The Fast Iterated Bootstrap

by

Russell Davidson

Department of Economics and CIREQ
McGill University
Montréal, Québec, Canada
H3A 2T7

GREQAM
Centre de la Vieille Charité
2 Rue de la Charité
13236 Marseille cedex 02, France

russell.davidson@mcgill.ca

and

Mirza Trokić

Department of Economics
Bilkent University
Bilkent, Ankara, Turkey
06800

trokic@bilkent.edu.tr

Abstract

The standard forms of bootstrap iteration are very computationally demanding. As a result, there have been several attempts to alleviate the computational burden by use of approximations. In this paper, we extend the fast double bootstrap of Davidson and MacKinnon (2007) to higher orders of iteration, and provide algorithms for their implementation. The new methods make computational demands that increase only linearly with the level of iteration, unlike standard procedures, whose demands increase exponentially. In a series of simulation experiments, we show that the fast triple bootstrap improves on both the standard and fast double bootstraps, in the sense that it suffers from less size distortion under the null with no accompanying loss of power.

Key words: bootstrap iteration, fast iterated bootstrap

JEL codes: C10, C12, C15, C63

This research was supported by the Canada Research Chair program (Chair in Economics, McGill University), and by grants from the Social Sciences and Humanities Research Council of Canada and the Fonds Quebecois de Recherche sur la Societe et la Culture. We are grateful to seminar participants at the Third French Econometric Conference, the Singapore Management University, the Hong Kong University of Science and Technology, and the University of Sydney, for useful comments.

December 2016

1. Introduction

Bootstrap iteration has been discussed in a wide variety of contexts from not long after the invention of the bootstrap by Efron (1979). The first article to mention the topic of the iterated bootstrap to our knowledge is Hall (1986), a paper followed quickly by two articles by Beran (1987) and (1988), in which the double bootstrap is introduced. An extensive theoretical discussion, with examples, is given in Hall and Martin (1988), in which the approaches of Hall and Beran are unified. Bootstrap iteration is also mentioned in Hall's (1992) influential book, in which Edgeworth expansions play a central role.

However, the computational burden of bootstrap iteration was then, and still is, although to a lesser extent, very heavy. Consequently, a number of attempts have been made to lighten this burden. An early one is found in DiCiccio, Martin, and Young (1992). It makes no use of Edgeworth expansions, preferring instead other asymptotic arguments. Lee and Young (1995) develop procedures for bootstrap confidence intervals with an iterated bootstrap that does not involve any resampling, replacing it with analytic asymptotic approximations. Chan and Lee (2001) derive an algorithm for infinitely iterated bootstrap bias correction, by considering bootstrap iteration as a Markov process. In Lee and Young (2003), possibilities are investigated of weighted resampling to mimic the results of bootstrap iteration. Ho and Lee (2005) consider bootstrap iteration in conjunction with smoothing of the discrete distribution associated with conventional resampling.

Another attempt to alleviate the computation burden is the fast double bootstrap (FDB) exposited in Davidson and MacKinnon (2007). The technique had been used previously: Davidson and MacKinnon (2002) use it to improve the reliability of bootstrap tests of non-nested linear regression models; Omtzigt and Fachin (2006) show that it gives better size correction than the single bootstrap with or without Bartlett correction in the cointegrated VAR model; Lamarche (2004) investigates some of its numerical properties; Davidson (2006) uses it in developing bootstrap tests of cointegration with fractionally integrated time series. A recent working paper, Ouysse (2009) develops a version of the FDB for the purposes of bias correction. Giacomini, Politis, and White (2013) provide a formal analysis of the FDB, under the name of a "warp-speed" method, and give conditions under which the method is justified asymptotically. Most recently, Chang and Hall (2015) revert to the use of Edgeworth expansions to explore rates of convergence of the FDB, and show that it improves on the single bootstrap for the purpose of bias correction, but not for constructing bootstrap confidence intervals.

The starting point for this paper is found in Davidson (2010), in which the FDB is studied and compared with the standard double bootstrap and some other procedures aimed at improving bootstrap reliability. In the context of a much simplified unit root test, it is shown that estimating the distribution of the FDB P value leads to a reduction in the size distortion of the FDB, which is less distorted than the single bootstrap, and roughly comparable to the standard double bootstrap. The procedure is more or less a bootstrap of the FDB, which gives rise to an approximation to the bootstrapped double bootstrap, that is, the twice iterated, or triple, bootstrap. The FDB algorithm is not a nested bootstrap procedure. But when it is bootstrapped, what results is a nested bootstrap procedure, essentially as computationally intensive as the double bootstrap.

It is therefore interesting to see whether it is possible to make use of the approximations that simplify the original double bootstrap algorithm to the FDB, but to the bootstrapped FDB. One would then effectively obtain an approximation to the standard triple bootstrap, an approximation that can reasonably be termed the fast triple bootstrap (FTB). In this paper, after having discussed bootstrap iteration and developing some suitable notation, we show how fast versions of any order of iterated bootstrap can be defined. Further, whereas almost all of the papers cited

above restrict attention to the conventional resampling bootstrap, or varieties of the block bootstrap, which also involve resampling, our development here is applicable more widely, and covers the parametric bootstrap, the wild bootstrap, and other bootstrap procedures quite generally.

