An Improved Fast Double Bootstrap

by

Russell Davidson

Department of Economics and CIREQ
McGill University
Montreal, Quebec, Canada
H3A 2T7

Aix-Marseille Université
CNRS, EHESS, AMSE
13205 Marseille cedex 01, France

email: **russell.davidson@mcgill.ca**

and

Andrea Monticini

Catholic University
Via Necchi 5
20123 Milan, Italy

email: **andrea.monticini@gmail.com**

Abstract

The fast double bootstrap can improve considerably on the single bootstrap when the bootstrapped statistic is approximately independent of the bootstrap DGP. This is because, among the approximations that underlie the fast double bootstrap (FDB), is the assumption of such independence. In this paper, use is made of a discrete formulation of bootstrapping in order to develop a conditional version of the FDB, which makes use of the joint distribution of a statistic and its bootstrap counterpart, rather than the full joint distribution of the statistic and the bootstrap data-generating process (DGP), which is available only by means of a simulation as costly as the full double bootstrap. Simulation evidence shows that the conditional FDB can greatly improve on the performance of the FDB when the statistic and the bootstrap DGP are far from independent, while giving similar results in cases of near independence.

Keywords: Bootstrap inference, fast double bootstrap, discrete model, conditional fast double bootstrap

JEL codes: C12, C22, C32

This research was supported by the Canada Research Chair program (Chair in Economics, McGill University) and by grants from the Fonds de Recherche du Québec - Société et Culture. This work was also supported by the French National Research Agency Grant ANR-17-EURE-0020

February, 2023

1. Introduction

The double bootstrap proposed by Beran (1988) can permit more reliable inference than the ordinary or single bootstrap, by means of bootstrapping the single bootstrap P value. If the statistic that is bootstrapped is pivotal for the null model under test, then bootstrap inference is exact up to simulation randomness. Similarly, an approximately pivotal statistic, for instance an asymptotically pivotal statistic, benefits when bootstrapped from bootstrap refinements; see Hall (1992). Beran shows how the bootstrap P value obtained by bootstrapping any statistic, approximately pivotal or not, is closer to being pivotal than the original statistic, and thus benefits from asymptotic refinements. Further iterations are expected in general to improve on the double bootstrap, although their implementation is usually extremely computationally costly. The idea is that, if the (finite-sample) distribution of a statistic is known, it can yield exact inference, and iterated bootstraps provide better and better approximations to the distribution of the P value of the previous iteration.

The fast double bootstrap (FDB) studied in detail by Davidson and MacKinnon (2007) has been shown to improve the reliability of bootstrap inference in many practical situations; see among others Davidson and MacKinnon (2002), Lamarche (2004), Davidson (2006), Omtzigt and Fachin (2006), Ahlgren and Antell (2008), and Ouysse (2013). Like the double bootstrap, it makes use of an approximation to the distribution of the single bootstrap P value, but without introducing an additional layer of bootstrapping. Rather, at each step of the single layer of simulation, joint realisations are obtained of the ordinary bootstrap statistic, which we call τ, and of the bootstrap statistic, denoted τ^1, generated by the bootstrap data-generating process (DGP) that is distributed jointly with τ. Let the distribution function (CDF) of the statistic τ under some DGP μ be characterised by the function $R_0(\cdot,\mu)$, and let the CDF of the second-layer statistic τ^1 be $R^1(\cdot,\mu)$. Under the assumption that both statistics have absolutely continuous distributions on their support, let the uniquely defined quantile functions that are inverse to R_0 and R^1 be $Q_0(\cdot,\mu)$ and $Q^1(\cdot,\mu)$ respectively.

Let the bootstrap DGP that is generated jointly with τ from the data be denoted β. The bootstrap P value can then be written as $p_1 = R_0(\tau,\beta)$, where, without loss of generality and for notational convenience, the rejection region is presumed to be to the left of the statistic. We write $R_1(\cdot,\mu)$ for the CDF of p_1 under DGP μ. Beran's double bootstrap approximates this by $R_1(\cdot,\beta)$ at the cost of a nested layer of simulation. The FDB approximation of $R_1(\cdot,\mu)$ is $R_1^f(\cdot,\beta) \equiv R_0(Q^1(\cdot,\beta),\beta)$, so that a simulation estimate of R_1^f can be obtained using the simulated distributions of τ and τ^1. However, this approximation depends only on the *marginal* distributions. In this paper, we see how to make use of the *joint* distribution of τ and τ^1 to get a better approximation, and hence more reliable inference.

The FDB approximation relies on two assumptions. The first is that the statistic τ and the bootstrap DGP β are approximately (usually asymptotically) independent. This assumption is correct in many, but by no means all, commonly occurring situations. The other assumption, which we will not spell out in detail at this point, is reasonable more generally, and can usually be checked by an asymptotic argument.

The paper is laid out as follows. In Section 2, the maximum-entropy (ME) bootstrap of Vinod (2006) is discussed, and its failings noted. Analysis of these failings is undertaken by use of a discrete formulation of bootstrapping introduced by Davidson (2017b), which facilitates much analysis of the bootstrap by making it possible to obtain exact expressions for quantities without any asymptotic approximation. Some results are established concerning the joint distribution of the statistics τ and τ^1. These results are extended in Section 3, to show how the joint distribution of τ and β affects the bootstrap discrepancy, that is, the error in the rejection probability of a bootstrap test. A link is then forged to a diagnostic procedure proposed by Davidson (2017a), and it is shown that this procedure can detect possible over- or under-rejection by a bootstrap test. Then, in Section 4, the fast double bootstrap (FDB) of Davidson and MacKinnon (2007) is formulated in the discrete formulation used for the theoretical parts of this paper, and an exact expression for the CDF of the FDB P value given. However, the main contribution of this section is to modify the FDB by making explicit use of the joint distribution of τ and τ^1. This is an approximate version of a result in Davidson and MacKinnon (1999) that depends on the joint distribution of τ and β. Whereas that earlier result involves distributions that cannot be estimated without great computational cost, the modified FDB, called CFDB for *conditional* FDB, involves no more cost than the FDB itself. An explicit algorithm is given for the CFDB that is not restricted to the discrete formulation. Simulation evidence is provided in Section 5 to show that the CFDB gives results little different from those given by the FDB when τ and τ^1 are nearly independent, but that it greatly reduces the bootstrap discrepancy when they are dependent, as is the case with the ME bootstrap, and rivals the full double bootstrap in its reliability. Section 6 concludes, and discusses how the work of this paper can be extended in various directions.

2. Joint Distribution of the Statistic and the Bootstrap Statistic

A technique for bootstrapping time series, called the maximum-entropy bootstrap, was introduced by Vinod (2006) and further studied by Vinod and López-de Lacalle (2009). However, Bergamelli, Novotný, and Urga (2015) analysed the performance of the technique for a unit-root test and found that it was very poor indeed. They were able to show that this was due to the near perfect rank correlation of the test statistic (our τ) and the bootstrap statistics (realisations of our τ^1) generated by the ME bootstrap.

Davidson (2017a) proposed a diagnostic procedure for evaluating bootstrap performance, based on a simulation experiment. This was used in a working paper, Davidson and Monticini (2014), to examine the ME bootstrap for testing hypotheses in the context of a linear regression model with strongly serially correlated data. They found that it was characterised by serious under-rejection for all conventional levels, and over-rejection for levels greater than a half. Like Bergamelli *et al*, they also saw that statistics τ realised with serially correlated data and their ME bootstrap counterparts τ^1 were strongly positively correlated. In this section, it will be shown that these phenomena are not at all unrelated, and that, quite generally, correlation between a statistic and its bootstrap version aggravates the size distortion of a bootstrap test, in a way qualitatively similar to what was seen with the ME bootstrap.

