p \sigma^2 \Sigma_{WW} =: \Omega$ and $\bar{d}_{M,n} \to_p \bar{d}_M$. Hence, $\text{Var}^*(T_n^* - \hat{B}_n) = \bar{d}'_{M,n} \hat{\Omega}_n \bar{d}_{M,n} \to_p v^2$ and Assumption 2′(i) is verified in this example. However, $\hat{B}_n$ is not close to B_n as n increases. In fact, we can show that

$$\hat{B}_n - B_n = A_n Z n^{1/2}(\hat{\delta}_n - \delta) \xrightarrow{d} \xi_2 := N(0, v_{22}), \tag{4.6}$$

given the asymptotic normality of $n^{1/2}(\hat{\delta}_n - \delta)$ and the fact that $A_n Z$ converges in probability to $\sum_{m=1}^M \omega_m \Sigma_{xx.Z_m}^{-1} \Sigma_{xZ.Z_m}$. This result implies that T_n^* does not mimic the asymptotic distribution of the original statistic T_n. 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. This implies the asymptotic invalidity of $\hat{p}_n = P^*(T_n^* \leq T_n)$, as we show next. To this end, define $\bar{b}'_{M,n} := \sum_{m=1}^M \omega_m S_{xx.Z_m}^{-1} S_{xZ.Z_m} S_{ZZ.x}^{-1}(-S_{Zx} S_{xx}^{-1}, I_q)$ and let $\bar{b}'_M$ denote its probability limit.

Lemma 4.5 *Under Assumption MA it holds that (i) $T_n^* - \hat{B}_n \xrightarrow{d^*}_p N(0, v^2)$, and (ii) imposing also $\mathsf{H}_0 : \beta = 0$,*

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} = \begin{pmatrix} \bar{d}'_{M,n} \\ \bar{b}'_{M,n} \end{pmatrix} n^{1/2} S_{W\varepsilon} \xrightarrow{d} N(0, V),$$

where

$$V = \begin{pmatrix} \bar{d}'_M \Omega \bar{d}_M & \bar{d}'_M \Omega \bar{b}_M \\ \bar{b}'_M \Omega \bar{d}_M & \bar{b}'_M \Omega \bar{b}_M \end{pmatrix} =: \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}$$

is positive definite and $v^2 = v_{11}$.

Lemma 4.5 implies that Assumptions 1′ and 2′ are satisfied in this example with $v^2 = v_{11}$ and $v_d^2 = v_{11} + v_{22} - 2v_{12} > 0$. Hence, by Theorem 3.1, the bootstrap p-value satisfies $H_n(u) := P(\hat{p}_n \leq u) \to H(u) = \Phi(m^{-1}\Phi^{-1}(u))$, where $m^2 := v_d^2/v^2$. This proves the invalidity of the standard bootstrap p-value, $\hat{p}_n$.

Restoring bootstrap validity. One approach is to estimate $H(u)$ consistently using $\hat{H}_n(u) := \Phi(\Phi^{-1}(u)/\hat{m}_n)$, where $\hat{m}_n := (\hat{v}_{d,n}^2/\hat{v}_n)^{1/2}$ and $\hat{v}_{d,n}^2 := \hat{v}_{11,n} + \hat{v}_{22,n} - 2\hat{v}_{12,n}$ with $\hat{v}_{ij,n}$ denoting the sample analogues of v_{ij} defined in Lemma 4.5; for instance, $\hat{v}_{11,n} = \hat{v}_n^2 = \bar{d}'_{M,n} \hat{\Omega}_n \bar{d}'_{M,n}$. By Assumption MA, $\hat{m}_n \to_p m$, implying that a valid plug-in modified p-value is

$$\tilde{p}_n = \hat{H}_n(\hat{p}_n) = \Phi(\Phi^{-1}(\hat{p}_n)/\hat{m}_n)$$

by application of Corollary 3.2.

Alternatively, we can estimate $H_n(u)$ using the double bootstrap estimator $\hat{H}_n(u) := P^*(\hat{p}_n^{**} \leq u)$, where $\hat{p}_n^{**} := P^{**}(T_n^{**} \leq T_n^*)$ as described in Section 3. Specifically, let the double bootstrap sample be

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

where $(\hat{\beta}_n^*, \hat{\delta}_n^*)$ is the OLS estimator of (β, δ) obtained from the full model and $\varepsilon^{**}|\{D_n, D_n^*\} \sim N(0, \hat{\sigma}_n^{*2} I_n)$, with $D_n^* = \{y^*, W\}$, $\hat{\sigma}_n^{*2} = n^{-1}\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*$, and $\varepsilon^* = y^* - x\hat{\beta}_n^* - Z\hat{\delta}_n^*$.

The double bootstrap analogue of T_n is defined as

$$T_n^{**} := n^{1/2}(\tilde{\beta}_n^{**} - \hat{\beta}_n^{*}),$$

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 m^{th} model.

LEMMA 4.6 *Under Assumption MA, $T_n^{**} - \hat{B}_n^{*} \xrightarrow{d^{**}}_{p^*} N\left(0, v^2\right)$, in probability, where $\hat{B}_n^{*} := n^{1/2} A_n Z \hat{\delta}_n^{*}$ and*

$$\begin{pmatrix} T_n^{*} - \hat{B}_n \\ \hat{B}_n^{*} - \hat{B}_n \end{pmatrix} = \begin{pmatrix} \bar{d}'_{M,n} \\ \bar{b}'_{M,n} \end{pmatrix} n^{1/2} S_{W\varepsilon^*} \xrightarrow{d^*}_p N(0, V)$$

with $\hat{B}_n := n^{1/2} A_n Z \hat{\delta}_n$ and V as defined in Lemma 4.5.

Lemma 4.6 shows that Assumption 3′ is verified in this example. The asymptotic validity of the double bootstrap modified p-value, $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$, now follows from Lemmas 4.4–4.6 and Theorem 3.2.

EXTENSION TO THE PAIRS BOOTSTRAP. We now extend the previous results to the pairs bootstrap, which is widely used in applications. Interestingly, the main difference with respect to the FRB implemented above is that the centered pairs bootstrap statistic $T_n^{*} - \hat{B}_n$ no longer replicates the distribution of $T_n^{*} - B_n$, except if δ is local to zero (see Remark 4.5). Thus, the pairs bootstrap is invalid, but we can prove validity of an appropriate modification of the pairs bootstrap using the results in Section 3.5.

To simplify the discussion we consider the case with scalar z_t in (4.3) 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,

$$\tilde{\beta}_n = S_{xx}^{-1} S_{xy}.$$

This estimator corresponds to the model averaging estimator with $M = 1$, where all weight is put on the smallest model. The statistic of interest is

$$T_n = n^{1/2}(\tilde{\beta}_n - \beta).$$

In this special case, Lemma 4.4 simplifies as follows.

LEMMA 4.7 *Suppose Assumption MA holds. Then*

$$T_n - B_n \to_d N(0, v^2),$$

where

$$v^2 = \sigma^2 \Sigma_{xx}^{-1} \text{ and } B_n = S_{xx}^{-1} S_{xz} n^{1/2} \delta.$$

Consider now a pairs bootstrap sample $\{y^*, x^*, z^*\}$, 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 i.i.d. draw from $\hat{\varepsilon}_t = y_t - x_t\hat{\beta}_n - z_t\hat{\delta}_n$. The pairs bootstrap model averaging estimator is

$$\tilde{\beta}_n^* := S_{x^*x^*}^{-1} S_{x^*y^*} = \hat{\beta}_n + S_{x^*x^*}^{-1} S_{x^*z^*}\hat{\delta}_n + S_{x^*x^*}^{-1} S_{x^*\varepsilon^*},$$

and the pairs bootstrap statistic is

$$T_n^* := n^{1/2}(\tilde{\beta}_n^* - \hat{\beta}_n) = B_n^* + n^{1/2} S_{x^*x^*}^{-1} S_{x^*\varepsilon^*}$$

with

$$B_n^* := S_{x^*x^*}^{-1} S_{x^*z^*} n^{1/2}\hat{\delta}_n.$$

Therefore, and in contrast with the FRB, when the pairs bootstrap is implemented 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 $T_n - B_n$, as shown in the next lemma.

We strengthen the previous conditions by adding the following.

Assumption MA$_2$ *(i)* $\sup_t E\|w_t\|^4 < \infty$ *and* $E|\varepsilon_t^4| < \infty$*; (ii)* $n^{-1}\sum_{t=1}^n x_t^2\varepsilon_t^2 \to_p \sigma^2\Sigma_{xx}$, $n^{-1}\sum_{t=1}^n x_t^2 w_t w_t' \to_p \Sigma_r > 0$, *and* $n^{-1}\sum_{t=1}^n x_t^2 w_t\varepsilon_t \to_p 0$.

Lemma 4.8 *Suppose Assumptions MA and MA$_2$ hold. Then*

$$T_n^* - \hat{B}_n \xrightarrow{d^*}_p N(0, v^2 + \kappa^2),$$

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 constrast 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, which demonstrates that the double bootstrap p-values are asymptotically uniformly distributed.