In the next section, we set up notation for dealing with bootstrap iteration, and detail some assumptions made in the subsequent material. In section 3, bootstrap iteration to any order is defined formally, and a rather crude sufficient condition given for the bootstrap to perform prepivoting. Then, in section 4, we review the theory of the FDB, and give algorithms for the computation of FDB P values and confidence intervals. Section 5 is where we pursue an approach analogous to that which leads to the FDB in order to motivate and define the fast triple bootstrap, for which a computational algorithm is detailed. Once we can see how to get as far as the FTB, the road is clear to defining arbitrary orders of fast iterated bootstraps; that is done in section 6. In section 7, we report the results of three sets of simulation experiments, one based on the experimental design of Davidson (2010) for a unit root test, the next on a test for ARCH effects, and the third on a test for serial correlation of the disturbances of a regression model. Section 8 concludes.

2. Concepts and Notations

We denote by $\mathbb{M}_0$ the set of data-generating processes (DGPs) that satisfy a null hypothesis we wish to test. The test statistic used is denoted by τ. In general, τ has a different distribution under the different DGPs in $\mathbb{M}_0$, and it certainly has a different distribution under DGPs in the model, $\mathbb{M}$ say, that represents the alternative hypothesis. Here $\mathbb{M}_0 \subset \mathbb{M}$. It is conventional to suppose that τ is defined as a random variable on some suitable probability space, on which we define a different probability measure for each different DGP.

Rather than using this approach, we define a probability space $(\Omega, \mathcal{F}, P)$, with just one probability measure, P. Then we treat the test statistic τ as a stochastic process with the model $\mathbb{M}$ as index set. We have

$$\tau \ : \ \mathbb{M} \times \Omega \to \mathbb{R}. \tag{1}$$

Since most of the discussion of the paper is couched in the language of simulation, the probability space can, for our present purposes, be taken to be that of a random number generator. A realisation of the test statistic is therefore written as $\tau(\mu, \omega)$, for some $\mu \in \mathbb{M}$ and $\omega \in \Omega$.

Since this is an unconventional way of proceeding, it is probably necessary to explain why it is useful for our present purposes. First, it accurately represents how one sets about doing a simulation experiment. The notation ω stands for the random (or pseudo-random) numbers obtained from the computer's random number generator, and μ stands for the non-random computer program that transforms the raw random numbers into realisations of the statistic or estimator under study. It makes it easy to describe experiments in which one uses the *same* random numbers with different DGPs, as well as the operation used whenever the bootstrap is implemented by simulation, namely generating a large number of IID realisations from one and the same DGP. In addition, it will be seen later that our notation allows an accurate description of bootstrap iteration.

It may be thought that the approach nonetheless does violence to our conception of real-world DGPs. This point is taken up in Davidson (2015), where the thesis is that, by using models to describe the real world, we invent a sort of virtual reality which can, if we so wish, be implemented on a computer.

Assumption 1: measurability and continuity

For all $\mu \in \mathbb{M}$, the mapping

$$\tau_\mu : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}), \quad \tau_\mu(\omega) = \tau(\mu, \omega),$$

where $\mathcal{B}$ is the Borel sigma-algebra defined on the real line $\mathbb{R}$, is measurable. The probability measure induced on $\mathbb{R}$ by τ_μ from the measure P on Ω is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, and has a continuous density.

Assumption 2: approximate *P* value

For all $\mu \in \mathbb{M}$, the range of τ_μ is the $[0, 1]$ interval, and its density is nowhere zero in the interior of the interval. The statistic takes the form of an approximate P value, which leads to rejection when the statistic is too small.

Remarks:

Measurability in Assumption 1 imposes no meaningful restriction. Absolute continuity, on the other hand, is restrictive, but it makes for a considerable simplification of the theoretical analysis, and serves as a useful approximation when the actual distribution is discrete. The specification of the range of τ_μ in Assumption 2 is made purely for notational convenience, and and, if Assumption 1 is satisfied, is completely unrestrictive. The requirement of a non-zero density is made so that the corresponding distribution and quantile functions are strictly increasing on $[0, 1]$.

Let $R_0 : [0, 1] \times \mathbb{M} \to [0, 1]$ be the cumulative distribution function (CDF) of τ under the DGP μ:

$$R_0(x, \mu) = P\{\omega \in \Omega \,|\, \tau(\mu, \omega) \leq x\}. \tag{2}$$

Let $Q_0 : [0, 1] \times \mathbb{M} \to [0, 1]$ be the corresponding quantile function. Since by Assumption 2 R_0 and Q_0 are continuous and strictly increasing, we have the relations

$$R_0(Q_0(x, \mu), \mu) = x = Q_0(R_0(x, \mu), \mu). \tag{3}$$

Suppose that we have a statistic computed from a data set that was generated by a DGP $\mu \in \mathbb{M}$. Denote this statistic as $t \equiv \tau(\mu, \omega)$. If μ belongs to the null-hypothesis model $\mathbb{M}_0$, the ideal P value, $R_0(t, \mu)$, is a drawing from $\mathrm{U}(0, 1)$, the uniform distribution on $[0, 1]$. Exact inference on the basis of the ideal P value is of course in general infeasible, because μ is unknown.

Remark:

If τ is a pivotal statistic with respect to $\mathbb{M}_0$, then the distribution of $\tau(\mu, \cdot)$ is the same for all $\mu \in \mathbb{M}_0$. In such a case, exact inference can be based on an Monte Carlo test; see for instance Dufour and Khalaf (2001).