The diagnostic procedure works as follows: for each of N replications, simulated data are drawn from the chosen DGP, and are then used to generate a realisation τ_i, $i = 1, \ldots, N$, of a test statistic, and, at the same time, a realisation of the corresponding bootstrap DGP, β_i say, which is then used to generate a single realisation τ_i^1 of the bootstrap statistic.

The marginal distributions of the variables τ and τ^1 of which the τ_i and τ_i^1 are realisations should, if the bootstrap is to perform satisfactorily, be similar: Davidson (2017a) suggests comparing kernel-density plots of the simulated statistics to see whether this is the case. The diagnostic that more particularly concerns us here is given by an OLS regression of the τ_i^1 on a constant and the τ_i. A significant coefficient for the regressor τ indicates correlation in the joint distribution of τ and τ^1, with the same sign as that of the estimated coefficient.

The discrete formulation

In order to analyse the consequences of a correlation of τ and τ^1, it is convenient to use the discrete formulation of bootstrapping developed in Davidson (2017b). It is only fair to warn readers that this formulation is unconventional, although it accurately reflects very general bootstrap procedures. Its advantage for the current study is that the discreteness makes it straightforward to derive tractable expressions for the joint distribution of τ and τ^1, without any need for limiting arguments or any asymptotic reasoning. In particular, there is no need for regularity conditions in order to derive results like the asymptotic validity of some bootstrap procedure. Indeed, the discrete approach is not designed to treat anything asymptotic at all. However, it is worth emphasising that it can encompass situations that in a more conventional approach would be considered instances of bootstrap failure, as well as situations in which the bootstrap would have desirable properties. It would be necessary to impose further structure on the model described below before any judgements of bootstrap success or failure could be made. Without additional structure, the model is limited to discussion of purely formal properties of bootstrapping in general, although these include bootstrap iteration, as shown in Davidson (2017b).

It may be thought that theoretical results derived from a discretised model are not much related to the usual case in which both the statistic and the bootstrap discrepancy vary over a continuous domain. In order to refute this idea, it is useful, first, to recall that computers work with discrete quantities, and, second, that nothing in our arguments prevents the discretisation from being arbitrarily fine, and so approximating the continuous setup arbitrarily well. The virtue of the discrete formulation is that it enables analysis that is perfectly rigorous within its own limitations, and with no appeal to asymptotic approximations. In fact, the sample size is a quantity that appears nowhere in the basic discrete model.

It is assumed that the statistic τ, in approximate P value form, can take on only the values π_i, $i = 0, 1, \ldots, n$, with

$$0 = \pi_0 < \pi_1 < \pi_2 < \ldots < \pi_{n-1} < \pi_n = 1.$$

For instance, if $n = 100$, with $\pi_i = i/100$, $i = 0, 1, \ldots, n$, P values would thereby be limited to integer percentages. Further, we assume that there are only m possible DGPs in the model that represents the null hypothesis. Thus the outcome space on which the random variables τ, the test statistic, and the bootstrap DGP β, are defined consists of just $m(n+1)$ points, labelled by two integer coordinates (i, j), $i = 0, 1, \ldots, n$, $j = 1, \ldots, m$. Any scalar- or vector-valued function of the coordinates (i, j) is a random variable defined on this outcome space.

The third component of a probability space is a probability measure. The probability space we use here is completely characterised by the discrete set of probabilities p_{kij}, $k, j = 1, \ldots, m$, $i = 0, 1, \ldots, n$, where, under the DGP indexed by k,

$$p_{kij} = \Pr{}_k\big[\tau = \pi_i \text{ and } \beta = j\big].$$

It follows for all $k = 1, \ldots, m$, that

$$\sum_{i=0}^{n}\sum_{j=1}^{m} p_{kij} = 1. \tag{1}$$

Make the following definitions for $k = 1, \ldots, m$ and $i = 0, \ldots, n$:

$$P_{ki} = \sum_{j=1}^{m} p_{kij}, \quad a_{kij} = \sum_{l=0}^{i-1} p_{klj}, \quad A_{ki} = \sum_{j=1}^{m} a_{kij}, \quad b_{kj} = \sum_{i=0}^{n} p_{kij}. \tag{2}$$

Clearly, P_{ki} is the probability under DGP k of the event $\tau = \pi_i$; a_{kij} the probability that simultaneously $\tau < \pi_i$ and $\beta = j$; A_{ki} the probability that $\tau < \pi_i$; b_{kj} the probability that $\beta = j$. It follows directly from (1) and (2) that, for $k = 1, \ldots, m$,

$$\sum_{j=1}^{m} b_{kj} = 1 \quad \text{and} \quad A_{ki} = \sum_{l=0}^{i-1} P_{kl}. \tag{3}$$

We begin by expressing the probability under DGP k that $\tau = \pi_i$ and $\tau^1 = \pi_l$. With probability p_{kij}, $\tau = \pi_i$ and $\beta = j$. If $\beta = j$, the probability that $\tau^1 = \pi_l$ is P_{jl}. Thus

$$P_k\big((\tau = \pi_i) \wedge (\tau^1 = \pi_l)\big) = \sum_{j=1}^{m} p_{kij} P_{jl}. \tag{4}$$

In addition, we have for the marginal distribution of τ^1 that

$$P_k(\tau^1 = \pi_l) = \sum_{i=0}^{n}\sum_{j=1}^{m} p_{kij} P_{jl} = \sum_{j=1}^{m} b_{kj} P_{jl}. \tag{5}$$

The following theorem shows that independence of τ and τ^1 follows from that of τ and β, and that the implication goes in the other direction as well under stronger regularity conditions.

Theorem 1

If, under DGP k, the statistic τ and the bootstrap DGP β are independent, then τ and τ^1 are also independent. The converse implication holds as well if the $m \times (n+1)$ matrix $\boldsymbol{P}$ with element jl given by P_{jl} has full row rank.

Proof:

Independence of τ and β under DGP k means that $p_{kij} = b_{kj}P_{ki}$, and so the joint probability (4) becomes

$$\sum_{j=1}^{m} b_{kj}P_{ki}P_{jl} = P_{ki}\sum_{j=1}^{m} b_{kj}P_{jl},$$

and this is the product of the probability P_{ki} that $\tau = \pi_i$, and the probability that $\tau^1 = \pi_l$, by (5). Thus independence of τ and β implies that of τ and τ^1.

Next, suppose that τ and τ^1 are independent under k. This implies that, for all $i, l = 0, \ldots, n$,

$$P_k(\tau^1 = \pi_l \bigm| \tau = \pi_i) = P_k(\tau^1 = \pi_l) = \sum_{j=1}^{m} b_{kj}P_{jl}.$$

The conditional probability above is equal to $\sum_{j=1}^{m} p_{kij}P_{jl}/P_{ki}$, and so the condition is equivalent to

$$\sum_{j=1}^{m} P_{jl}(b_{kj}P_{ki} - p_{kij}) = 0 \qquad \text{for all } i, l = 0, \ldots, n. \tag{6}$$

Now P_{jl} is element jl of the matrix $\boldsymbol{P}$, which is supposed to have full row rank. Thus (6) implies that, for all $i = 0, \ldots, n$ and $j = 1, \ldots, m$, $p_{kij} = b_{kj}P_{ki}$. This expresses the independence of τ and β under k. ■

Remarks:

The most interesting result of this theorem is the first. It is hard to derive any intuition about the full-rank condition. In some sense, it implies that the different DGPs in the model are sufficiently different. In addition, it is necessary for the full-rank condition that $m \leq n + 1$. In particular, if τ is an exact pivot for the model, then all the rows of $\boldsymbol{P}$ are the same.

In Davidson and MacKinnon (1999), it is shown that, if τ and β are asymptotically independent, the bootstrap benefits from an additional asymptotic refinement relative to cases without such approximate independence. The result of the present theorem requires exact independence, but applies exactly in finite samples.