4.4 Inference based on ridge estimators

Setup. Consider the linear regression model $y_t = \theta' x_t + \varepsilon_t$, $t = 1, \ldots, n$, where x_t is a $p \times 1$ non-stochastic vector and $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$. Interest is in inference on a linear combination of θ, such as the simple null hypothesis $\mathsf{H}_0 : g'\theta = r$, based on ridge estimation of θ. Specifically, a classic lasso-type/penalized least squares estimator of θ (or 'bridge', see Frank and Friedman, 1993) has the form $\tilde{\theta}_n^{(\eta)} := \arg\min_{\theta\in\mathbb{R}^p}\{\sum_{t=1}^n (y_t - \theta' x_t)^2 + c_n \|\theta\|_\eta^\eta\}$, where $\|x\|_\eta$ denotes the ℓ_η-norm of the vector x. The tuning parameter c_n controls the degree of shrinkage toward zero for the estimator. For $\eta = 2$ we obtain the so-called ridge estimator, which we denote by $\tilde{\theta}_n$ and we discuss in the following; see Giglio and Xiu (2021) for a recent application to estimation of asset pricing (factor) models. The ridge estimator has closed form expression

$$\tilde{\theta}_n = \tilde{S}_{xx}^{-1} S_{xy}, \tag{4.7}$$

where $\tilde{S}_{xx} := S_{xx} + n^{-1}c_n I_p$. Clearly, $c_n = 0$ corresponds to the OLS estimator, $\hat{\theta}_n$. Notice also that $\tilde{\theta}_n$ can be written as $\tilde{\theta}_n = \Omega_n \hat{\theta}_n$, where $\Omega_n := \tilde{S}_{xx}^{-1} S_{xx}$ is a weighting matrix which shrinks the OLS estimator $\hat{\theta}_n$ towards zero. As in Knight and Fu (2000) we assume the following.

Assumption RE *(i) $\varepsilon_t \sim$ i.i.d.$(0,\sigma^2)$; (ii) $\max_{t=1,\dots,n} x_t' x_t = o(n)$; (iii) S_{xx} is nonsingular for any n and converges to a positive definite matrix, Σ_{xx}; (iv) $\theta = \delta n^{-1/2}$; and (v) $n^{-1}c_n \to c_0 \geq 0$.*

The asymptotic properties of $\tilde{\theta}_n$ depend crucially on the shrinkage parameter c_n as well as on the magnitude of the coefficient vector θ. In particular, rearranging (4.7) yields

$$n^{1/2}(\tilde{\theta}_n - \theta) = n^{1/2}\tilde{S}_{xx}^{-1} S_{x\varepsilon} + b_n,$$

where $b_n := -c_n n^{-1/2}\tilde{S}_{xx}^{-1}\theta = E(n^{1/2}(\tilde{\theta}_n - \theta))$ is a bias term. Note that b_n can either converge or diverge, depending on the magnitudes of c_n and θ. We are interested in the case where the regressors have limited explanatory power, i.e. where $\theta = \delta n^{-1/2}$ is local to zero as in Assumption RE(iv). Indeed, this case can be taken as a motivation for shrinkage towards zero and hence for ridge estimation. If, in addition, $n^{-1}c_n \to c_0 \geq 0$ as in Assumption RE(v) (see also Knight and Fu, 2000), then b_n does not vanish as $n \to \infty$ unless $c_0 = 0$. Hence, for $c_0 > 0$, $\tilde{\theta}_n$ is asymptotically biased and the bias term cannot be consistently estimated.

Consider a (test) stastistic for H_0 of the form

$$T_n := n^{1/2}(g'\tilde{\theta}_n - r).$$

It is not difficult to see that, under H_0, T_n satisfies Assumption $1'$ and, in particular,

$$T_n = n^{1/2}g'(\tilde{\theta}_n - \theta) = n^{1/2}g'\tilde{S}_{xx}^{-1}S_{x\varepsilon} + B_n,$$

where $B_n := g'b_n = -c_n n^{-1/2}g'\tilde{S}_{xx}^{-1}\theta$. With $\tilde{\Sigma}_{xx} := \Sigma_{xx} + c_0 I_p$, under Assumption RE and H_0,

$$\xi_{1,n} := T_n - B_n = n^{1/2}g'\tilde{S}_{xx}^{-1}S_{x\varepsilon} \xrightarrow{d} \xi_1 \sim N(0, v^2), \tag{4.8}$$

where $v^2 := \sigma^2 g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g$. Hence, inference based on the quantiles of the $N(0,v^2)$ distribution is invalid unless $c_0 = 0$.

Bootstrap. Consider a pairs bootstrap sample $\{y_t^*, x_t^*\}$ built by i.i.d. resampling of the tuples $\{y_t, x_t; t = 1,\dots,n\}$. The bootstrap analogue of the ridge estimator is $\tilde{\theta}_n^* := \tilde{S}_{x^*x^*}^{-1}S_{x^*y^*}$, where $\tilde{S}_{x^*x^*} := S_{x^*x^*} + n^{-1}c_n I_p$. With $\varepsilon_t^* := y_t^* - \hat{\theta}_n' x_t^*$, we consider the following bootstrap analogue of the original test statistic,

$$T_n^* := n^{1/2}g'(\tilde{\theta}_n^* - \hat{\theta}_n) = B_n^* + \xi_{1,n}^*,$$

where

$$B_n^* := -c_n n^{-1/2}g'\tilde{S}_{x^*x^*}^{-1}\hat{\theta}_n \text{ and } \xi_{1,n}^* := n^{1/2}g'\tilde{S}_{x^*x^*}^{-1}S_{x^*\varepsilon^*}. \tag{4.9}$$

Note that T_n^* is centered using $\hat{\theta}_n$ to guarantee that ε_t^* and x_t^* are uncorrelated in the bootstrap world.

As we show next, the bootstrap fails to approximate the asymptotic distribution of T_n (see also Chatterjee and Lahiri, 2010, 2011). The reason is that, although $\xi_{1,n}^*$ mimics $\xi_{1,n}$

(the stochastic part of T_n), B_n^* does not approximate B_n in large samples. To determine the distribution of the bootstrap p-value, we verify the high-level conditions in Section 3. We do this through the following two lemmas, which are based on the fact that in large samples B_n^* is close to its non-bootstrap analogue, $\hat{B}_n := -c_n n^{-1/2} g' \tilde{S}_{xx}^{-1} \hat{\theta}_n$.

LEMMA 4.9 *Under Assumption RE and the null hypothesis* $\mathsf{H}_0 : g'\theta = r$ *it holds that*

$$\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} = \begin{pmatrix} g'\tilde{S}_{xx}^{-1} \\ -c_n n^{-1} g' \tilde{S}_{xx}^{-1} S_{xx}^{-1} \end{pmatrix} n^{1/2} S_{x\varepsilon} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim N(0, V), \quad V = (v_{ij}), \tag{4.10}$$

with $v_{11} = v^2$ *as defined above,* $v_{12} = -c_0 g' \tilde{\Sigma}_{xx}^{-1} \tilde{\Sigma}_{xx}^{-1} g$*, and* $v_{22} = c_0^2 g' \tilde{\Sigma}_{xx}^{-1} \Sigma_{xx}^{-1} \tilde{\Sigma}_{xx}^{-1} g$.

LEMMA 4.10 *Under Assumption RE and the null hypothesis* $\mathsf{H}_0 : g'\theta = r$*, and assuming also* $E\varepsilon_t^4 < \infty$ *and* $\max_{t=1,\ldots,n} x_t' x_t = o(n^{1/2})$*, it holds that* $T_n^* - B_n^* = T_n^* - \hat{B}_n + o_p^*(1) \xrightarrow{d^*} N(0, v^2)$*, in probability.*

Hence, we can rely on Corollary 3.4, which implies that the bootstrap p-value satisfies

$$\hat{p}_n \xrightarrow{d} \Phi(m\Phi^{-1}(U_{[0,1]})) \tag{4.11}$$

with m^2 given by

$$m^2 = \frac{g' \Sigma_{xx}^{-1} g}{g' \tilde{\Sigma}_{xx}^{-1} \Sigma_{xx} \tilde{\Sigma}_{xx}^{-1} g}, \tag{4.12}$$

see Appendix B.4. Notice that (4.11) holds irrespectively of θ being fixed or local to zero. Thus, the bootstrap is invalid unless $c_0 = 0$ which implies $m = 1$.

REMARK 4.6 *In the univariate case (where we set* $g = 1$*) it follows from (4.12) that*

$$m = 1 + c_0/\Sigma_{xx} \geq 1,$$

with equality holding only for $c_0 = 0$*. Hence, a left-sided test with nominal level* $\alpha \in (0, 1/2)$ *that rejects when* $\hat{p}_n \leq \alpha$ *is over-sized. Specifically, its asymptotic size follows from (4.11) and is given by*

$$P(\hat{p}_n \leq \alpha) \to P(\Phi(m\Phi^{-1}(U_{[0,1]})) \leq \alpha) = \Phi(\Phi^{-1}(\alpha) \tfrac{\Sigma_{xx}}{\Sigma_{xx} + c_0}),$$

which is greater than α *unless* $c_0 = 0$.

REMARK 4.7 *It is worth noting that the cdf of the ridge test statistic, conditionally on the data, is random in the limit, see Remark 3.3, unless* $c_0 = 0$ *(i.e.,* $c_n = o(n)$*). Specifically, using the fact that* $\hat{B}_n - B_n \xrightarrow{d} \xi_2$ *and* $B_n \to B := -c_0 g' \tilde{\Sigma}_{xx} \delta$*, the distribution of* T_n^**, conditionally on the data, converges (weakly) to the distribution of* $\xi_1^* + \xi_2 + B$ *conditionally on* ξ_2*, where* ξ_1^* *is distributed as* ξ_1 *and is independent of* ξ_2*. That is,* $P^*(T_n^* \leq u) \to_w \Phi((u - \xi_2 - B)/v)$.