The principle of the bootstrap is that, when we want to use some function or functional of an unknown DGP μ, we use an estimate in its place. This estimate is the bootstrap DGP. Analogously to (1), define the DGP-valued stochastic process

$$\beta \;:\; \mathbb{M} \times \Omega \to \mathbb{M}_0,$$

where, although the model $\mathbb{M}$ on the left-hand side may include the entire alternative hypothesis, we insist that the $\mathbb{M}_0$ on the right-hand side is the null model under test (the first golden rule of bootstrapping). The bootstrap DGP computed from the data set for which the statistic is $\tau(\mu, \omega)$

is then $b \equiv \beta(\mu, \omega)$, with the *same* realisation ω. The notation neatly encapsulates the fact that the distribution of statistics generated by the bootstrap DGP is *conditional* on the realised sample. The bootstrap statistic, or bootstrap P value, is then $R_0(t, b)$. Since we have assumed that the distribution of τ is absolutely continuous for all $\mu \in \mathbb{M}_0$, and since the bootstrap DGP $b \in \mathbb{M}_0$, the bootstrap statistic has an absolutely continuous distribution.

Unless the function R_0 is known analytically, simulation is necessary for the computation of the bootstrap statistic. We make the definition

$$\hat{R}_0(x, \mu) = \frac{1}{B} \sum_{j=1}^{B} \mathrm{I}(\tau(\mu, \omega_j^*) < x), \tag{4}$$

where the ω_j^* are IID drawings from the random number generator, and $\mathrm{I}(\cdot)$ is the indicator function, equal to one when the Boolean argument is true, and to zero when it is false. As $B \to \infty$, $\hat{R}_0(\alpha, \mu)$ tends almost surely to $R_0(\alpha, \mu)$ by the strong law of large numbers. Accordingly, we estimate the bootstrap statistic by $\hat{R}_0(t, b)$. However, in the theoretical discussion that follows, we imagine that B is infinite, so that the function R_0 is available.

3. Bootstrap Iteration

It is useful at this point to define a stochastic process that represents the bootstrap statistic. Since $t = \tau(\mu, \omega)$ and $b = \beta(\mu, \omega)$, the definition is

$$p_1 \ : \ \mathbb{M} \times \Omega \to \mathbb{R}, \quad p_1(\mu, \omega) = R_0\big(\tau(\mu, \omega), \beta(\mu, \omega)\big). \tag{5}$$

Since by absolute continuity R_0 is a continuous function, it follows that p_1 also has an absolutely continuous distribution. We denote the continuous CDF of $p_1(\mu, \omega)$ by $R_1(\cdot, \mu)$.

The random variable $R_1\big(p_1(\mu, \omega), \mu\big)$ is, by construction, distributed as U(0,1). But, as with $R_0\big(\tau(\mu, \omega), \mu\big)$, which is also distributed as U(0,1), this fact is not enough to allow exact inference, because the actual μ that generates the data is unknown, except of course in the context of a simulation experiment. However, the bootstrap principle can again be applied, and the unknown μ replaced by the estimate b, which can be computed from observed data. This leads to the double bootstrap, of which the P value, for realisations t and b, can be written as

$$R_1\big(R_0(t, b), b\big). \tag{6}$$

This is just the bootstrap estimate of the probability mass in the distribution of the single bootstrap statistic to the left of the realisation $R_0(t, b)$. Expressed as a random variable, this double bootstrap P value is

$$p_2(\mu, \omega) \equiv R_1\big(R_0(\tau(\mu, \omega), \beta(\mu, \omega)), \beta(\mu, \omega)\big) \tag{7}$$

If we write the right-hand side above as $R_1\big(p_1(\mu, \omega), b(\mu, \omega)\big)$, the analogy with the definition (5) of $p_1(\mu, \omega)$ is complete. This demonstrates that the double bootstrap effectively bootstraps the single bootstrap P value.

From that observation, it is clear that we can define iterated bootstraps as follows.

Definition 1: iterated bootstrap

For $k = 0, 1, 2, \ldots$, the k^{th} order bootstrap P value and its distribution are given by recurrence relation

$$R_k(\alpha, \mu) = P\{\omega \in \Omega \,|\, p_k(\mu, \omega) \leq \alpha\}, \tag{8}$$
$$p_{k+1}(\mu, \omega) = R_k(p_k(\mu, \omega), \beta(\mu, \omega)), \tag{9}$$

where we initialise the recurrence by the definition $p_0(\mu, \omega) = \tau(\mu, \omega)$.

Thus $p_{k+1}(\mu, \omega)$ is the bootstrap P value obtained by bootstrapping the k^{th} order P value $p_k(\mu, \omega)$. It estimates the probability mass in the distribution of the k^{th} order P value to the left of its realisation.

In all discussions about bootstrap iteration, it is implicitly or explicitly assumed that iteration leads to more reliable inference. In a wide variety of circumstances, it has been possible to demonstrate that asymptotic refinements accrue with bootstrap iteration. As the bootstrap is not an asymptotic procedure, it is desirable to analyse the situation in finite samples.