The virtues and limitations of the discrete approach may be better appreciated by comparing the proof above of the first part of the theorem with the one given in the Appendix for the absolutely continuous case. This alternative proof illustrates how the discrete and absolutely continuous cases are related.

Corollary 1

If the statistic τ is a pivot for the model, so is τ^1, both τ and τ^1 have the same distribution, and the two are independent.

Proof:

If τ is a pivot, then $P_{ki} = P_i$ independent of k. From (5), for $k = 1, \ldots, m$,

$$P_k(\tau^1 = \pi_l) = \sum_{j=1}^{m} b_{kj} P_{jl} = P_l \sum_{j=1}^{m} b_{kj} = P_l,$$

where the last equality follows from (3). This implies that τ^1 has the same distribution as τ for all k, and so is a pivot. The joint distribution (4) becomes

$$P_k\big((\tau = \pi_i) \wedge (\tau^1 = \pi_l)\big) = \sum_{j=1}^{m} p_{kij} P_{jl} = P_l \sum_{j=1}^{m} p_{kij} = P_l P_i,$$

which demonstrates the independence of τ and τ^1. ■

3. The Bootstrap *P* Value and the Error in Rejection Probability

Consider next the bootstrap P value with joint realisation (i, j). It is the probability mass under the bootstrap DGP j of a value of τ less than π_i, that is, A_{ji}. Note that this makes no mention of the DGP that generated the realisation – in practice the true DGP is unknown. But we can nonetheless compute the distribution of the bootstrap P value under a given DGP k. Denote by $R_1(\cdot, k)$ the CDF of this distribution under k. Then

$$R_1(x, k) = P_k(A_{ji} \le x) = \sum_{i=0}^{n} \sum_{j=1}^{m} p_{kij} \mathrm{I}(A_{ji} \le x), \tag{7}$$

where $\mathrm{I}(\cdot)$ is an indicator function. Let $q_j^1(x)$ be defined by

$$q_j^1(x) = \max_{i=0,\ldots,n+1} \{i : A_{ji} \le x\}. \tag{8}$$

Then, as shown explicitly in Davidson (2017b), (7) becomes

$$R_1(x, k) = \sum_{j=1}^{m} a_{k q_j^1(x) j}. \tag{9}$$

When A_{ji} is the realised bootstrap P value, the bootstrap test rejects the null hypothesis at nominal level α if $A_{ji} < \alpha$. Under DGP k, the actual probability of rejection is $R_1(\alpha, k)$. For $\alpha = A_{ki}$ with arbitrary i, this becomes

$$R_1(A_{ki}, k) = \sum_{j=1}^{m} a_{kq_j^1(A_{ki})j} = \sum_{j=1}^{m} \sum_{l=0}^{q_j^1(A_{ki})-1} p_{klj}. \tag{10}$$

The bootstrap discrepancy at level A_{ki} under k, that is, the difference between the actual rejection probability and the nominal level, is $R_1(A_{ki}, k) - A_{ki}$.

We have seen that, when τ and β are independent, $p_{kij} = P_{ki}b_{kj}$. In general, without independence, define $\delta_{kij} > 0$ such that

$$p_{kij} = b_{kj}P_{ki}\delta_{kij}. \tag{11}$$

Then, since $\sum_{j=1}^{m} p_{kij} = P_{ki}$, we see by summing (11) over j that

$$P_{ki} = P_{ki}\sum_{j=1}^{m} b_{kj}\delta_{kij},$$

whence we conclude that, for any $k = 1, \ldots, m$ and $i = 0, \ldots, n$

$$\sum_{j=1}^{m} b_{kj}\delta_{kij} = 1. \tag{12}$$

In addition, (10) can be written as

$$R_1(A_{ki}, k) = \sum_{j=1}^{m} b_{kj} \sum_{l=0}^{q_j^1(A_{ki})-1} P_{kl}\delta_{klj}.$$

Thus the bootstrap discrepancy under k at level A_{ki} is

$$\sum_{j=1}^{m} b_{kj} \sum_{l=0}^{q_j^1(A_{ki})-1} P_{kl}\delta_{klj} - \sum_{j=1}^{m} b_{kj} \sum_{l=0}^{i-1} P_{kl}, \tag{13}$$

where we have used the relation $\sum_{j=1}^{m} b_{kj} = 1$ from (3).

If τ and β are independent under k, $\delta_{kij} = 1$ for all relevant i and j, and the discrepancy (13) is

$$\sum_{j=1}^{m} b_{kj} \Big[\sum_{l=0}^{q_j^1(A_{ki})-1} P_{kl} - \sum_{l=0}^{i-1} P_{kl} \Big]. \tag{14}$$

The interpretation of (14) is interesting. Notice first that it follows at once from the definition (8) that $q_k^1(A_{ki}) = i$. Thus the term with $j = k$ in (14) vanishes. If for some j, $q_j^1(A_{ki}) < i$, the corresponding term in (14) is negative, and, if $q_j^1(A_{ki}) > i$, it is positive. In the former case, (8) implies that, for all $l > q_j^1(A_{ki})$, $A_{jl} > A_{ki}$, and so, in particular, $A_{ji} > A_{ki}$. This means that DGP j assigns more probability mass to the event $\tau < \pi_i$ than does DGP k.

In many cases, there will be some positive and some negative terms in (14). Suppose then that, under DGP k, the realisation is (i, j) with $q_j^1(A_{ki}) < i$. The bootstrap P value is A_{ji}. The “ideal” P value, that is, the true probability mass for the event $\tau < \pi_i$, is A_{ki}, which is less than A_{ji}. The bootstrap P value is greater than the ideal one, and so the realisation (π_i, j) corresponds to under-rejection. Similarly, if $q_j^1(A_{ki}) > i$, this corresponds to over-rejection.

The presence of factors $\delta_{kq_j^1(A_{ki})j}$ different from 1 in (13) complicates the story, as we will see shortly. Notice here that the difference between the discrepancies (13) and (14) is

$$\sum_{j=1}^{m} b_{kj} \sum_{l=0}^{q_j^1(A_{ki})-1} P_{kl}(\delta_{klj} - 1). \tag{15}$$

The linear regression

In order to study the consequences of a correlation of τ and τ^1, as revealed by the diagnostic regression of realisations of τ^1 on joint realisations of τ and a constant, it is useful to express the covariance of τ and τ^1 in terms of the notation we have been developing. First, the expectations of these two random variables. We see immediately that

$$\mathrm{E}_k(\tau) = \sum_{i=0}^{n} P_{ki} \pi_i.$$

From (5), we find that

$$\mathrm{E}_k(\tau^1) = \sum_{j=1}^{m} \sum_{l=0}^{n} b_{kj} P_{jl} \pi_l = \sum_{j=1}^{m} b_{kj} \mathrm{E}_j(\tau).$$

The expectation of the product $\tau\tau^1$ can be calculated by use of the joint distribution given by (4):

$$\begin{aligned}\mathrm{E}_k(\tau\tau^1) &= \sum_{i=0}^{n}\sum_{l=0}^{n}\pi_i\pi_l P_k\big((\tau=\pi_i)\wedge(\tau^1=\pi_l)\big)\\ &= \sum_{i=0}^{n}\sum_{l=0}^{n}\pi_i\pi_l\sum_{j=1}^{m}p_{kij}P_{jl} = \sum_{i=0}^{n}\pi_i\sum_{j=1}^{m}p_{kij}\mathrm{E}_j(\tau).\end{aligned}$$