REMARK 4.8 *The use of the pairs bootstrap (which implicitly sets* $\hat{\theta}_n$ *as true parameter under the bootstrap probability measure) guarantees that* $\hat{B}_n - B_n$ *is centered around zero as* $n \to \infty$*, as required in Assumption 2′.*

Restoring bootstrap validity. Post-ridge estimation valid inference can be performed using either the plug-in approach or the double bootstrap. For the plug-in method, a simple consistent estimator of m is given by

$$\hat{m}_n := \sqrt{\frac{g' S_{xx}^{-1} g}{g' \tilde{S}_{xx}^{-1} S_{xx} \tilde{S}_{xx}^{-1} g}}.$$

Inference based on the plug-in modified p-value,

$$\tilde{p}_n = \hat{H}_n(\hat{p}_n) = \Phi(\Phi^{-1}(\hat{p}_n)/\hat{m}_n),$$

is then asymptotically valid by Corollary 3.2.

To implement the double bootstrap, it suffices to construct the double bootstrap sample mimicking the first-level bootstrap scheme. That is, we can draw the double bootstrap sample $\{y_t^{**}, x_t^{**}; t = 1, \ldots, n\}$ as i.i.d. 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 assumptions, prepivoting based on the double bootstrap provides valid inference. In particular, the following lemma verifies Assumption $3'$ and validity of the double bootstrap modified p-value follows by application of Theorem 3.2.

Lemma 4.11 *Under the assumptions of Lemma 4.10, it holds that (i)* $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} N(0, v^2)$, *in probability, with* $\hat{B}_n^* := -c_n n^{-1/2} g' \tilde{S}_{x^*x^*}^{-1} \hat{\theta}_n^*$, *and (ii)*

$$\begin{pmatrix} T_n^* - \hat{B}_n \\ \hat{B}_n^* - \hat{B}_n \end{pmatrix} \xrightarrow{d^*}_p N(0, V),$$

where $\hat{B}_n$ *and* V *are defined in (4.9) and (4.10), respectively.*

4.5 Nonlinear dynamic panel data models with incidental parameter bias

Another example that fits our framework is inference based on panel data estimators subject to incidental parameter bias. We consider the properties of the cross-sectional pairs bootstrap considered by Kaffo (2014), Dhaene and Jochmans (2015), and Gonçalves and Kaffo (2015) in the context of a general nonlinear panel data model. Although this bootstrap cannot replicate the bias, we show that our prepivoting approach based on a plug-in estimator of the bias is valid. Recently, Higgins and Jochmans (2022) proposed a (double) bootstrap procedure that retains asymptotic validity without an explicit plug-in estimator of the bias, but their procedure relies heavily on the parametric distribution assumption.

Setup. Let z_{it} denote a vector of random variables for a set of n individuals, $i = 1, \ldots, n$, over T time periods, $t = 1, \ldots, T$. Given a model for the density function $f_{it}(\theta, \alpha_i) := f(z_{it}, \theta, \alpha_i)$, the parameter of interest is $\theta \in \Theta$, which is common to all the individuals, while $\alpha_i \in \mathcal{A}$ denote the individual fixed effects. The fixed effects estimator of θ is the maximum likelihood estimator

(MLE) defined as

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n}\sum_{t=1}^{T} \log f_{it}(\theta, \hat{\alpha}_i(\theta)), \text{ where } \hat{\alpha}_i(\theta) = \arg\max_{\alpha_i\in\mathcal{A}} \sum_{t=1}^{T} \log f_{it}(\theta, \alpha_i). \tag{4.13}$$

Under certain regularity conditions (see, e.g., Hahn and Kuersteiner, 2011), including letting $n, T \to \infty$ jointly such that $n/T \to \rho < \infty$,

$$T_n := \sqrt{nT}(\hat{\theta}_n - \theta) \xrightarrow{d} N(B, v^2), \tag{4.14}$$

where B denotes the incidental parameter bias and v^2 is the asymptotic variance of $\hat{\theta}_n$. Hence, Assumption 1 is satisfied with $\xi_1 := N(0, v^2)$ (equivalently, Assumption $1'$ is satisfied).

The exact forms of B and v^2 may be quite involved and depend on the type of heterogeneity and dependence assumptions imposed on z_{it}. A standard assumption is that z_{it} is independent across i while allowing for time series dependence of unknown form; see Hahn and Kuersteiner (2011).

Bootstrap. Given the cross sectional independence assumption, a natural bootstrap method in this context is the cross sectional pairs bootstrap. The idea is to resample $z_i = (z_{i1}, \ldots, z_{iT})'$ in an i.i.d. fashion in the cross sectional dimension. If $z_{it} = (y_{it}, x_{it})'$ and $f(z_{it}, \theta, \alpha_i) = f(y_{it}|x_{it}, \theta, \alpha_i)$ is the conditional density of y_{it} given x_{it}, this is equivalent to a cross sectional pairs bootstrap. As the results of Kaffo (2014, Theorem 3.1) show, this bootstrap fails to capture the bias term B. In particular, letting $\hat{\theta}_n^*$ denote the bootstrap analogue of $\hat{\theta}_n$, we have that

$$T_n^* := \sqrt{nT}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d^*}_p N(0, v^2),$$

which implies that[6]

$$\hat{p}_n := P^*(T_n^* \le T_n) = \Phi(v^{-1}T_n) + o_p(1) \xrightarrow{d} \Phi(v^{-1}B + \Phi^{-1}(U_{[0,1]})).$$

Thus,

$$P(\hat{p}_n \le u) \to H(u) := P(\Phi(\Phi^{-1}(U_{[0,1]}) + v^{-1}B) \le u) = \Phi(\Phi^{-1}(u) - v^{-1}B),$$

which shows that the bootstrap test based on $\hat{p}_n$ is asymptotically invalid since its limiting distribution is not uniform.

Remark 4.9 *Note that, in this example, $\hat{L}_n(u) := P^*(T_n^* \le u) \to_p \Phi(u/v)$, showing that the bootstrap conditional distribution of T_n^* is not random in the limit. The invalidity of $\hat{p}_n$ is due to the fact that the cross sectional pairs bootstrap induces $\hat{B}_n = 0$, whereas $B \neq 0$. This implies that $\hat{B}_n - B = -B := \xi_2$ is not random. The fact that ξ_2 is not zero is the cause of the bootstrap invalidity. See Remark 3.1, which contains this example as a special case.*

Contrary to previous examples (see, e.g., Remarks 4.1 and 4.3), B and v can both be consistently estimated. Hence, in this example we can restore bootstrap validity by modifying the

[6] Note that this result can also be obtained by an application of Theorem 3.1 by setting $\xi_1 \sim N(0, v^2)$, $\xi_2 = -B$ a.s., $G_\gamma(u) = \Phi(v^{-1}u)$, and $F_\phi(u) = P(\xi_1 - \xi_2 \le u) = \Phi(v^{-1}u - v^{-1}B)$. Although Assumption $2'$(ii) is not satisfied here, because it requires $\hat{B}_n - B$ to be asymptotically centered at zero, we can still appeal to Theorem 3.1 since Assumptions 1 and 2 hold.

bootstrap p-value using a plug-in approach. More specifically, let $\tilde{B}_n$ and $\hat{v}_n$ denote consistent estimators of B and v, respectively.[7] By Corollary 3.2,

$$\tilde{p}_n = \hat{H}_n(\hat{p}_n) = \Phi(\Phi^{-1}(\hat{p}_n) - \hat{v}_n^{-1}\hat{B}_n) \xrightarrow{d} U_{[0,1]},$$

where $\hat{H}_n(u) := \Phi(\Phi^{-1}(u) - \hat{v}_n^{-1}\hat{B}_n)$ is a consistent estimator of $H(u)$.

Remark 4.10 *A double bootstrap modified p-value version of $\tilde{p}_n$ is not valid in this setting. The reason is that the double bootstrap mimics the behavior of the first-level bootstrap, i.e.*

$$T_n^{**} := \sqrt{nT}(\hat{\theta}_n^{**} - \hat{\theta}_n^*) \xrightarrow{d^{**}}_p N(0, v^2),$$

so that $\hat{B}_n^*$ in Assumption 3(i) is zero. Since $\hat{B}_n = 0$, Assumption 3(ii) holds with $\hat{B}_n^* - \hat{B}_n = 0$, whereas Assumption 2(ii) has $\hat{B}_n - B_n = -B$. Then,

$$\hat{p}_n^* = P^{**}(v^{-1}T_n^{**} \leq v^{-1}T_n^*) = \Phi(v^{-1}T_n^*) \xrightarrow{d^*}_p \Phi(\Phi^{-1}(U_{[0,1]})) = U_{[0,1]},$$

whereas

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

Thus, $\hat{H}_n(u) := P^*(\hat{p}_n^* \leq u)$ is not a consistent estimator of $H(u)$, invalidating $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$.

Remark 4.11 *A special case of the previous setup is a linear panel dynamic model, where $z_{it} = (y_{it}, x_{it}')'$ and x_{it} is a vector containing lags of y_{it} (Hahn and Kuersteiner, 2002). In this case, the plug-in modified p-value, $\tilde{p}_n$, based on the cross sectional pairs bootstrap can be implemented using any consistent estimator of B, as described above. However, we can also use a recursive bootstrap that exploits the linearity of the model to obtain an asymptotically valid standard bootstrap p-value, $\hat{p}_n$. The validity of $\hat{p}_n$ follows from the fact that the recursive bootstrap estimates B consistently, contrary to the pairs bootstrap (Gonçalves and Kaffo, 2015). In light of this, prepivoting $\hat{p}_n$ by computing a double bootstrap modified p-value $\tilde{p}_n = \hat{H}_n(\hat{p}_n)$ is not needed in this example, but it is still a valid alternative.*

5 Concluding remarks

Estimators with (asymptotic) bias arise in all areas of economics and statistics. Inference is challenging because the bias typically cannot be estimated.