We wish to consider a measure defined on the product space $[0, 1] \times \mathbb{M}_0$, the space in which the two random variables $\tau(\mu, \omega)$ and $\beta(\mu, \omega)$ are jointly realised. The component $[0, 1]$ of the product space can be equipped with the Borel σ-algebra $\mathcal{B}$, as in Assumption 1 but restricted to the interval $[0, 1]$ on account of Assumption 2, but it is less obvious how to associate a σ-algebra to the model $\mathbb{M}$. The easiest way to do so is to introduce a metric on the space of CDFs of random variables defined on $[0, 1]$. It results from the combination of the uniform or L^∞ metrics on the CDFs and their corresponding quantile functions.

Definition 2: metric on $\mathbb{M}$

For any two DGPs μ and ν in $\mathbb{M}$, let the distance between them be defined as

$$\lambda_0(\mu, \nu) = \frac{1}{2}\Big\{ \sup_{x \in [0.1]} |R_0(x, \mu) - R_0(x, \nu)| + \sup_{x \in [0,1]} |Q_0(x, \mu) - Q_0(x, \nu)| \Big\}. \tag{10}$$

The choice of the function R_0 in the definition of the metric is motivated by the fact that the different behaviour of the function when evaluated at a DGP μ and when evaluated at the bootstrap DGP $\beta(\mu, \omega)$ is what accounts for the bootstrap discrepancy, that is, the difference between the RP of a bootstrap test and the nominal significance level.

The metric λ_0 can be used to define the σ-algebra of Borel sets on $\mathbb{M}$, which we write as $\mathcal{B}(\mathbb{M})$, generated by the open sets in the topology defined by the metric. Then we can define the product σ-algebra $\mathcal{B}[0, 1] \times \mathcal{B}(\mathbb{M})$ on the product space $[0, 1] \times \mathbb{M}_0$, thus giving the measurable space $\mathcal{M} \equiv \big([0, 1] \times \mathbb{M}_0, \mathcal{B}[0, 1] \times \mathcal{B}(\mathbb{M}_0)\big)$.

A DGP $\mu \in \mathbb{M}_0$ defines an induced measure P_μ on $\mathcal{M}$ as follows: for any Borel set $A \in \mathcal{B}[0, 1] \times \mathcal{B}(\mathbb{M}_0)$,

$$P_\mu(A) = P\{\omega \in \Omega \,|\, (\tau(\mu, \omega), \beta(\mu, \omega)) \in A\}.$$

The measure P_μ has a projection measure B_μ defined on $\mathcal{B}(\mathbb{M}_0)$: for $B \in \mathcal{B}(\mathbb{M}_0)$,

$$B_\mu(B) = P\{\omega \in \Omega \,|\, \beta(\mu, \omega) \in B\}.$$

Assumption 3: existence of carrier measure

There exists a measure $m : \mathcal{B}(\mathbb{M}_0) \to [0,1]$ such that, for all $\mu \in \mathbb{M}_0$, the measure B_μ is absolutely continuous with respect to m.

The product measure M is defined naturally on $\mathcal{M}$ as $\mathcal{L}[0,1] \times m$, where $\mathcal{L}[0,1]$ is Lebesgue measure on $[0,1]$. For $A \in \mathcal{B}[0,1]$ and $B \in \mathcal{B}(\mathbb{M}_0)$, we have $M(A \times B) = \mathcal{L}(A)m(B)$.

Assumption 4: existence of density

The probability measure P_μ has a density $f_\mu : [0,1] \times \mathbb{M}_0 \to \overline{R}_+$, such that, if $C \in \mathcal{B}[0,1] \times \mathcal{B}(\mathbb{M}_0)$,

$$P_\mu(C) = \int_0^1 \int_{\mathbb{M}_0} \mathrm{I}((t,b) \in C) f_\mu(t,b)\, \mathrm{d}t\, m(\mathrm{d}b).$$

It is convenient to write the density $f_\mu(t,b)$ as $f(t,b,\mu)$.

The next definition is made so as to be able to provide conditions under which bootstrapping performs pre-pivoting, in the sense of Beran (1988). The extent to which the statistic τ is *not* pivotal is conveniently measured by the quantity $\bar{\lambda}_0 \equiv \sup_{\mu,\nu \in \mathbb{M}_0} \lambda_0(\mu,\nu)$.

Definition 3: metric induced by bootstrap *P* value

The distance between two DGPs μ and ν in $\mathbb{M}_0$ in the metric induced by bootstrapping is defined as

$$\lambda_1(\mu,\nu) = \frac{1}{2}\{ \sup_{x \in [0,1]} |R_1(x,\mu) - R_1(x,\nu)| + |Q_1(x,\mu) - Q_1(x,\nu)| \}.$$

Recall that R_1 and Q_1 are respectively the CDF and quantile function of the bootstrap P value $p_1(\mu,\omega)$. The extent to which the bootstrap P value $p_1(\mu,\omega)$ is not pivotal can be quantified by $\bar{\lambda}_1 \equiv \sup_{\mu,\nu \in \mathbb{M}_0} \lambda_1(\mu,\nu)$.

Definition 4: prepivoting

The bootstrap is said to perform prepivoting for the model $\mathbb{M}_0$ if $\bar{\lambda}_1 < \bar{\lambda}_0$.

The next theorem provides a crude sufficient condition for the bootstrap to perform prepivoting.