Then

$$\begin{aligned}\mathrm{cov}_k(\tau,\tau^1) &= \mathrm{E}_k(\tau\tau^1)-\mathrm{E}_k(\tau)\mathrm{E}_k(\tau^1)\\ &= \sum_{i=0}^{n}\pi_i\sum_{j=1}^{m}p_{kij}\mathrm{E}_j(\tau)-\mathrm{E}_k(\tau)\sum_{j=1}^{m}b_{kj}\mathrm{E}_j(\tau)\\ &= \sum_{j=1}^{m}\mathrm{E}_j(\tau)\Big[\sum_{i=0}^{n}\pi_i p_{kij}-b_{kj}\mathrm{E}_k(\tau)\Big].\end{aligned}$$

We check quickly that independence implies a zero correlation, as shown in Theorem 1. Since with independence $p_{kij}=P_{ki}b_{kj}$,

$$\mathrm{cov}_k(\tau,\tau^1)=\sum_{j=1}^{m}\mathrm{E}_j(\tau)\Big[b_{kj}\Big(\sum_{i=0}^{n}\pi_iP_{ki}-\mathrm{E}_k(\tau)\Big)\Big]=0.$$

In general,

$$\mathrm{cov}_k(\tau,\tau^1)=\sum_{j=1}^{m}\mathrm{E}_j(\tau)\,b_{kj}\sum_{i=0}^{n}\pi_iP_{ki}(\delta_{kij}-1). \tag{16}$$

Corollary 2

If the expectations $\mathrm{E}_k(\tau)$ are all equal to $\bar\tau$, independent of k, the covariance of τ and τ^1 is zero.

Proof:

The result follows immediately from Corollary 1 if τ is a pivot. Under the weaker condition used here, from (16) we see that

$$\mathrm{cov}_k(\tau,\tau^1)=\bar\tau\sum_{i=0}^{n}\pi_iP_{ki}\sum_{j=1}^{m}b_{kj}(\delta_{kij}-1)=0,$$

where the last equality follows from (12). ■

Theorem 2

Let the expectations $\mathrm{E}_j(\tau)$ $j=1,\ldots,m$, of the statistic under the DGPs of the model be bounded above by the positive number M. Then, if the covariance (16) is positive, the differences (15) between the bootstrap discrepancy and what it would be in the case of independence of τ and β average to negative values in the lower part of the distribution, and to positive values in the upper part.

Proof:

Observe first that, by (12),

$$\sum_{i=0}^{n}\sum_{j=1}^{m} b_{kj}P_{ki}(\delta_{kij}-1)=0. \tag{17}$$

Thus some of the terms $\sum_{j=1}^{m} b_{kj}P_{ki}(\delta_{kij}-1)$ in the sum over i are positive and some negative. For given A_{ki}, expression (15) is the sum of some of these terms with upper limits for the index l less than n.

The covariance (16) is bounded above by M times

$$\sum_{i=0}^{n}\sum_{j=1}^{m} b_{kj}\pi_i P_{ki}(\delta_{kij}-1), \tag{18}$$

and so, if (16) is positive, so too is the sum (18). By comparing (17) and (18), it can be seen that the weights π_i, increasing in i, push the sum from being zero, as in (17), to being positive, as in (18). This implies that the terms $\sum_{j=1}^{m} b_{kj}P_{ki}(\delta_{kij}-1)$ average to a negative value for small values of i and to a positive value for large values of i. Thus the values of expression (15), for increasing values of the nominal level A_{ki}, progressively incorporate fewer negative terms and more positive terms. This implies the statement of the Theorem. ■

Remarks:

Our assumption that τ takes the form of an approximate P value guarantees that $M \leq 1$. The seemingly redundant requirement in the statement of the Theorem is made in order to indicate that, if raw statistics are used, as would normally be the case in practice, the result continues to hold under this mild condition that $M < \infty$.

Nothing in the model requires the bootstrap discrepancies to move smoothly as a function of the nominal level. That is why the theorem speaks of averages of the discrepancies rather than their precise values. Especially if the correlation of τ and τ^1 is slight, particular values of the discrepancy may differ from the average effects.

The theorem tells us how the pattern of the bootstrap discrepancy differs from what it would be if, other things being equal, τ and β were independent. This means that non-independence may well reduce the magnitude of the discrepancy for particular levels. If, however, the discrepancy with independence is small, the introduction of dependence is bound to increase its magnitude for most levels.

In particular, the pattern of under-rejection by a test based on the ME bootstrap for conventional levels, and over-rejection for greater levels, follows from the strong positive correlation of τ and τ^1.

4. The Fast Double Bootstrap

It was remarked in the introduction that the FDB relies on the marginal distributions of τ and τ^1. In this section, we will see that the FDB can be made more accurate by use of their joint distribution. Of course, if the two are nearly independent, there is little or no gain, but when there is substantial correlation, the improvement can be considerable.

It may be useful here to present the algorithm that can be used to implement the FDB, where it is assumed without loss of generality that the rejection region is to the left of the statistic.

Algorithm for FDB

1. From the data set under analysis, compute the realisations of the statistic τ and the bootstrap DGP β.
2. Draw B bootstrap samples and use them to compute B independent realisations of the bootstrap statistic τ_j^*, $j = 1, \ldots, B$, and of the bootstrap DGP β_j^*.
3. Compute B second-level bootstrap statistics τ_j^{1*} by drawing from the simulated bootstrap DGPs β_j^*.
4. Compute the estimated first-level bootstrap P value $\hat{p}_1$ as the proportion of the τ_j^* smaller than τ. This is our estimate of $R_0(\tau, \beta)$.
5. Compute an estimate of the $\hat{p}_1$-quantile of the τ^{1*}; denote it by $\hat{q}^1$. This is our estimate of $Q^1\big(R_0(\tau, \beta), \beta\big)$.
6. Compute the estimated FDB P value $\hat{p}_2^f$ as the proportion of the τ_j^* smaller than $\hat{q}^1$. This is our estimate of the FDB P value $R_0\big(Q^1(\hat{p}_1, \beta), \beta\big)$.

Remark on the algorithm:

It is quite possible to use in the above algorithm a statistic that rejects to the right, like a χ^2 or F test, without first converting it to an approximate P value. The needed modifications are these: The statistic τ, as also the τ_j^* and the τ_j^{1*} are all constructed to reject to the right; in step 4, $\hat{p}_1$ is the proportion of the τ_j^* *greater* than τ; in step 5, it is the $(1 - \hat{p}_1)$-quantile that is needed; in step 6, $\hat{p}_2^f$ is the proportion of the τ_j^* greater than $\hat{q}^1$.

We can now express the FDB P value in the notation of our discrete model. Suppose that the DGP called μ in the introduction is the DGP k. The CDF of the single bootstrap P value is written as $R_1(x, k)$, and it can be approximated by $R_1^f(x, k) = R_0\big(Q^1(x, k), k\big)$, where $R_0(x, k)$ is the CDF under k of τ, and $Q^1(x, k)$ is the quantile function of τ^1. For $0 \leq x \leq 1$, let

$$i(x) = \max\big\{i : \pi_i < x\big\} + 1.$$

Note that $i(\pi_l) = l$. For given x, the function returns the index of the greatest value of τ that is no greater than x; in particular, for $\pi_i \le x < \pi_{i+1}$, $i(x) = i$. Then, from the definitions (2), we can see that $R_0(x,k) = A_{ki(x)}$. Similarly, from (5) we have

$$R^1(x,k) = \sum_{j=1}^{m} b_{kj} A_{ji(x)} \equiv A^*_{ki(x)}, \tag{19}$$

thus defining the probabilities A^*_{ki}. Analogously to (8), define

$$q^*_k(x) = \max_{i=0,\ldots,n} \{ i : A^*_{ki} \le x \}, \tag{20}$$

and observe that $q^*_k(A^*_{ki}) = i$. A possible definition of the quantile function Q^1, and the one we use here, is then $Q^1(x,k) = \pi_{q^*_k(x)}$. With these definitions, it follows that

$$R^f_1(x,k) = A_{kq^*_k(x)} = \sum_{j=1}^{m} a_{kq^*_k(x)j}. \tag{21}$$

This can be compared with the expression (9) for $R_1(x,k)$ itself. The nature of the approximation may be made still clearer by noting that, when τ and β are independent, so that $p_{kij} = P_{ki} b_{kj}$, we have

$$R_1(x,k) = \sum_{j=1}^{m} A_{kq^1_j(x)} b_{kj} \quad \text{and} \quad R^f_1(x,k) = \sum_{j=1}^{m} A_{kq^*_k(x)} b_{kj}.$$