In this paper, we have shown that in statistical problems involving bias terms that cannot be estimated, the bootstrap can be modified to provide asymptotically valid inferences. 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

[7] Since we reserve the notation $\hat{B}_n$ for the bootstrap-induced bias estimator (which is zero for the cross sectional pairs bootstrap), we use the notation $\tilde{B}_n$ to denote any consistent estimator of B in this setup. For instance, $\tilde{B}_n$ could be the plug-in estimator proposed by Hahn and Kuersteiner (2011), which is based on a closed-form expression of B_1. Another option is the half-split panel jackknife estimator of Dhaene and Jochmans (2015).

form of a bound on the bias of the estimator in question. These bounds appear, e.g., 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.

Appendix

A Proofs of main results

Proof of Theorem 3.1. First notice that $\hat{p}_n$ and $G_\gamma(T_n - \hat{B}_n)$ have the same asymptotic distribution because Assumption 2(i) and continuity of G_γ imply that

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

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

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

by the continuous mapping theorem using continuity of G_γ. Since $\xi_1 - \xi_2$ has continuous cdf F_ϕ, it holds that $\xi_1 - \xi_2 \sim F_\phi^{-1}(U_{[0,1]})$, which completes the proof.

Proof of Theorem 3.2. To prove this result, recall that $\hat{H}_n(u) = P^*(\hat{p}_n^* \leq u)$ and $H_n(u) = P(\hat{p}_n \leq u)$, where $H_n(u) \to H(u) = F_\phi(G_\gamma^{-1}(u))$ uniformly in $u \in \mathbb{R}$, since H is a continuous distribution function by Assumptions 1 and 2. We have that

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

where $G_\gamma(F_\phi^{-1}(U_{[0,1]}))$ is a random variable whose distribution function is $H(u)$. Hence,

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

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

B Additional proofs

B.1 Proof of results from Section 2

Derivation of $v_{1,n}^2$. It suffices to note that $T_n - B_n$ satisfies

$$T_n - B_n = (1-\omega)S_{xx}^{-1}, \omega S_{xx.z}^{-1})n^{1/2} \begin{pmatrix} S_{x\varepsilon} \\ S_{x\varepsilon.z} \end{pmatrix} \sim ((1-\omega)S_{xx}^{-1}, \omega S_{xx.z}^{-1})N\left(0, \begin{pmatrix} S_{xx} & S_{xx.z} \\ S_{xx.z} & S_{xx.z} \end{pmatrix}\right),$$

which implies that $v_{1,n}^2 = \omega^2 S_{xx.z}^{-1} + (1-\omega^2)S_{xx}^{-1}$, which can be written as

$$v_{1,n}^2 = \frac{1}{S_{xx}} \frac{1-(1-\omega^2)\hat{\rho}_{xz}^2}{1-\hat{\rho}_{xz}^2} = \frac{1}{S_{xx}}\left(1 + \omega^2 \frac{\hat{\rho}_{xz}^2}{1-\hat{\rho}_{xz}^2}\right)$$

with $\hat{\rho}_{xz} = S_{xx}^{-1/2} S_{xz} S_{zz}^{-1/2}$.

Derivation of $v_{2,n}^2$. Noting that $\hat{B}_n - B_n = (1-\omega) S_{xx}^{-1} S_{xz} S_{zz.x}^{-1} n^{1/2} S_{z\varepsilon.x}$, we obtain

$$v_{2,n}^2 = V(\hat{B}_n - B_n) = (1-\omega)^2 S_{xx}^{-2} S_{xz}^2 S_{zz.x}^{-1} = (1-\omega)^2 S_{xx}^{-1} \frac{\hat{\rho}_{xz}^2}{1-\hat{\rho}_{xz}^2}.$$

Derivation of $v_{d,n}^2$ and m_n. First, notice that $v_{d,n}^2 = v_{1,n}^2 + v_{2,n}^2 - 2v_{12,n}$, where $v_{12,n} = \mathrm{Cov}(T_n - B_n, \hat{B}_n - B_n)$. The latter term is given by

$$\begin{aligned}
v_{12,n} &= \mathrm{Cov}((1-\omega) S_{xx}^{-1} n^{1/2} S_{x\varepsilon} + \omega S_{xx.z}^{-1} n^{1/2} S_{x\varepsilon.z}, (1-\omega) S_{xx}^{-1} S_{xz} S_{zz.x}^{-1} n^{1/2} S_{z\varepsilon.x}) \\
&= E((1-\omega)^2 S_{xx}^{-1} S_{x\varepsilon} n^{1/2} S_{xx}^{-1} S_{xz} S_{zz.x}^{-1} n^{1/2} S_{z\varepsilon.x}) + E(\omega(1-\omega) S_{xx.z}^{-1} n^{1/2} S_{x\varepsilon.z} S_{xx}^{-1} S_{xz} S_{zz.x}^{-1} n^{1/2} S_{z\varepsilon.x}) \\
&= \omega\,(1-\omega)\, S_{xx.z}^{-1} S_{xx}^{-1} S_{xz}^2 S_{zz.x}^{-1} \left(\rho_{xz}^2 - 1\right),
\end{aligned}$$

where we used the facts that $E(S_{x\varepsilon} S_{z\varepsilon.x}) = 0$ and $E(n S_{x\varepsilon.z} S_{z\varepsilon.x}) = S_{xz}(\rho_{xz}^2 - 1)$. Because

$$S_{xx.z}^{-1} S_{xx}^{-1} S_{xz}^2 S_{zz.x}^{-1} = \frac{S_{xz}^2}{S_{xx}(S_{zz} - S_{zx}^2 S_{xx}^{-1})(S_{xx} - S_{zx}^2 S_{zz}^{-1})} = \frac{\rho_{zx}^2}{S_{xx}(1-\rho_{zx}^2)^2},$$

we have $v_{12,n} = -\omega(1-\omega) S_{xx}^{-1} \rho_{zx}^2 (1-\rho_{zx}^2)^{-1}$, and it then follows that

$$\begin{aligned}
v_{d,n}^2 = v_{1,n}^2 + v_{2,n}^2 - 2v_{12,n} &= \frac{1}{S_{xx}}\left(1 + \omega^2 \frac{\hat{\rho}_{xz}^2}{1-\hat{\rho}_{xz}^2}\right) + (1-\omega)^2 \frac{1}{S_{xx}} \frac{\hat{\rho}_{xz}^2}{1-\hat{\rho}_{xz}^2} + 2\omega(1-\omega) \frac{\rho_{zx}^2}{S_{xx}(1-\rho_{zx}^2)} \\
&= \frac{1}{S_{xx}} \frac{1}{1-\hat{\rho}_{xz}^2} (1 - \hat{\rho}_{xz}^2 + \omega^2 \hat{\rho}_{xz}^2 + (1-\omega)^2 \hat{\rho}_{xz}^2 + 2\omega(1-\omega)\rho_{zx}^2) = \frac{1}{S_{xx}} \frac{1}{1-\hat{\rho}_{xz}^2}.
\end{aligned}$$

Finally,

$$\begin{aligned}
m_n^2 = v_{d,n}^2 / v_{1,n}^2 &= \frac{1}{S_{xx}} \frac{1}{1-\hat{\rho}_{xz}^2} \left(S_{xx}^{-1}\left(1 + \omega^2 \frac{\hat{\rho}_{xz}^2}{1-\hat{\rho}_{xz}^2}\right)\right)^{-1} \\
&= \frac{1}{1-\hat{\rho}_{xz}^2} \left(\frac{1 - \hat{\rho}_{xz}^2 + \omega^2 \hat{\rho}_{xz}^2}{1-\hat{\rho}_{xz}^2}\right)^{-1} = \left(\frac{1}{1-(1-\omega^2)\hat{\rho}_{xz}^2}\right).
\end{aligned}$$

B.2 Proofs of the results in Section 4.2

Proof of Lemma 4.1. Under H_0 and our normalization conditions on the regressors,

$$n^{1/2} S_{xy} = n^{1/2} S_{xZ} \delta + n^{1/2} S_{x\varepsilon} = S_{xZ} c + n^{1/2} S_{x\varepsilon}$$

since $\delta = n^{-1/2} c$. Hence,

$$T_n = \frac{n^{1/2} S_{xy}}{\tilde{\sigma}_n} = \tilde{\sigma}_n^{-1} S_{xZ} c + \tilde{\sigma}_n^{-1} n^{1/2} S_{x\varepsilon},$$

so the result follows by Assumption OC if we can show that

$$\tilde{\sigma}_n^2 := n^{-1} \tilde{\varepsilon}' \tilde{\varepsilon} \to_p \sigma^2. \tag{B.1}$$

To prove (B.1) we write $\tilde{\varepsilon} = y - W\tilde{b}_n = \varepsilon - W(\tilde{b}_n - b)$, which implies that

$$\tilde{\sigma}_n^2 = n^{-1} \varepsilon' \varepsilon + S_{\varepsilon W}(b - \tilde{b}_n) + (b - \tilde{b}_n)' S_{W\varepsilon} + (b - \tilde{b}_n)' S_{WW} (b - \tilde{b}_n).$$

Under Assumption OC it holds that $n^{-1}\varepsilon'\varepsilon \to_p \sigma^2$, $S_{W\varepsilon} = O_p(n^{-1/2})$, and $S_{WW} = O_p(1)$. Thus, (B.1) follows if $\tilde{b}_n - b = o_p(1)$. To see that $\tilde{b}_n - b = o_p(1)$, note that

$$\tilde{b}_n = \hat{b}_n - \hat{\Delta}_n,$$

where $\hat{\Delta}_n := A_n\hat{\delta}_n$ and $A_n := (-S'_{xZ}, 0, I_p)'$. Under Assumption OC, $\hat{b}_n \to_p b$ and $\hat{\Delta}_n \to_p \Delta := (-\Sigma_{xZ}\delta, 0', \delta')' = O(n^{-1/2})$ because $\delta = O(n^{-1/2})$. Thus, $\tilde{b}_n \to_p b$, which proves (B.1).