Theorem PP: prepivoting

Let $\sup_{\mu \in \mathbb{M}_0} \lambda_0(\mu, \beta(\mu,\omega)) = \varepsilon$ almost surely. Further let the derivative R_0' of R_0 with respect to x be uniformly bounded: $\sup(|R_0'(x,\mu)| : x \in [0,1], \mu \in \mathbb{M}_0) = r$. Then, if $2r\varepsilon < \bar{\lambda}_0$, the bootstrap performs prepivoting for the model $\mathbb{M}_0$.

Proof : in the Appendix ■

Remarks:

It is clear that this theorem imposes much too stringent restrictions on the bootstrap. Since the main aim of this paper is to study the fast iterated bootstraps, we have made no attempt to prove a less crude theorem. Nonetheless, its relevance to iterated bootstrapping is clear. Distances $\lambda_k(\mu,\nu)$ can be defined using R_k instead of R_1 in Definition 3, and then one can seek conditions under which $\bar{\lambda}_{k+1} < \bar{\lambda}_k$. This can then be interpreted as iterated prepivoting.

4. The Fast Double Bootstrap

The fast double bootstrap (FDB) of Davidson and MacKinnon (2007) is based on two approximations. The first is to assume that, for any $\mu \in \mathbb{M}_0$, the random variables $\tau(\mu, \omega)$ and $\beta(\mu, \omega)$ are independent. The assumption is of course false except in special circumstances, but it holds asymptotically in many commonly encountered situations. By definition,

$$R_1(x, \mu) = P\{\omega \in \Omega \,|\, p_1(\mu, \omega) < x\} = \mathrm{E}[\mathrm{I}(R_0(\tau(\mu, \omega), \beta(\mu, \omega)) < x)]. \tag{11}$$

Use of the relation (3) between R_0 and Q_0 lets us write (11) as

$$R_1(x, \mu) = \mathrm{E}[\mathrm{I}(\tau(\mu, \omega) < Q_0(x, \beta(\mu, \omega))]$$

If $\tau(\mu, \omega)$ and $\beta(\mu, \omega)$ are treated as though they were independent, then we have

$$\begin{aligned} R_1(x, \mu) &= \mathrm{E}\Big[\mathrm{E}[\mathrm{I}(\tau(\mu, \omega) < Q_0(x, \beta(\mu, \omega)) \,|\, \beta(\mu, \omega)]\Big] \\ &= \mathrm{E}[R_0(Q_0(x, \beta(\mu, \omega)), \mu)] \end{aligned} \tag{12}$$

where the last step follows from the "Independence Lemma"– Lemma A.3 in Schilling and Partzsch (2012). Since in general $\tau(\mu, \omega)$ and $\beta(\mu, \omega)$ are not independent, (12) is taken as an approximation.

Assumption FDB1: first FDB approximation

There exists a (small) quantity $\eta_1 > 0$ such that, for all $\mu \in \mathbb{M}_0$, and for all $x \in [0, 1]$,

$$|R_1(x, u) - \mathrm{E}[R_0(Q_0(x, \beta(\mu, \omega)), \mu)] < \eta. \tag{13}$$

Consider now two identical probability spaces $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$, and their product space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P_1 \times P_2)$. Define the stochastic process

$$\tau^1 \;:\; \mathbb{M} \times (\Omega_1 \times \Omega_2) \to \mathbb{R}$$

by the formula

$$\tau^1(\mu, \omega_1, \omega_2) = \tau\big(\beta(\mu, \omega_1), \omega_2\big). \tag{14}$$

Thus $\tau^1(\mu, \omega_1, \omega_2)$ can be thought of as a realisation of the bootstrap statistic when the underlying DGP is μ. We denote the CDF of τ^1 under μ by $R^1(\cdot, \mu)$. Thus

$$\begin{aligned} R^1(x, \mu) &= (P_1 \times P_2)\big\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 \,|\, \tau\big(\beta(\mu, \omega_1), \omega_2\big) < x\big\} \\ &= \mathrm{E}[\mathrm{I}(\tau(\beta(\mu, \omega_1), \omega_2) < x)] \\ &= \mathrm{E}\Big[\mathrm{E}[\mathrm{I}(\tau(\beta(\mu, \omega_1), \omega_2) < x) \,|\, \mathcal{F}_1]\Big] \\ &= \mathrm{E}[R_0(x, \beta(\mu, \omega_1))]. \end{aligned} \tag{15}$$

In the third line above, $\mathcal{F}_1$ denotes the σ-algebra generated by deterministic functions of ω_1, and the last step follows from the Independence Lemma.

The next assumption underlying the FDB approximation extends Assumption 2 by requiring the function R_0 to be close to the CDF of the U(0,1) distribution and also requiring that R_0 should have a derivative uniformly close to unity.

Assumption FDB2: nearly uniform distribution

There exists a (small) quantity $\delta > 0$ such that, for all $\mu \in \mathbb{M}_0$ and for all $x \in [0,1]$,

$$|R_0(x,\mu) - x| < \delta \quad \text{and} \quad |R_0'(x,\mu) - 1| < \delta, \tag{16}$$

where $R_0'(x,\mu)$ is the derivative with respect to x.