The double bootstrap P value, p_2, is $R_1(p_1, \beta)$. With absolutely continuous distributions, the random variable $R_1(p_1, \mu)$ has the U(0,1) distribution, and so for p_2 the unknown true DGP μ is replaced by its estimate, namely the bootstrap DGP β. In exactly the same way, the FDB P value can be written as $R^f_1(p_1, \beta)$, and, since in our current notation $p_1 = A_{ji}$, this is

$$p^f_2 = A_{jq^*_j(A_{ji})}. \tag{22}$$

Theorem 3

The CDF of the fast double bootstrap P value under DGP k is given by

$$P_k(p^f_2 \le x) = \sum_{j=1}^{m} a_{ky(j,x)j}, \tag{23}$$

where $y(j,x)$ is defined to be $q^1_j(A^*_{jq^1_j(x)+1})$.

Proof:

We have

$$P_k(p_2^f \le x) = \sum_{i=0}^{n} \sum_{j=1}^{m} p_{kij} \mathrm{I}(A_{jq_j^*(A_{ji})} \le x).$$

The condition in the indicator function is equivalent to the condition $q_j^*(A_{ji}) \le q_j^1(x)$. Suppose that i is such that $A_{ji} \in [A^*_{jl}, A^*_{jl+1}[$ for some l. Then $q_j^*(A_{ji}) = l$. The condition is thus equivalent to $l \le q_j^1(x)$. This last inequality is satisfied if and only if $A_{ji} < A^*_{jq_j^1(x)+1}$ (with strict inequality), and this is equivalent to the requirement that $i < q_j^1(A^*_{jq_j^1(x)+1})$, that is, $i < y(j,x)$, again with strict inequality. The result (23) follows at once. ■

Remark:

At this point, we have imposed no structure at all on the model defined by the three-dimensional array of the quantities p_{kij}. It is therefore unsurprising that the result (23) by itself sheds no light on the good or bad performance of the FDB. The challenge is to discover conditions that make sense in the context of an econometric model and influence this performance.

Using the joint distribution

In Davidson and MacKinnon (1999) a finite-sample calculation leads to an expression for the bootstrap discrepancy at a given nominal level α in terms of the expectation of the α-quantile of the distribution of the statistic under the bootstrap DGP β, conditional on the statistic τ. If τ and β are independent, the conditional expectation is just the unconditional one, but otherwise the extent of dependence influences the bootstrap discrepancy. This suggests that the performance of the FDB, relying as it does on the assumption of near independence, can be enhanced in cases of dependence by explicitly taking account of it.

The simulations needed to estimate the FDB P value do not yield realisations of the quantile. Although those needed for the full double bootstrap do so, it appears that any technique for simulating these realisations must be as computationally expensive as the double bootstrap. However, the FDB does make explicit use, through its quantile function, of the distribution of τ^1, generated by the average of the bootstrap DGPs, and this may serve as a proxy for the distributions of the statistics generated by the different bootstrap DGPs severally.

It is therefore tempting to replace the unconditional quantile of τ^1 in the definition of the FDB P value by the quantile conditional on the realised value of τ. That realised value is generated by the true unknown DGP μ, but we can use the bootstrap principle to replace μ by the bootstrap DGP β, and obtain an estimated conditional quantile from the joint distribution of random variables τ and τ^1 generated by β.

We have

$$P_k(\tau^1 < \pi_i \mid \tau = \pi_l) = \sum_{i'=0}^{i-1} P_k(\tau^1 = \pi'_i \mid \tau = \pi_l).$$

Then it is natural to make the definition of the CDF of τ^1 conditional on $\tau = \pi_l$ as follows:

$$R^1(x, k \mid l) = \sum_{i'=0}^{i(x)-1} P_k(\tau^1 = \pi'_i \mid \tau = \pi_l) = \sum_{i'=0}^{i(x)-1} \sum_{j=1}^{m} p_{klj} P_{ji'}/P_{kl},$$

the last equality following from (4). Let $b_{kj|l} = p_{klj}/P_{kl}$, the probability under k that $\beta = j$ conditional on $\tau = \pi_l$. Then

$$R^1(x, k \mid l) = \sum_{j=1}^{m} \sum_{i=0}^{i(x)-1} b_{kj|l} P_{ji} = \sum_{j=1}^{m} b_{kj|l} A_{ji(x)}.$$

By analogy with (19) and (20), we make the definitions

$$A^*_{ki|l} = \sum_{j=1}^{m} b_{kj|l} A_{ji} \quad \text{and} \quad q^*_{k|l}(x) = \max_{i=0,\ldots,n} \big\{ i : A^*_{ki|l} \le x \big\}.$$

We are led to define the following conditional approximation:

$$R_1^{cf}(x, k \mid l) = R_0\big(Q^1(x, k \mid l), k\big), \tag{24}$$

where $Q^1(x, k \mid l) = \pi(q^*_{k|l}(x))$. Explicitly,

$$R_1^{cf}(x, k \mid l) = A_{kq^*_{k|l}(x)} = \sum_{j=1}^{m} a_{kq^*_{k|l}(x)j};$$

compare (21). We may now define the conditional FDB P value as

$$p_2^{cf} = R_1^{cf}(A_{ji}, j \mid i) = A_{jq^*_{j|i}(A_{ji})}, \tag{25}$$

analogously to (22).

Corollary to Theorem 3:

The CDF of the conditional FDB P value under DGP k is given by

$$P_k(p_2^{cf} \le x) = \sum_{j=1}^{m} \sum_{i < y_c(j,x|i)} p_{kij},$$

where $y_c(j, x \mid i)$ is $q^1_j(A^*_{jq^1_j(x)+1|i})$.

Proof:

Most of the algebra is identical to that used in the proof of Theorem 3. In the very last step of that proof, however, the sum over i cannot be performed explicitly, since the upper limit $i_{\max}$ is determined implicitly by

$$i_{\max} = \max_{i=0,\ldots,n} \big\{ i : i < y_c(j, x \mid i) \big\}. \qquad \blacksquare$$

The algorithm

Since it would be unusual to wish to use either the FDB or its conditional version with a discrete model, it is preferable here to give the algorithm for computing a conditional fast double bootstrap (CFDB) P value in a general way, allowing for both continuous and discrete models. In terms of the quantities computed from the data, the statistic τ, which is represented in the discrete model by the index i, and the bootstrap DGP β, represented by j, the P value p_2^{cf}, as given by (25), can be written, using (24), as

$$p_2^{cf} = R_1^{cf}\big(R_0(\tau,\beta),\beta \mid \tau\big) = R_0\big(Q^1(R_0(\tau,\beta),\beta \mid \tau),\beta\big). \tag{26}$$

In order to implement this formula, the unconditional CDF R_0 and the conditional CDF R^1 must be estimated for the bootstrap DGP, and the latter then inverted to obtain the conditional quantile function Q^1. The procedure is almost identical to that given in the Algorithm for FDB. The difference is that Steps 5 and 6 are to be replaced by the steps below:

5′. Each pair (τ_j^*, τ_j^{1*}) is a drawing from the joint distribution under β of (τ, τ^1). Compute an estimate of the $\hat{p}_1$-quantile of the τ^{1*} conditional on $\tau^* = \tau$; denote it by $\hat{q}^1$. This is our estimate of $Q^1\big(R_0(\tau,\beta),\beta \mid \tau\big)$.