Proof of Lemma 4.2. Proof of (i). Because

$$\tilde{\beta}^*_n = S^{-1}_{xx.Q} S_{xy^*.Q} = S_{xy^*.Q} = \hat{\beta}_n + S_{xZ}\hat{\delta}_n + S_{x\varepsilon^*},$$

we can write

$$T^*_n := (\tilde{\beta}^*_n - \hat{\beta}_n)/s^*(\tilde{\beta}^*_n) = (\tilde{\beta}^*_n - \hat{\beta}_n)/\tilde{\sigma}^*_n = S_{xZ}\tilde{\sigma}^{*-1}_n n^{1/2}\hat{\delta}_n + \tilde{\sigma}^{*-1}_n n^{1/2} S_{x\varepsilon^*},$$

and the result follows by proving that (a) $n^{1/2}S_{x\varepsilon^*} \xrightarrow{d^*}_p N(0, \sigma^2)$ and (b) $\tilde{\sigma}^{*2}_n \xrightarrow{p^*}_p \sigma^2$.

Proof of (a). By construction $\varepsilon^*|D_n \sim N(0, \hat{\sigma}^2_n I_n)$ and $n^{1/2}S_{x\varepsilon^*}|D_n \sim N(0, \hat{\sigma}^2_n)$. The result follows because $\hat{\sigma}^2_n \to_p \sigma^2$ (as in Lemma 4.1). Note that we could have used $\varepsilon^*|D_n \sim N(0, \tilde{\sigma}^2_n I_n)$ since $\tilde{\sigma}^2_n \to_p \sigma^2$ by (B.1).

Proof of (b). We write

$$\tilde{\varepsilon}^* = y^* - W\tilde{b}^*_n = \varepsilon^* - W(\tilde{b}^*_n - \hat{b}_n),$$

where $\tilde{b}^*_n := (\tilde{\beta}^*_n, \tilde{\eta}^{*\prime}_n, 0')'$, implying that

$$\tilde{\sigma}^{*2}_n := n^{-1}\tilde{\varepsilon}^{*\prime}\tilde{\varepsilon}^* = n^{-1}\varepsilon^{*\prime}\varepsilon^* + S_{\varepsilon^* W}(\hat{b}_n - \tilde{b}^*_n) + (\hat{b}_n - \tilde{b}^*_n)'S_{W\varepsilon^*} + (\hat{b}_n - \tilde{b}^*_n)'S_{WW}(\hat{b}_n - \tilde{b}^*_n). \quad \text{(B.2)}$$

Here, $n^{-1}\varepsilon^{*\prime}\varepsilon^* \xrightarrow{p^*}_p \sigma^2$ because $E^*(n^{-1}\varepsilon^{*\prime}\varepsilon^*) = \tilde{\sigma}^2_n \to_p \sigma^2$ and $\mathrm{Var}^*(n^{-1}\varepsilon^{*\prime}\varepsilon^*) = n^{-1}2\tilde{\sigma}^4_n \to_p 0$. The result then follows by showing that the remaining terms on the right-hand side of (B.2) are $o_{p^*}(1)$, in probability. First, note that $S_{\varepsilon^* W} = O_{p^*}(n^{-1/2})$, in probability, by part (a), whereas $S_{WW} = O_p(1)$ by Assumption OC. Thus, part (b) follows by showing that $\hat{b}_n - \tilde{b}^*_n = o_{p^*}(1)$, in probability. To show this, we write

$$\tilde{b}^*_n = \hat{b}^*_n - \hat{\Delta}^*_n,$$

where $\hat{\Delta}^*_n := A_n\hat{\delta}^*_n$ with $A_n := (-S'_{xZ}, 0, I_p)'$. Letting $\hat{\Delta}_n := A_n\hat{\delta}_n$, as in the proof of Lemma 4.1, yields

$$\tilde{b}^*_n - \hat{b}_n = (\hat{b}^*_n - \hat{b}_n) + (\hat{\Delta}^*_n - \hat{\Delta}_n) - \hat{\Delta}_n.$$

Under Assumption OC,

$$\begin{aligned}
\hat{b}^*_n - \hat{b}_n &= S^{-1}_{WW}S_{W\varepsilon^*} = o_{p^*}(1), \text{ in probability,} \\
\hat{\Delta}^*_n - \hat{\Delta}_n &= A_n(\hat{\delta}^*_n - \hat{\delta}_n) = o_{p^*}(1), \text{ in probability, and} \\
\hat{\Delta}_n &= A_n\delta + A_n(\hat{\delta}_n - \delta) = O_p(n^{-1/2}) = o_p(1),
\end{aligned}$$

since $\delta = O(n^{-1/2})$, which shows that $\hat{b}_n - \tilde{b}^*_n = o_{p^*}(1)$, in probability.

Proof of (ii). Note that we can write

$$\left(\begin{array}{c} T_n - B \\ \hat{B}_n - B \end{array}\right) = \sigma^{-1}\left(\begin{array}{c} n^{1/2}S_{x\varepsilon} \\ \Sigma_{xZ}n^{1/2}(\hat{\delta}_n - \delta) \end{array}\right) + o_p(1),$$

where

$$n^{1/2}(\hat{\delta}_n - \delta) = S_{ZZ.x}^{-1}n^{1/2}S_{Z\varepsilon.x} = S_{ZZ.x}^{-1}(n^{1/2}S_{Z\varepsilon} - S_{Zx}n^{1/2}S_{x\varepsilon}).$$

By Assumption OC, it follows that

$$n^{1/2}(\hat{\delta}_n - \delta) = \Sigma_{ZZ.x}^{-1}(n^{1/2}S_{Z\varepsilon} - \Sigma_{Zx}n^{1/2}S_{x\varepsilon}) + o_p(1),$$

which implies that

$$\left(\begin{array}{c} T_n - B \\ \hat{B}_n - B \end{array}\right) = \left(\begin{array}{cc} 1 & 0 \\ -\Sigma_{xZ}\Sigma_{ZZ.x}^{-1}\Sigma_{Zx} & \Sigma_{xZ}\Sigma_{ZZ.x}^{-1} \end{array}\right)\left(\begin{array}{c} \sigma^{-1}n^{1/2}S_{x\varepsilon} \\ \sigma^{-1}n^{1/2}S_{Z\varepsilon} \end{array}\right) + o_p(1).$$

The result then follows by direct application of Assumption OC.

Proof of Lemma 4.3. This result follows by arguments similar to those used in the proof of Lemma 4.2.

Proof of (i). We write

$$T_n^{**} := (\tilde{\beta}_n^{**} - \hat{\beta}_n^{*})/s(\tilde{\beta}_n^{**}) = S_{xZ}\tilde{\sigma}_n^{**-1}n^{1/2}\hat{\delta}_n^{*} + \tilde{\sigma}_n^{**-1}n^{1/2}S_{x\varepsilon^{**}},$$

using the facts that $\tilde{\beta}_n^{**} - \hat{\beta}_n^{*} = S_{xZ}\hat{\delta}_n^{*} + S_{x\varepsilon^{**}}$ and $s^2(\tilde{\beta}_n^{**}) = n^{-1}\tilde{\sigma}_n^{**2}$. We then show that (a) $n^{1/2}S_{x\varepsilon^{**}} \xrightarrow{d^{**}}_{p^*} N\left(0,\sigma^2\right)$ and (b) $\tilde{\sigma}_n^{**2} \xrightarrow{p^{**}}_{p^*} \sigma^2$, both in probability. Proof of (a): By construction, $n^{1/2}S_{x\varepsilon^{**}}|\{D_n^*, D_n\} \sim N(0,\tilde{\sigma}_n^{*2})$, and the result follows because $\tilde{\sigma}_n^{*2} \xrightarrow{p^*}_p \sigma^2$ as shown in Lemma 4.2. Proof of (b): As in the proof of Lemma 4.2, we can write

$$\tilde{\sigma}_n^{**2} = n^{-1}\varepsilon^{**\prime}\varepsilon^{**} + S_{\varepsilon^{**}W}(\hat{b}_n^{*} - \tilde{b}_n^{**}) + (\hat{b}_n^{*} - \tilde{b}_n^{**})'S_{W\varepsilon^{**}} + (\hat{b}_n^{*} - \tilde{b}_n^{**})'S_{WW}(\hat{b}_n^{*} - \tilde{b}_n^{**}). \quad \text{(B.3)}$$

Here it holds that $n^{-1}\varepsilon^{**\prime}\varepsilon^{**} \xrightarrow{p^{**}}_{p^*} \sigma^2$, in probability, since $E^{**}(n^{-1}\varepsilon^{**\prime}\varepsilon^{**}) = \tilde{\sigma}_n^{*2} \xrightarrow{p^*}_p \sigma^2$ and $\text{Var}^{**}(n^{-1}\varepsilon^{**\prime}\varepsilon^{**}) = n^{-1}2\tilde{\sigma}_n^{*4} \xrightarrow{p^*}_p 0$. The result then follows by showing that the remaining terms in (B.3) are $o_{p^{**}}(1)$. First, note that $S_{W\varepsilon^{**}} = O_{p^{**}}(n^{-1/2})$, in probability, by part (a) and $S_{WW} = O_p((1)$ by Assumption OC. Second, note that $\hat{b}_n^{*} - \tilde{b}_n^{**} = o_{p^{**}}(1)$, in probability, by completely analogous arguments to those in the proof of Lemma 4.2.