Then we can show that

Lemma FDB: second FDB approximation

Let $\varepsilon = \delta/(1-\delta)$ with δ as defined in Assumption FDB2. Then for all $\mu \in \mathbb{M}_0$ and for all $x \in [0,1]$,

$$|\mathrm{E}[R_0(Q_0(x,\beta(\mu,\omega)),\mu)] - R_0(Q^1(x,\mu),\mu)| < 4\varepsilon^2, \tag{17}$$

where $Q^1(\cdot,\mu)$ is the quantile function inverse to the CDF $R^1(\cdot,\mu)$.

Proof : in the Appendix. ■

There follows from Assumption FDB1, Assumption FDB2, and Lemma FDB the approximation that underlines the FDB.

Theorem FDB: FDB approximation

Make the definition:

$$R_1^f(x,\mu) = R_0(Q^1(x,\mu),\mu). \tag{18}$$

Then, for all $\mu \in \mathbb{M}_0$ and for all $x \in [0,1]$,

$$|R_1(x,\mu) - R_1^f(x,\mu)| < \eta + 4\varepsilon^2, \tag{19}$$

where η is defined in Assumption FDB1, and ε in Lemma FDB.

Proof : in the Appendix. ■

Remarks:

If the argument leading to the FDB approximation were to be couched in asymptotic terms, then, for sample size n, η would be $O(n^{-1})$, and ε would be $O(n^{-1/2})$, and this would establish an asymptotic refinement for the FDB. But the assumptions underlying such an asymptotic argument are rather different from those used here in the finite-sample context.

If the realisations of $\tau(\mu,\omega)$ and $\beta(\mu,\omega)$ are respectively t and b, then, by analogy with the double-bootstrap P value (6), the FDB P value is

$$p_2^f \equiv R_1^f(R_0(t,b),b) = R_0(Q^1(t,b),b). \tag{20}$$

The algorithm for estimating the FDB P value by simulation is as follows.

Algorithm FDB

1. From the data set under analysis, compute the realised statistic t and the bootstrap DGP b.

2. Draw B bootstrap samples and compute B bootstrap statistics $t_j^* = \tau(b, \omega_j^*)$, $j = 1, \ldots, B$, and B iterated bootstrap DGPs $b_j^* = \beta(b, \omega_j^*)$, for B independent realisations ω_j^* from $(\Omega, \mathcal{F}, P)$.

3. Compute B second-level bootstrap statistics $t_j^{1*} = \tau(b_j^*, \omega_j^{**})$, where the ω_j^{**} are further independent realisations from $(\Omega, \mathcal{F}, P)$, and sort them in increasing order.

4. Compute the estimated first-level bootstrap P value $\hat{p}_1(t, b)$ as the proportion of the t_j^* smaller than t.

5. Obtain the estimate $\hat{Q}^1(\hat{p}_1(t, b), b)$ as the order statistic of the t_j^{1*} of rank $\lceil B\hat{p}_1(t, b) \rceil$. Denote this estimate as $\hat{Q}^{1*}$.

6. The estimated FDB P value is the proportion of the t_j^* that are smaller than $\hat{Q}^{1*}$.

The above algorithm produces the FDB P value for a test of some specific hypothesis. The FDB can also be used to construct a confidence interval for a parameter, θ say, that is defined within a model $\mathbb{M}$. The idea is conventional: for a confidence interval at confidence level $1 - \alpha$, the FDB hypothesis test is inverted in order to find the limits of the interval. Let the FDB P value used to test the hypothesis that $\theta = \theta_0$ be denoted as $p_2^f(\theta_0)$. The limit(s) of a confidence interval are then the solution(s) of the equation $p_2^f(\theta) = \alpha$, or, explicitly, the equation

$$R_0(Q^1(p_1(\theta_0), b), b) = \alpha, \tag{21}$$

where $p_1(\theta_0)$ is the first-level bootstrap P value. The following algorithm produces the FDB confidence interval based on an asymptotic t statistic.

Algorithm FDB confidence interval

1. From the data set under analysis, compute a point estimate $\hat{\theta}$ of the parameter of interest, along with a suitable standard error $\hat{\sigma}_\theta$, and a bootstrap DGP b for which the true value of the parameter is $\hat{\theta}$. The test statistic for testing the hypothesis that $\theta = \theta_0$ is $\tau(\theta_0) \equiv (\hat{\theta} - \theta_0)/\hat{\sigma}_\theta$.

2. Draw B bootstrap samples and compute B bootstrap statistics $t_j^* = (\theta^* - \hat{\theta})/(\sigma_\theta^*)_j$, where θ^* and $(\sigma_\theta^*)_j$ are respectively the point estimate and standard error from the j^{th} bootstrap sample, $j = 1, \ldots, B$. Also compute B iterated bootstrap DGPs b_j^*, as in the above Algorithm.

3. Compute the second-level bootstrap statistics $t_j^{1*} = (\theta^{**} - \theta^*)/(\sigma_\theta^{**})_j$, where θ^{**} and $(\sigma_\theta^{**})_j$ are computed from the sample generated by b_j^*.

4. Estimate the distribution of the statistic under the null by the empirical distribution of the bootstrap statistics t_j^*: $\hat{R}_0(x) = B^{-1} \sum_{j=1}^B \mathrm{I}(t_j^* < x)$. This is the estimate of the function $R_0(\cdot, b)$ needed for (21).