6′. Compute the estimated CFDB P value $\hat{p}_2^{cf}$ as the proportion of the τ_j^* smaller than $\hat{q}^1$. This is our estimate of the CFDB P value (26).

Remark on the algorithm:

For step 5′, there are various ways to estimate the conditional quantile. The most obvious way is first to estimate the conditional CDF, using a kernel estimate, with a suitable kernel function K, perhaps Gaussian or Epanechnikov, and a bandwidth h. The estimate of the CDF of τ^1, evaluated at t^1, conditional on $\tau = t$, is

$$\hat{F}(t^1 \mid t) = \frac{\sum_{j=1}^B K\big((\tau_j^* - t)/h\big)\mathrm{I}(\tau_j^{1*} < t^1)}{\sum_{j=1}^B K\big((\tau_j^* - t)/h\big)}$$

This is just the Nadaraya-Watson estimate for the non-parametric regression of the indicator $\mathrm{I}(\tau^{1*} < t^1)$ on τ^*. It would no doubt be better to use a locally linear estimator instead. It has also been suggested that it might help to smooth the discontinuous indicator function, replacing it by a cumulative kernel, L say, evaluated at $(t^1 - \tau_j^{1*})/b$, where b is a bandwidth. Some experimentation showed that this led to a deterioration in the accuracy of the estimate, and so we made no use of this idea.

Whatever choice is made for the estimation of the conditional CDF, the conditional α-quantile can be estimated by solving the equation $\hat{F}(t^1 \mid t) = \alpha$ for t^1 given t. A root-finding algorithm, such as bisection or Brent, can be used for this purpose.

Alternatively, the check-function approach can be used. It is well known that the α-quantile of a distribution can be defined as the solution of the problem

$$\underset{q}{\operatorname{argmin}}\, \mathrm{E}\big(\rho_\alpha(Y - q)\big),$$

where Y is a random variable with the distribution of which the quantile is sought, and where the check function is defined as $\rho_\alpha(u) = u\big(\alpha - \mathrm{I}(u < 0)\big)$. For the conditional quantile we seek, we define another kernel estimator:

$$\hat{S}_\alpha(q \mid t) = \sum_{j=1}^{B} K\big((\tau_j^* - t)/h\big)\rho_\alpha(\tau_j^{1*} - q),$$

If this estimator is minimised with respect to q, the minimising q estimates the α-quantile of the τ^{1*} conditional on $\tau^* = t$. Again, a locally linear estimator may be preferred: it is given by minimising the function

$$\hat{S}_\alpha(q, \beta \mid t) = \sum_{j=1}^{B} K\big((\tau_j^* - t)/h\big)\rho_\alpha(\tau_j^{1*} - q - (\tau_j^* - t)\beta)$$

with respect to q and β, the minimising q being the estimate of the conditional quantile.

Recently, Racine and Li (2017) have reviewed a number of methods for estimating conditional quantiles, and have proposed a new approach, in which the quantile is estimated directly, rather than by inverting an estimated conditional CDF. More experience will be needed to determine which of the various methods is best adapted to the current problem.

5. Simulation Evidence

In seeking evidence for the performance of the CFDB by means of simulations, the first experiments were designed just to see if the CFDB was no worse than the ordinary FDB when there is approximate independence of τ and β. For this purpose, we used a setup considered in Davidson and MacKinnon (2007) as an illustration of the FDB.

Consider the linear regression model

$$\begin{aligned} y_t &= \boldsymbol{X}_t\boldsymbol{\beta} + u_t, \quad u_t = \sigma_t\varepsilon_t, \quad t = 1, \ldots, n, \\ \sigma_t^2 &= \sigma^2 + \gamma u_{t-1}^2 + \delta\sigma_{t-1}^2, \quad \varepsilon_t \sim \mathrm{IID}(0, 1). \end{aligned}$$

The disturbances of this model follow the GARCH(1,1) process introduced by Bollerslev (1986). The hypothesis that the u_t are IID in this model is tested by running the regression

$$\hat{u}_t^2 = b_0 + b_1\hat{u}_{t-1}^2 + \text{residual},$$

where $\hat{u}_t$ is the t^{th} residual from an OLS regression of y_t on $\boldsymbol{X}_t$. The null hypothesis that $\gamma = \delta = 0$ can be tested by testing the hypothesis that $b_1 = 0$. Besides the ordinary t statistic for b_1, a commonly used statistic is n times the centred R^2 of the regression, which has a limiting asymptotic distribution of χ_1^2 under the null hypothesis.

The experimental design is copied from Davidson and MacKinnon (2007). Since in general one is unwilling to make any restrictive assumptions about the distribution of the ε_t, a resampling bootstrap seems the best choice. In all cases, $\boldsymbol{X}_t$ consists of a constant and two independent, standard normal random variates. In order to have a non-negligible bootstrap discrepancy, the ε_t are drawn from the χ^2_2 distribution, subsequently centred and rescaled to have variance 1. For the same reason, the sample size of 40 is rather small. Without loss of generality, we set $\boldsymbol{\beta} = \mathbf{0}$ and $\sigma^2 = 1$, since the test statistic is invariant to changes in the values of these parameters. The invariance means that we can use a straightforward resampling bootstrap DGP, with the y^*_t in a bootstrap sample IID drawings from the empirical distribution of the y_t. For the iterated bootstrap, y^{**}_t is resampled from the y^*_t.

The experiment consisted of 10,000 replications, each with 9,999 bootstrap repetitions. The distributions of the P values of the single bootstrap, the FDB, and the CFDB were estimated; the results are displayed in Figure 1 as P value discrepancy plots. Despite the number of replications, there remains non-negligible simulation randomness. However, it can be seen that the discrepancies of the FDB and CFDB are of the same order of magnitude, with the FDB actually performing better than the CFDB for nominal levels of practical interest. The overall discrepancy of the single bootstrap is, as expected, somewhat greater.

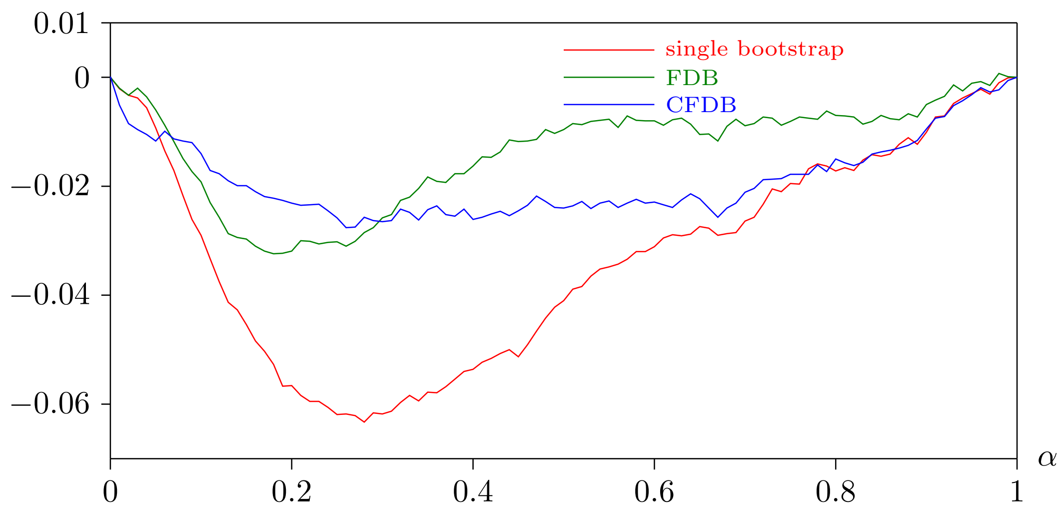

Figure 1: *P* value discrepancy plots for the bootstrap, FDB, and CFDB

A much more interesting experiment was then undertaken, with the ME bootstrap, for which τ and β are greatly correlated, so that the discrepancies are huge. The model considered was

$$y_t = \boldsymbol{X}_1\boldsymbol{\beta}_1 + x_2\beta_2 + u_t. \tag{27}$$

The disturbances follow an AR(1) process with autoregressive parameter $\rho = 0.9$ and Gaussian innovations. The regressors in the matrix $\boldsymbol{X}_1$ include a constant and two other regressors, serially correlated with autoregressive parameter 0.8 as also the last regressor x_2. Without loss of generality, all the slope coefficients, $\boldsymbol{\beta}_1$ and β_2 are set to zero. The null hypothesis that $\beta_2 = 0$ is tested using a χ^2 with one degree of freedom, and with a Newey-West HAC covariance matrix estimator; see Newey and

West (1987). The sample size is 64, and the lag truncation parameter for the HAC estimator is 16.