Proof of (ii). We write

$$\left(\begin{array}{c} T_n^{*} - \hat{B}_n \\ \hat{B}_n^{*} - \hat{B}_n \end{array}\right) = \tilde{\sigma}_n^{-1}\left(\begin{array}{c} n^{1/2}S_{x\varepsilon^{*}} \\ S_{xZ}n^{1/2}(\hat{\delta}_n^{*} - \hat{\delta}_n) \end{array}\right) + o_{p^*}(1),$$

in probability, where

$$n^{1/2}(\hat{\delta}_n^{*} - \hat{\delta}_n) = S_{ZZ.x}^{-1}n^{1/2}S_{Z\varepsilon^{*}.x} = S_{ZZ.x}^{-1}(n^{1/2}S_{Z\varepsilon^{*}} - S_{Zx}n^{1/2}S_{x\varepsilon^{*}}).$$

This implies that

$$\left(\begin{array}{c} T_n^{*} - \hat{B}_n \\ \hat{B}_n^{*} - \hat{B}_n \end{array}\right) = \left(\begin{array}{cc} 1 & 0 \\ -S_{xZ}S_{ZZ.x}^{-1}\Sigma_{Zx} & S_{xZ}S_{ZZ.x}^{-1} \end{array}\right)\left(\begin{array}{c} \tilde{\sigma}_n^{-1}n^{1/2}S_{x\varepsilon^{*}} \\ \tilde{\sigma}_n^{-1}n^{1/2}S_{Z\varepsilon^{*}} \end{array}\right) + o_{p^*}(1),$$

in probability. Since $\varepsilon^*|D_n \sim N(0, \hat{\sigma}_n^2)$, we have that

$$\hat{\sigma}_n^{-1}\begin{pmatrix} n^{1/2}S_{x\varepsilon^*} \\ n^{1/2}S_{Z\varepsilon^*}\end{pmatrix}|D_n \sim N(0,\hat{\Omega}_n), \quad \hat{\Omega}_n := \begin{pmatrix} 1 & S_{xZ} \\ S_{Zx} & I_p \end{pmatrix}.$$

The result then follows by Assumption OC.

B.3 Proofs of the results in Section 4.3

Proof of Lemma 4.4. To show that $Y_n := n^{1/2}A_n\varepsilon \to_d N(0, v^2)$, notice that

$$\begin{aligned} S_{x\varepsilon.Z_m} &= x'M_{Z_m}\varepsilon = n^{-1}(x'\varepsilon - x'Z_m(Z_m'Z_m)^{-1}R_m'Z'\varepsilon) \\ &= ([1, S_{xZ_m}(S_{Z_mZ_m})^{-1}R_m')S_{W\varepsilon} =: \hat{d}_m'S_{W\varepsilon}, \end{aligned}$$

where, under the stated assumptions,

$$\bar{d}_{M,n} := \sum_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}\hat{d}_m \to_p \bar{d}_M := \sum_{m=1}^{M}\omega_m\Sigma_{xx.Z_m}^{-1}d_m. \tag{B.4}$$

Hence,

$$Y_n = \sum_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}\hat{d}_m'n^{1/2}S_{W\varepsilon} = \bar{d}_{M,n}'(n^{1/2}S_{W\varepsilon}) \xrightarrow{d} N(0,v^2) \tag{B.5}$$

with $v^2 := \bar{d}_M'\Omega\bar{d}_M$.

Proof of Lemma 4.5. The first statement of the lemma follows straightforwardly using previous arguments. Next, note that

$$n^{1/2}(\hat{\delta}_n - \delta) = S_{ZZ.x}^{-1}S_{Z\varepsilon.x} = S_{ZZ.x}^{-1}(-S_{Zx}S_{xx}^{-1}, I_q)n^{1/2}S_{W\varepsilon},$$

from which it follows that

$$\begin{aligned} \hat{B}_n - B_n &= A_nZn^{1/2}(\hat{\delta}_n - \delta) = A_nZS_{ZZ.x}^{-1}(-S_{Zx}S_{xx}^{-1}, I_q)n^{1/2}S_{W\varepsilon} \\ &= \sum_{m=1}^{M}\omega_m S_{xx.Z_m}^{-1}S_{xZ.Z_m}S_{ZZ.x}^{-1}(-S_{Zx}S_{xx}^{-1}, I_q)n^{1/2}S_{W\varepsilon} =: \bar{b}_{M,n}'n^{1/2}S_{W\varepsilon}, \end{aligned}$$

where $\bar{b}_{M,n}$ satisfies

$$\bar{b}_{M,n}' \to_p \bar{b}_M' := \sum_{m=1}^{M}\omega_m\Sigma_{xx.Z_m}^{-1}\Sigma_{xZ.Z_m}\Sigma_{ZZ.x}^{-1}(-\Sigma_{Zx}\Sigma_{xx}^{-1}, I_q).$$

Hence, using (B.5),

$$\begin{aligned} \begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} &= \begin{pmatrix} \bar{d}_{M,n}' \\ \bar{b}_{M,n}' \end{pmatrix} n^{-1/2}W'\varepsilon \xrightarrow{d} N(0,V), \\ V &= \begin{pmatrix} \bar{d}_M'\Omega\bar{d}_M & \bar{d}_M'\Omega\bar{b}_M \\ \bar{b}_M'\Omega\bar{d}_M & \bar{b}_M'\Omega\bar{b}_M \end{pmatrix} =: \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}, \end{aligned}$$

where $v_{11} = v^2$, $v_{12} = v_{21} = \bar{d}_M'\Omega\bar{b}_M$, and

$$v_{22} := \bar{b}_M'\Omega\bar{b}_M = \left(\sum_{m=1}^{M} \omega_m \Sigma_{xx.Z_m}^{-1}\Sigma_{xZ.Z_m}\right)\Sigma_{ZZ.x}^{-1}\left(\sum_{m=1}^{M} \omega_m \Sigma_{xx.Z_m}^{-1}\Sigma_{xZ.Z_m}\right)'.$$

Proof of Lemma 4.6. First note that we can write $\tilde{\beta}_n^{**} = A_n y^{**}$, which implies that

$$\tilde{\beta}_n^{**} = A_n x\hat{\beta}_n^* + A_n Z\hat{\delta}_n^* + A_n\varepsilon^{**}.$$

It follows that

$$T_n^{**} := n^{1/2}(\tilde{\beta}_n^{**} - \hat{\beta}_n^*) = \hat{B}_n^* + n^{1/2}A_n\varepsilon^{**},$$

where $\hat{B}_n^* := n^{1/2}A_n Z\hat{\delta}_n^*$ and

$$\begin{aligned} Y_n^{**} := n^{1/2}A_n\varepsilon^{**} &= \sum_{m=1}^{M} \omega_m S_{xx.Z_m}^{-1} n^{1/2} S_{x\varepsilon^{**}.Z_m} \\ &= \sum_{m=1}^{M} \omega_m S_{xx.Z_m}^{-1}\hat{d}_m' n^{1/2} S_{W\varepsilon^{**}} = \bar{d}_{M,n}' n^{1/2} S_{W\varepsilon^{**}}. \end{aligned}$$

Since $\varepsilon^{**}|\{D_n, D_n^*\} \sim N(0, \hat{\sigma}_n^{*2} I_n)$, we have that

$$Y_n^{**}|\{D_n, D_n^*\} \sim N(0, \hat{v}_n^{*2}) \text{ or } \hat{v}_n^{*-1}Y_n^{**}|\{D_n, D_n^*\} \sim N(0,1),$$

where

$$\hat{v}_n^{*2} = \bar{d}_{M,n}'(\mathrm{Var}^*(n^{1/2}S_{W\varepsilon^{**}}))\bar{d}_{M,n} = \bar{d}_{M,n}'(\hat{\sigma}_n^{*2}S_{WW})\bar{d}_{M,n}.$$