5. Estimate the distribution of the double bootstrap statistic by the empirical distribution of the t_j^{1*}: $\hat{R}^1(x) = B^{-1} \sum_{j=1}^B \mathrm{I}(t_j^{1*} < x)$, and the corresponding quantile function $\hat{Q}^1(p) = t^{1*}_{(\lceil p(B+1) \rceil)}$, where the notation means the order statistic of the t_j^{1*} of order $\lceil p(B+1) \rceil$. $\hat{Q}^1(\cdot)$ is the estimate of the function $Q^1(\cdot, b)$ in (21). In addition, we have the estimate $\hat{R}_1^f(x) = \hat{R}_0(\hat{Q}^1(x))$.

6. The (estimated) single bootstrap P value for testing the hypothesis $\theta = \theta_0$ is $\hat{p}_1(\theta_0) = \hat{R}_0((\hat{\theta} - \theta_0)/\hat{\sigma}_\theta)$. The (estimated) FDB P value to test $\theta = \theta_0$ is then, as in (21), $\hat{R}_1^f(\hat{p}_1(\theta_0))$.

7. The equation $\hat{R}_1^f(\hat{p}_1(\theta_0)) = \alpha$ can then be solved. A direct method is to use a one-dimensional root finder, such as bisection or *regula falsi*. Alternatively, we may take the estimate of equation (21) apart by using a chain of equivalent equations, as follows.

$$\begin{aligned}\hat{p}_2^f(\theta) = \alpha &\iff \hat{R}_1^f(\hat{p}_1(\theta)) = \alpha \iff \hat{R}_0\Big(\hat{Q}^1(\hat{R}_0(\tau(\theta)))\Big) = \alpha \\ &\iff \hat{Q}^1(\hat{R}_0(\tau(\theta))) = \hat{Q}_0(\alpha) \iff \hat{R}_0(\tau(\theta)) = \hat{R}^1(\hat{Q}_0(\alpha)) \\ &\iff \tau(\theta) = \hat{Q}_0\Big(\hat{R}^1(\hat{Q}_0(\alpha))\Big).\end{aligned}$$

Here the function $\hat{Q}_0(p)$ is defined as an order statistic of the t_j^*: $\hat{Q}_0(p) = t^*(\lceil p(B+1) \rceil)$.

Remarks:

The two methods of solving the equation for the limit of a confidence interval do not give numerically identical results in general. This is just a consequence of the discreteness of the estimated bootstrap distribution, which means that there is a range of values of θ for which the P value is equal to any given α.

Chang and Hall (2015) show, using Edgeworth expansions, that the FDB does not improve asymptotically on the single bootstrap for the construction of confidence intervals. Thus Algorithm FDB confidence interval may not be as reliable as Algorithm FDB for P values. If this is a concern, the same reliability as that given by FDB P values can be had by solving, not (21), but rather the equation

$$R_0(Q^1(p_1(\theta_0), b(\theta_0)), b(\theta_0)) = \alpha,$$

where $b(\theta_0)$ is a bootstrap DGP for which the true value of the parameter θ is θ_0. For full details of this method, see Davidson and MacKinnon (2010).

5. The Fast Triple Bootstrap

In order to study the distribution of the FDB P value (20), we wish to evaluate the expression

$$P(p_2^f < x) = \mathrm{E}\Big[\mathrm{I}\Big(R_0(Q^1(R_0(\tau(\mu,\omega), \beta(\mu,\omega)), \beta(\mu,\omega)), \beta(\mu,\omega)) < x\Big)\Big],$$

which is the probability, under the DGP μ, that the FDB P value is less than x. The inequality that is the argument of the indicator above is equivalent to several other inequalities, as follows:

$$\begin{aligned}&Q^1(R_0(\tau(\mu,\omega), \beta(\mu,\omega)), \beta(\mu,\omega) < Q_0(x, \beta(\mu,\omega)) \\ \iff\ &R_0(\tau(\mu,\omega), \beta(\mu,\omega)) < R^1(Q_0(x, \beta(\mu,\omega)), \beta(\mu,\omega)) \\ \iff\ &\tau(\mu,\omega) < Q_0(R^1(Q_0(x, \beta(\mu,\omega)), \beta(\mu,\omega)), \beta(\mu,\omega)).\end{aligned}$$

At this point, we can again invoke an approximation that would be exact if $\tau(\mu,\omega)$ and $\beta(\mu,\omega)$ were independent. The final inequality above separates $\tau(\mu,\omega)$ from $\beta(\mu,\omega)$ on the left- and right-hand sides respectively, and so the expectation of the indicator of that inequality is approximated by

$$\mathrm{E}\Big[R_0(Q_0(R^1(Q_0(x, \beta(\mu,\omega)), \beta(\mu,\omega)), \beta(\mu,\omega)), \mu)\Big]. \tag{22}$$

It is quite possible to estimate (22) for given x and μ by simulation. The function $R_0(\cdot, \mu)$ can be estimated as in (4). But, for the other needed functions, $Q_0(\cdot, \beta(\mu, \omega))$ and $R^1(\cdot, \beta(\mu, \omega))$, which both depend on ω, a second, inner, loop is necessary for each iteration of the main simulation loop. Thus the straightforward estimation procedure is about as computationally intensive as the double bootstrap.

However, we can make a further approximation in the spirit of (17), the second of the approximations that lead to the FDB. The aim is to use our smallness assumptions in order to eliminate all mention of the bootstrap DGP $\beta(\mu, \omega)$, the presence of which requires the inner loop.