As with the GARCH experiment, the conditional quantile was estimated by first estimating the conditional CDF using a locally linear estimator, and then inverting it using the bisection method. A preliminary experiment was undertaken with a bivariate normal distribution, for which the true value of the conditional quantile was known analytically. It was found that smoothing the indicator was counter-productive, and led to significantly greater bias than with the indicator itself, and so in the experiment with the model (27) the indicator was not smoothed.

The details of the ME algorithm used can be found in Davidson and Monticini (2014) and in Bergamelli *et al* (2015). In Figure 2 P value discrepancy plots are shown for the test of $\beta_2 = 0$ by the single ME bootstrap and its FDB and CFDB versions. We used 10,000 replications each with 999 bootstrap repetitions. It can be seen that the FDB is at best marginally better than the single ME bootstrap, but that the CFDB gives rise to a very considerable improvement, without being really reliable.

A very much costlier experiment was undertaken for which, in addition to the bootstrap versions examined in Figure 2, the discrepancy for the full double bootstrap was estimated. Again, 10,000 replications were performed, with 999 bootstrap repetitions at the first level, and 399 at the second level. In order for the experiment to be completed in a reasonable timeframe, 50 concurrent processes were used, with different random number generators in each. The mean running time of the processes was around 14 hours, but with a fairly wide variance, so that results were available after around 16 hours. As an indication of the relative computing costs, each process for the experiment for Figure 2 took on average only a little more than 5 minutes.

The results of this experiment, shown in Figure 3, show the same results as in Figure 2, with the addition of those for the full double bootstrap. It emerges clearly that the full double bootstrap performs only slightly better than the CFDB.

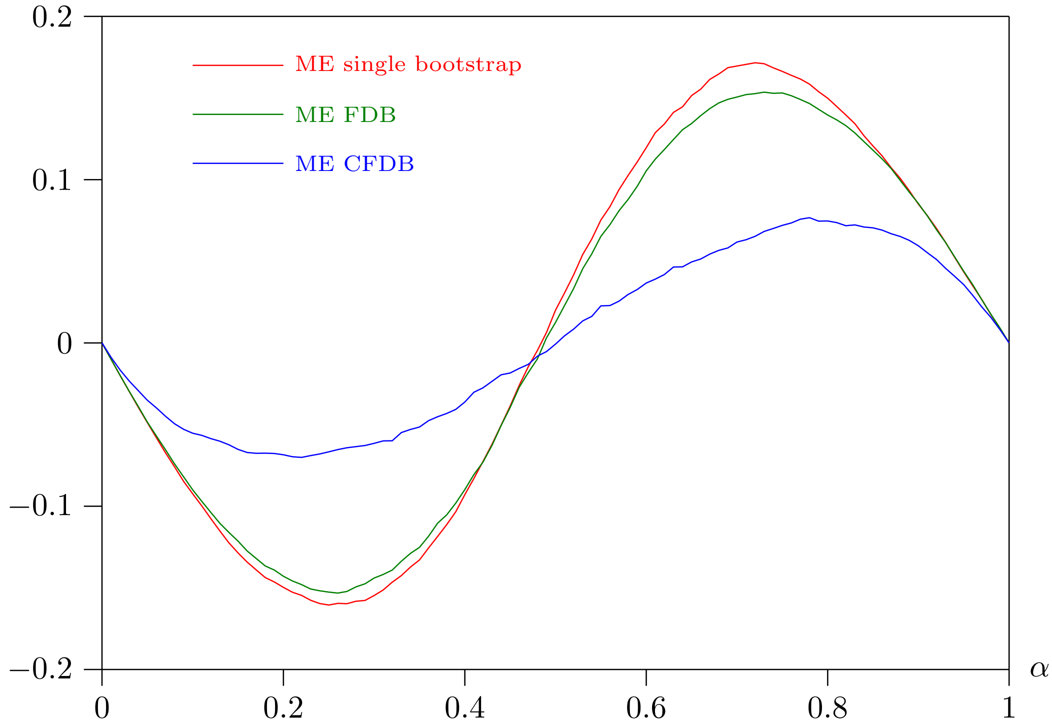

Figure 2: Results for the single ME bootstrap and its FDB and CFDB versions

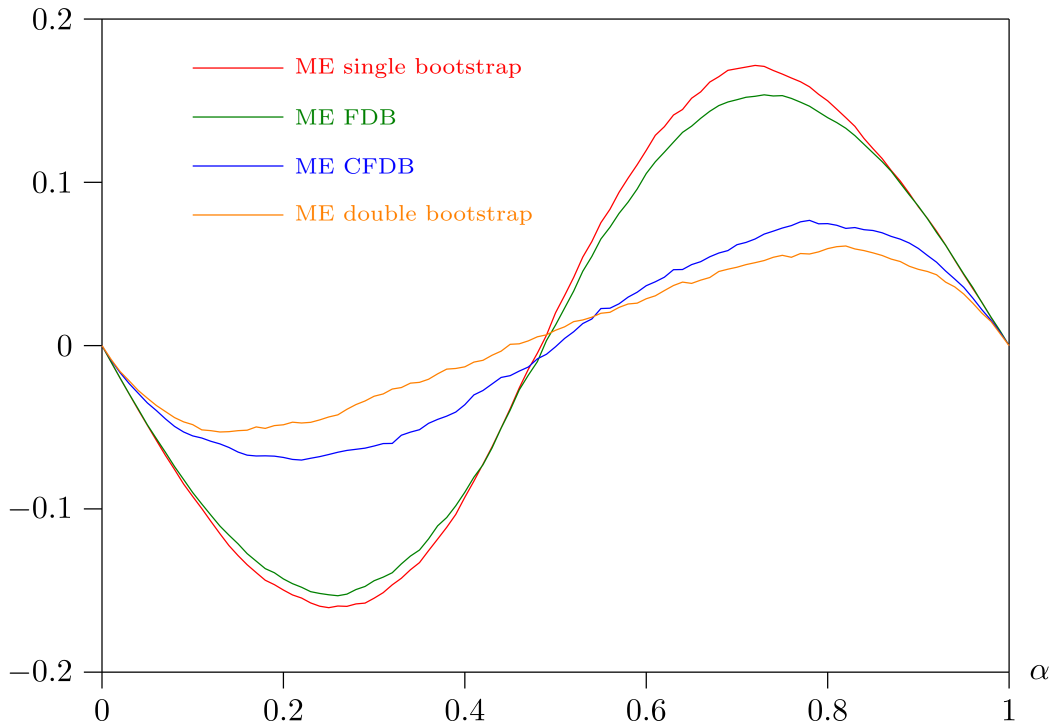

Figure 3: As in Figure 2, with the addition of the full double bootstrap

6. Conclusion

Although the focus of everything here has been on bootstrap hypothesis tests, it is entirely possible to use the techniques of both the FDB and the CFDB for the construction of confidence intervals. It was pointed out by Chang and Hall (2015) that, although the FDB benefits from asymptotic refinements for testing, it does not do so for confidence intervals. This seeming defect can be overcome at a certain computational cost, by a procedure similar to Hansen's (1999) grid bootstrap, or the one used in Davidson and MacKinnon (2010).