Under our assumptions, $\hat{\sigma}_n^{*2} \xrightarrow{p^*}_p \sigma^2$, $S_{WW} \to_p \Sigma_{WW}$, and $\bar{d}_{M,n} \to_p d_M$. This implies that $\hat{v}_n^{*2} \xrightarrow{p^*}_p v^2$, and hence

$$\begin{aligned} Y_n^{**} &= (v + \hat{v}_n^* - v)\hat{v}_n^{*-1}Y_n^{**} = v\hat{v}_n^{*-1}Y_n^{**} + (\hat{v}_n^* - v)\hat{v}_n^{*-1}Y_n^{**} \\ &= v\hat{v}_n^{*-1}Y_n^{**} + o_{p^{**}}(1) \xrightarrow{d^{**}}_{p^*} N(0, v^2), \end{aligned}$$

in probability. This proves that $T_n^{**} - \hat{B}_n^* \xrightarrow{d^{**}}_{p^*} N(0, v^2)$, in probability. For the joint convergence of $(T_n^* - \hat{B}_n^*, \hat{B}_n^* - \hat{B}_n)'$, we write

$$n^{1/2}(\hat{\delta}_n^* - \hat{\delta}_n) = S_{ZZ.x}^{-1}S_{Z\varepsilon^*.x} = S_{ZZ.x}^{-1}(-S_{Zx}S_{xx}^{-1}, I_q)n^{1/2}S_{W\varepsilon^*},$$

from which it follows that

$$\begin{pmatrix} T_n^* - \hat{B}_n \\ \hat{B}_n^* - \hat{B}_n \end{pmatrix} = \begin{pmatrix} \bar{d}_{M,n}' \\ \bar{b}_{M,n}' \end{pmatrix} n^{1/2}S_{W\varepsilon^*}|\{D_n, D_n^*\} \sim N(0, \hat{V}_n)$$

since $n^{1/2}S_{W\varepsilon^*}|\{D_n, D_n^*\} \sim N(0, \hat{\Omega}_n)$. Here, $\hat{\Omega}_n := \hat{\sigma}_n^2 S_{WW}$ because $\varepsilon^*|\{D_n, D_n^*\} \sim N(0, \hat{\sigma}_n^2 I_n)$. This implies that

$$\hat{V}_n = \begin{pmatrix} \bar{d}_{M,n}'\hat{\Omega}_n\bar{d}_{M,n} & \bar{d}_{M,n}'\hat{\Omega}_n\bar{b}_{M,n} \\ \bar{b}_{M,n}'\hat{\Omega}_n\bar{d}_{M,n} & \bar{b}_M'\hat{\Omega}_n\bar{b}_{M,n} \end{pmatrix} =: \begin{pmatrix} \hat{v}_{11,n} & \hat{v}_{12,n} \\ \hat{v}_{21,n} & \hat{v}_{22,n} \end{pmatrix}$$

is such that $\hat{V}_n \to_p V$, from which the desired result follows.

Proof of Lemma 4.8. We first prove that

$$S_{W^*W^*} - S_{WW} \xrightarrow{p^*}_p 0, \tag{B.6}$$

$$S_n^* := \begin{pmatrix} n^{1/2} S_{x^*\varepsilon^*} \\ n^{1/2}(S_{x^*z^*} - S_{xz}) \\ n^{1/2}(S_{x^*x^*} - S_{xx}) \end{pmatrix} \xrightarrow{d^*}_p N(0, \Sigma_S), \quad \Sigma_S = \begin{pmatrix} \sigma^2 \Sigma_{xx} & 0 \\ 0 & \Sigma_r \end{pmatrix}. \tag{B.7}$$

Here, (B.6) follows by straightforward application of Chebyshev's LLN.

To prove (B.7), we first compute the mean and variance of S_n^*. Note that the mean of S_n^* is zero by construction; for example, $E^*(n^{1/2} S_{x^*\varepsilon^*}) = n^{-1/2} \sum_{t=1}^n E^*(x_t^* \varepsilon_t^*) = n^{1/2} S_{x\hat{\varepsilon}} = 0$ by the OLS first-order condition. In addition,

$$\mathrm{Var}^*(n^{1/2} S_{x^*\varepsilon^*}) = n^{-1} \sum_{t=1}^n E^*(x_t^{*2} \varepsilon_t^{*2}) = n^{-1} \sum_{t=1}^n x_t^2 \hat{\varepsilon}_t^2 \xrightarrow{p} \sigma^2 \Sigma_{xx}$$

under Assumptions MA and MA_2. Similarly, letting

$$\begin{pmatrix} n^{1/2}(S_{x^*z^*} - S_{xz}) \\ n^{1/2}(S_{x^*x^*} - S_{xx}) \end{pmatrix} = n^{1/2}(S_{x^*W^*} - S_{xW}),$$

we find that

$$\begin{aligned} \mathrm{Var}^*(n^{1/2}\,(S_{x^*W^*} - S_{xW})) &= n^{-1} \sum_{t=1}^n (x_t w_t - E^*(x_t^* w_t^*))(x_t w_t - E^*(x_t^* w_t^*))' \\ &= n^{-1} \sum_{t=1}^n x_t^2 w_t w_t' - S_{xW} S_{Wx} \xrightarrow{p} \Sigma_r - \Sigma_{xW} \Sigma_{Wx}. \end{aligned}$$

Note also that the covariance between $n^{1/2} S_{x^*\varepsilon^*}$ and $n^{1/2}(S_{x^*W^*} - S_{xW})$ is zero because

$$\begin{aligned} E^*(n S_{x^*\varepsilon^*} S_{x^*W^*}) &= n^{-1} E^* \left(\sum_{t=1}^n x_t^* \varepsilon_t^* \sum_{s=1}^n x_s^* w_s^* \right) = n^{-1} E^* \left(\sum_{t=1}^n x_t^{*2} w_t^* \varepsilon_t^* \right) \\ &= E^*(x_t^{*2} w_t^* \varepsilon_t^*) = n^{-1} \sum_{t=1}^n x_t^2 w_t \hat{\varepsilon}_t \xrightarrow{p} 0 \end{aligned}$$

by Assumption MA_2(ii). Thus, we have shown that $E^*(S_n^*) = 0$ and $E^*(S_n^* S_n^{*\prime}) \to_p \Sigma_S$. The result (B.7) now follows because the stated moment conditions imply the Lindeberg condition by standard arguments.

Next we can write

$$T_n^* - \hat{B}_n = n^{1/2} S_{x^*x^*}^{-1} S_{x^*\varepsilon^*} + B_n^* - \hat{B}_n,$$

where

$$B_n^* - \hat{B}_n = (S_{x^*x^*}^{-1} S_{x^*z^*} - S_{xx}^{-1} S_{xz}) n^{1/2} \hat{\delta}_n.$$

Adding and subtracting appropriately, we can write this difference as

$$B_n^* - \hat{B}_n = n^{1/2}(S_{x^*x^*}^{-1} S_{x^*z^*} - S_{xx}^{-1} S_{xz})\delta + (S_{x^*x^*}^{-1} S_{x^*z^*} - S_{xx}^{-1} S_{xz}) n^{1/2}(\hat{\delta}_n - \delta),$$

where $n^{1/2}(\hat{\delta}_n - \delta)$ is $O_p(1)$ by a CLT and $S_{x^*x^*}^{-1} S_{x^*z^*} - S_{xx}^{-1} S_{xz} = o_p^*(1)$, in probability, by (B.6).

The first term in $B_n^* - \hat{B}_n$ can be written as

$$\begin{aligned}
&S_{x^*x^*}^{-1} n^{1/2}(S_{x^*z^*} - S_{xz})\delta - S_{x^*x^*}^{-1} S_{xx}^{-1}(S_{x^*x^*} - S_{xx}) n^{1/2} S_{xz}\delta \\
&= \delta(\Sigma_{xx}^{-1}, -\Sigma_{xx}^{-2}\Sigma_{xz}) \begin{pmatrix} n^{1/2}(S_{x^*z^*} - S_{xz}) \\ n^{1/2}(S_{x^*x^*} - S_{xx}) \end{pmatrix} + o_p^*(1),
\end{aligned}$$

in probability, by application of (B.6) and Assumption MA(ii). It follows that

$$\begin{aligned}
T_n^* - \hat{B}_n &= S_{x^*x^*}^{-1} n^{1/2} S_{x\varepsilon}^* + \delta(\Sigma_{xx}^{-1}, -\Sigma_{xx}^{-2}\Sigma_{xz}) \begin{pmatrix} n^{1/2}(S_{x^*z^*} - S_{xz}) \\ n^{1/2}(S_{x^*x^*} - S_{xx}) \end{pmatrix} + o_p^*(1) \\
&= (\Sigma_{xx}^{-1}, \Sigma_{xx}^{-1}\delta, -\Sigma_{xx}^{-2}\Sigma_{xz}\delta) S_n^* + o_p^*(1),
\end{aligned}$$

in probability. The required result now follows from (B.7) because

$$\begin{aligned}
&(\Sigma_{xx}^{-1}, \Sigma_{xx}^{-1}\delta, -\Sigma_{xx}^{-2}\Sigma_{xz}\delta) \begin{pmatrix} \Sigma_s & 0 \\ 0 & \Sigma_r \end{pmatrix} (\Sigma_{xx}^{-1}, \Sigma_{xx}^{-1}\delta, -\Sigma_{xx}^{-2}\Sigma_{xz}\delta)' \\
&= \Sigma_{xx}^{-1}\Sigma_s\Sigma_{xx}^{-1} + d_r(\delta)'\Sigma_r d_r(\delta) = v^2 + \kappa^2.
\end{aligned}$$

B.4 Proofs of the results in Section 4.4

Proof of Lemma 4.9 and derivation of (4.12). First notice that, since $c_n n^{-1} \to c_0$, under Assumption RE we have that $n^{1/2} S_{x\varepsilon} \to_d N(0, \sigma^2\Sigma_{xx})$ and hence

$$\begin{aligned}
\begin{pmatrix} T_n - B_n \\ \hat{B}_n - B_n \end{pmatrix} &= (I_2 \otimes g'\tilde{S}_{xx}^{-1}) \begin{pmatrix} I_m \\ -n^{-1}c_n S_{xx}^{-1} \end{pmatrix} n^{1/2} S_{x\varepsilon} \xrightarrow{d} (I_2 \otimes g'\tilde{\Sigma}_{xx}^{-1}) N\left(0, \sigma^2 \begin{pmatrix} \Sigma_{xx} & -c_0 I \\ -c_0 I & c_0^2 \Sigma_{xx}^{-1} \end{pmatrix}\right) \\
&\sim N(0, V), \quad V = \sigma^2 \begin{pmatrix} g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1} g & -c_0 g'\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1} g \\ -c_0 g'\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1} g & c_0^2 g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1} g \end{pmatrix}. \qquad \text{(B.8)}
\end{aligned}$$

This immediately implies that $m^2 = v_{11} + v_{22} - 2v_{12}$ is given by

$$m^2 = \frac{g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1} g + 2c_0 g'\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1} g + c_0^2 g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}^{-1}\tilde{\Sigma}_{xx}^{-1} g}{g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1} g}. \qquad \text{(B.9)}$$

The numerator of m^2 in (B.9) can be written as

$$g'\tilde{\Sigma}_{xx}^{-1}(\Sigma_{xx} + 2c_0 I_m + c_0^2\Sigma_{xx}^{-1})\tilde{\Sigma}_{xx}^{-1} g = g'\tilde{\Sigma}_{xx}^{-1}(\tilde{\Sigma}_{xx}\Sigma_{xx}^{-1}\tilde{\Sigma}_{xx})\tilde{\Sigma}_{xx}^{-1} g = g'\Sigma_{xx}^{-1} g$$

and hence (4.12) follows.

Proof of Lemma 4.10. Note that $T_n^* - \hat{B}_n = \xi_{1,n}^* + B_n^* - \hat{B}_n$, where

$$\begin{aligned}
B_n^* - \hat{B}_n &= -c_n n^{-1/2} g'\tilde{S}_{x^*x^*}^{-1}\hat{\theta}_n + c_n n^{-1/2} g'\tilde{S}_{xx}^{-1}\hat{\theta}_n \\
&= -c_n n^{-1} g'(\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1}) n^{1/2}(\hat{\theta}_n - \theta_0) - c_n n^{-1} g'(\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1})\delta = O_p^*(\left\|\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1}\right\|),
\end{aligned}$$

in probability, such that $B_n^* - \hat{B}_n \xrightarrow{p^*}_p 0$ if $\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1} \xrightarrow{p^*}_p 0$. Because $\left\|\tilde{S}_{xx}^{-1}\right\| = O(1)$, which holds under the stated assumptions, it follows that $\left\|\tilde{S}_{x^*x^*}^{-1} - \tilde{S}_{xx}^{-1}\right\|$ has the same rate as $\left\|\tilde{S}_{x^*x^*} - \tilde{S}_{xx}\right\|$. Thus, $\tilde{S}_{x^*x^*} - \tilde{S}_{xx} = S_{x^*x^*} - S_{xx} = n^{-1}\sum_{t=1}^n (x_t^* x_t^{*\prime} - E^*(x_t^* x_t^{*\prime})) \xrightarrow{p^*}_p 0$ by a straightforward application of Chebyshev's LLN using that $\max_t x_t' x_t = o(n^{1/2})$.

The proof is completed by showing that $\xi^*_{1,n}$ satisfies the bootstrap CLT. By the above results it holds that $\xi^*_{1,n} = n^{1/2}g'\tilde{\Sigma}_{xx}^{-1}S_{x^*\varepsilon^*} + o_p^*(1)$, in probability, so it is only required to analyze the term $n^{1/2}g'\tilde{\Sigma}_{xx}^{-1}S_{x^*\varepsilon^*} = n^{1/2}S_{\tilde{x}^*\varepsilon^*}$, where $\tilde{x}^*_t := g'\tilde{\Sigma}_{xx}^{-1}x^*_t$. First, we have $E^*(n^{1/2}S_{x^*\varepsilon^*}) = n^{1/2}S_{x\hat{\varepsilon}} = 0$. Second,

$$\begin{aligned}\mathrm{Var}^*(n^{1/2}S_{\tilde{x}^*\varepsilon^*}) &= n^{-1}\sum_{t=1}^{n}\tilde{x}_t^2\hat{\varepsilon}_t^2 = n^{-1}\sum_{t=1}^{n}\tilde{x}_t^2(\hat{\varepsilon}_t^2 - \sigma^2 + \sigma^2)\\ &= \sigma^2 g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g + n^{-1}\sum_{t=1}^{n}\tilde{x}_t^2(\varepsilon_t^2 - \sigma^2) + o_p(1).\end{aligned}$$

Because ε_t is i.i.d. and $\tilde{x}_t^2$ non-stochastic, a sufficient condition for $n^{-1}\sum_{t=1}^n \tilde{x}_t^2(\varepsilon_t^2 - \sigma^2) \to_p 0$ is that $\lambda_{\min}(\sum_{t=1}^n \tilde{x}_t^2) \to \infty$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of the argument, and this is implied by $n^{-1}\sum_{t=1}^n \tilde{x}_t^2 \to g'\tilde{\Sigma}_{xx}^{-1}\Sigma_{xx}\tilde{\Sigma}_{xx}^{-1}g > 0$.

Third, we check the Lindeberg's condition, where we set $s_n^2 := nS_{xx}$. It holds that

$$\begin{aligned}\frac{1}{s_n^2}\sum_{t=1}^{n}E^*(\tilde{x}_t^{*2}\varepsilon_t^{*2}\mathbb{I}_{\{|\tilde{x}_t^*\varepsilon_t^*|>\epsilon s_n\}}) &= \frac{1}{S_{xx}}E^*(\tilde{x}_t^{*2}\varepsilon_t^{*2}\mathbb{I}_{\{(\tilde{x}_t^*\varepsilon_t^*)^2>\epsilon^2 nS_{xx}\}})\\ &\le \frac{1}{\epsilon^2 nS_{xx}^2}E^*(\tilde{x}_t^{*4}\varepsilon_t^{*4})\\ &= \frac{1}{\epsilon^2 n^2 S_{xx}^2}\sum_{t=1}^{n}\tilde{x}_t^4\hat{\varepsilon}_t^4 \le \frac{n^{-1}\max_t \tilde{x}_t^4}{\epsilon^2 S_{xx}^2}\frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^4 \xrightarrow{p} 0\end{aligned}$$

because $n^{-1}\max_t \tilde{x}_t^4 = o(1)$ and ε_t has bounded fourth-order moment.

Proof of Lemma 4.11. The proof follows closely the proofs of Lemmas 4.9 and 4.10 and is omitted for brevity.

References

Arcones, M.A., and E. Gine (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 25, 457–481.

Beran, R. (1987). Prepivoting to reduce level error in confidence sets. *Biometrika* 74, 457–468.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687–97.

Calonico, S., M.D. Cattaneo, and M.H. Farrell (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association* 113, 767–779.

Calonico, S., M.D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82, 2295–2326.

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

Cavaliere, G., and I. Georgiev (2020). Inference under random limit bootstrap measures. *Econometrica* 88, 2547–2574.

Chang, J., and P. Hall (2015). Double-bootstrap methods that use a single double-bootstrap simulation. *Biometrika* 102, 203–214.

Chatterjee, A., and S.N. Lahiri (2010). Asymptotic properties of the residual bootstrap for lasso estimators. *Proceedings of the American Mathematical Society* 138, 4497–4509.

Chatterjee, A., and S.N. Lahiri (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association* 106, 608–625.

Cornea-Madeira, A., and R. Davidson (2015). A parametric bootstrap for heavy-tailed distributions. *Econometric Theory* 31, 449–470.

Dhaene, G., and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* 82, 991–1030.

Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of American Statistical Association* 78, 316–331.

Frank, I.E., and J.H. Friedman (1993). A statistical view of some chemometrics regression tools. *Technometrics* 35, 109–135.

Giglio, S., and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy* 129, 1947–1990.

Goncalves, S., and M. Kaffo (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics* 186, 407–426.

Hahn, J., and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639–1657.

Hahn, J., and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.

Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics* 14, 1431–1452.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag: Berlin.

Hansen, B.E. (2007). Least squares model averaging. *Econometrica* 75, 1175–1189.

Higgins, A., and K. Jochmans (2022). Bootstrap inference for fixed-effect models. Working paper, Cambridge University.

Hjort, N., and G. Claeskens (2003). Frequentist model average estimators. *Journal of the American Statistical Association* 98, 879–899.

Horowitz, J.L. (2001). The bootstrap. In *Handbook of Econometrics* (Heckman, J.J., and E. Leamer, eds.), vol. 5, chp. 52, Elsevier: Amsterdam.

Hounyo, U. and K. Lahiri (2021). Estimating the variance of a combined forecast: Bootstrap-based approach. *Journal of Econometrics*, forthcoming.

Kaffo, M. (2014). Bootstrap inference for nonlinear dynamic panel data models with individual fixed effects. Manuscript.

Kallenberg, O. (1997). *Foundations of Modern Probability.* Springer: New York.

Kiefer, N.M., and T. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Annals of Statistics* 17, 1168–1175.

Knight, K., and W. Fu (2000). Asymptotics for lasso-type estimators. *Annals of Statistics* 28, 1356–1378.

Leeb, H., and B.M. Potscher (2005). Model selection and inference: facts and fiction. *Econometric Theory* 21, 21–59.

Li, C., and U. Muller (2021). Linear regression with many controls of limited explanatory power. *Quantitative Economics* 12, 405–442.

Liu, C.-A. (2015). Distribution theory of the least squares averaging estimator. *Journal of Econometrics* 186, 142–159.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37, 187–204.

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

Shao, X., and D.N. Politis (2013). Fixed b subsampling and the block bootstrap: Improved confidence sets based on p-value calibration. *Journal of the Royal Statistical Society B* 75, 161–184.