Davidson and Trokić (2020) propose fast versions of higher-order iterated bootstraps, beginning with the fast triple bootstrap, which they show to be able in some cases to reduce discrepancies relative to the FDB. It would be good to develop conditional fast iterated bootstraps. The use of the discrete formulation will almost certainly be a great help in this task, as it should be for many studies of the finite-sample properties of the bootstrap.

In Davidson (2017b), a discrete model is given for which the actual numerical values of the probabilities p_{kij} are available. It will be extremely interesting to use these numbers as a test bed for the new conditional procedure, and for the extensions that will be explored in future work.

Appendix

Theorem 1 in the continuous case

In keeping with the spirit of the paper, the regularity conditions used for this theorem will be kept to the bare minimum.

A.1 For convenience, we suppose that the test statistic τ is in approximate P value form, and can take on any real value in the interval [0,1], on which is defined the Borel σ-algebra $\mathcal{B}[0,1]$, and the usual Lebesgue measure $\mathcal{L}[0,1]$.

A.2 The model $\mathbb{M}$ is a set consisting of all the DGPs that respect the null hypothesis. A σ-algebra $\mathcal{B}(\mathbb{M})$ is defined on $\mathbb{M}$, along with a probability measure $m : \mathcal{B}(\mathbb{M}) \to [0,1]$.

A.3 The product σ-algebra $\mathcal{B}[0,1] \times \mathcal{B}(\mathbb{M})$ and the product measure $\mathcal{L}[0,1] \times m$ are well defined.

A.4 For each DGP $\mu \in \mathbb{M}$, a probability measure P_μ is defined on the product σ-algebra and is absolutely continuous with respect to $\mathcal{L}[0,1] \times m$.

A.5 For each $\mu \in \mathbb{M}$, the measure P_μ has a density $f(\mu, t, b)$, such that, for any subset $C \in \mathcal{B}[0,1] \times \mathcal{B}(\mathbb{M})$,

$$P_\mu(C) = \int_0^1 \int_{\mathbb{M}} \mathrm{I}\big((t,b) \in C\big) f(\mu,t,b)\, \mathrm{d}t\, m(\mathrm{d}b).$$

Here, I is the indicator function.

For any given μ, define the marginal densities:

$$g_\mu(t) = \int_{\mathbb{M}} f(\mu, t, b)\, m(\mathrm{d}b), \quad \text{and}$$
$$h_\mu(b) = \int_0^1 f(\mu, t, b)\, \mathrm{d}t,$$

and note that independence of τ and β implies that $f(\mu, t, b) = g_\mu(t)h_\mu(b)$. Then

$$P_\mu\big((\tau \le x) \wedge (\tau^1 \le x_1)\big) = \int_0^x \int_{\mathbb{M}} f(\mu, t, b)\, \mathrm{d}t\, m(\mathrm{d}b) \int_0^{x_1} g_b(t^1)\, \mathrm{d}t^1. \tag{28}$$

From this, by setting first $x_1 = 1$, and then $x = 1$, we derive easily that

$$P_\mu(\tau \le x) = \int_0^t g_\mu(t)\, \mathrm{d}t \text{ and } P_\mu(\tau^1 \le x_1) = \int_{\mathbb{M}} h_\mu(b) \int_0^{x_1} g_b(t^1)\, \mathrm{d}t^1\, m(\mathrm{d}b). \tag{29}$$

Now, if the joint density factorises in the case of independence, the probability in (28) becomes

$$\int_0^x g_\mu(t)\, \mathrm{d}t \int_{\mathbb{M}} h_\mu(b) \int_0^{x_1} g_b(t^1)\, \mathrm{d}t^1\, m(\mathrm{d}b),$$

which is just the product of the two marginal probabilities in (29). ■

References

Ahlgren, N. and J. Antell (2008). "Bootstrap and Fast Double Bootstrap Tests of Cointegration Rank with Financial Time Series", *Computational Statistics & Data Analysis*, **52**, 4754–4767.

Beran, R., (1988). "Prepivoting test statistics: a bootstrap view of asymptotic refinements", *Journal of the American Statistical Association* **83**, 687-697.

Bergamelli, M., J. Novotný, and G. Urga (2015). "Maximum Non-Extensive Entropy Block Bootstrap for Non-stationary Processes", *L'Actualité Economique*, **91(1-2)**, 115–139.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* **31**, 307–327.

Chang, J. and P. Hall (2015). "Double-bootstrap methods that use a single double-bootstrap simulation", *Biometrika* `doi:10.1093/biomet/asu060`

Davidson, J. (2006). "Alternative bootstrap procedures for testing cointegration in fractionally integrated processes," *Journal of Econometrics*, **133**, 741-777.

Davidson, R. (2017a). “Diagnostics for the Bootstrap and Fast Double Bootstrap”, *Econometric Reviews*, **36**, 1021–1038, `doi:10.1080/07474938.2017.1307918`

Davidson, R. (2017b). “A Discrete Model for Bootstrap Iteration”, *Journal of Econometrics* **201**, 228–236, `doi:10.1016/j.jeconom.2017.08.005`

Davidson, R. and J. G. MacKinnon (1999). “The Size Distortion of Bootstrap Tests”, *Econometric Theory*, **15**, 361-376.

Davidson, R. and J. G. MacKinnon (2002). “Fast double bootstrap tests of nonnested linear regression models,” *Econometric Reviews*, **21**, 417–427.

Davidson, R. and J. G. MacKinnon (2007). “Improving the Reliability of Bootstrap Tests with the Fast Double Bootstrap,” *Computational Statistics & Data Analysis*, **51**, 3259–3281.

Davidson, R. and J. G. MacKinnon (2010). “Wild bootstrap tests for IV regression”, *Journal of Business and Economic Statistics*, **28**, 128–144.

Davidson, R. and A. Monticini (2014). “Heteroskedasticity-and-autocorrelation-consistent bootstrapping. Technical report, Università Cattolica del Sacro Cuore, Dipartimenti e Istituti di Scienze Economiche (DISCE).

Davidson, R. and M. Trokić (2020). “The Fast Iterated Bootstrap”, *Journal of Econometrics*, `https://doi.org/10.1016/j.jeconom.2020.04.025`

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

Hansen, B. E. (1999). “The grid bootstrap and the autoregressive model,” *Review of Economics and Statistics*, 81, 594–607.

Lamarche, J.-F. (2004). “The numerical performance of fast bootstrap procedures,” *Computational Economics*, **23**, 379–389.

Newey, W. K. and K. D. West (1987). “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix”, *Econometrica* **55**, 703–708.

Omtzigt, P. and S. Fachin (2006). “The size and power of bootstrap and Bartlett-corrected tests of hypotheses on the cointegrating vectors”, *Econometric Reviews* **25**, 41–60.

Ouysse, R. (2013). “A Fast Iterated Bootstrap Procedure for Approximating the Small-Sample Bias”, *Communications in Statistics – Simulation and Computation*, **42**, doi:10.1080/03610918.2012.667473

Racine, J. S. and K. Li (2017). “Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach”, *Journal of Econometrics* **201**, 72–94. `http://dx.doi.org/10.1016/j.jeconom.2017.06.020`

Vinod, H. D. (2006). “Maximum entropy ensembles for time series inference in economics”, *Journal of Asian Economics* 17, 955–978.

Vinod, H. D. and J. López-de Lacalle (2009). “Maximum Entropy Bootstrap for Time Series: the meboot R Package”, *Journal of Statistical Software* **29**, 1–19.