Since the derivatives of R_0, Q_0, and R^1 with respect to their first argument are close to 1 by Assumption FDB2, it follows that the derivative of $R_0(Q_0(R^1(x, \beta(\mu, \omega)), \beta(\mu, \omega), \mu)$ with respect to x is also equal to 1 plus a small quantity, which we may call ε, as in the statement of Lemma FDB, although it is not in general exactly the same quantity. Therefore, by the mean-value theorem,

$$\begin{aligned} R_0(Q_0(R^1(Q_0(x, \beta(\mu, \omega)), \beta(\mu, \omega)), \beta(\mu, \omega)), \mu) - R_0(Q_0(R^1(Q^1(x, \mu), \beta(\mu, \omega)), \beta(\mu, \omega), \mu) \\ = (1 + \varepsilon)(Q_0(x, \beta(\mu, \omega)) - Q^1(x, \mu)), \end{aligned}$$

It is shown in the proof of Lemma FDB that the expectation of the right-hand side above is a small quantity of the order of ε^2, and so the expectation (22) is, up to an error of that order,

$$\mathrm{E}[R_0(Q_0(R^1(Q^1(x, \mu), \beta(\mu, \omega)), \beta(\mu, \omega)), \mu)]. \tag{23}$$

Analogously to (14), define the random variable

$$\tau^2(\mu, \omega_1, \omega_2, \omega_3) = \tau(\beta(\beta(\mu, \omega_1), \omega_2), \omega_3),$$

which can be thought of as a realisation of the second-order bootstrap statistic. The CDF of τ^2 under μ, denoted by $R^2(\cdot, \mu)$ is given by

$$\begin{aligned} R^2(x, \mu) &= \mathrm{E}[\mathrm{I}(\tau^2(\mu, \omega_1, \omega_2, \omega_3) < x)] \\ &= \mathrm{E}[\mathrm{E}[\mathrm{I}(\tau(\beta(\beta(\mu, \omega_1), \omega_2), \omega_3) < x) \,|\, \mathcal{F}_{12}]] \\ &= \mathrm{E}[R_0(x, \beta(\beta(\mu, \omega_1), \omega_2))] \\ &= \mathrm{E}[R^1(x, \beta(\mu, \omega_1))], \end{aligned} \tag{24}$$

where $\mathcal{F}_{12}$ denotes the product sigma-algebra defined on the probability space of ω_1 and ω_2. The third equality follows from the Independence Lemma and the definition of R_0, the fourth from the relation (12).

Now, an argument just like that leading to (23), but based on (24), shows that the expectation (23) is equal to

$$\mathrm{E}[R_0(Q_0(R^2(Q^1(x, \mu), \mu), \beta(\mu, \omega)), \mu)]$$

up to an error of order ε^2. Finally, we can use the argument in the proof of Lemma FDB to show that this last expression is, always up to an error of second order in ε,

$$R_2^f(x, \mu) \equiv R_0(Q^1(R^2(Q^1(x, \mu), \mu), \mu), \mu). \tag{25}$$

The theoretical FDB P value (20) is the approximation $R_1^f(x, \mu)$ to $R_1(x, \mu)$, as defined in (18), evaluated with x set equal to the first-level bootstrap P value, and μ replaced by the bootstrap

DGP. The theoretical fast triple bootstrap (FTB) P value is formed analogously from (25) by setting x equal to the FDB P value, and again replacing μ by the (first-level) bootstrap DGP, according to the bootstrap principle. The result is

$$p_3^f(\mu,\omega) \equiv R_0(Q^1(R^2(Q^1(p_2^f(\mu,\omega),\beta(\mu,\omega)),\beta(\mu,\omega)),\beta(\mu,\omega)),\beta(\mu,\omega)), \tag{26}$$

with p_2^f given by (20). The simulation estimate, which must be expressed as a function of the observed statistic t and bootstrap DGP b, is

$$\hat{p}_3^f(t,b) = \hat{R}_0(\hat{Q}^1(\hat{R}^2(\hat{Q}^1(\hat{p}_2^f(t,b),b),b),b),b), \tag{27}$$

with $\hat{p}_2^f(t,b)$ given by Algorithm FDB.

Here is the algorithm for the FTB P value.

Algorithm FTB

1. From the data set under analysis, compute the realised statistic t and the bootstrap DGP b.
2. Draw B bootstrap samples and compute B bootstrap statistics $t_j^* = \tau(b,\omega_j^*)$, $j = 1,\ldots,B$, and B iterated bootstrap DGPs $b_j^* = \beta(b,\omega_j^*)$.
3. Compute B second-level bootstrap statistics $t_j^{1*} = \tau(b_j^*,\omega_j^{**})$, and sort them in increasing order. At the same time, compute the corresponding second-level bootstrap DGPs $b_j^{**} = \beta(b_j^*,\omega_j^{**})$.
4. Compute B third-level bootstrap statistics $t_j^{2*} = \tau(b_j^{**},\omega_j^{***})$.
5. Compute the estimated first-level bootstrap P value $\hat{p}_1(t,b)$, as the proportion of the t_j^* smaller than t.
6. Obtain the estimate $\hat{Q}^{1*} \equiv \hat{Q}^1(\hat{p}_1(t,b),b)$ as the order statistic of the t_j^{1*} of rank $\lceil B\hat{p}_1(t,b)\rceil$.
7. Compute the estimated FDB P value $\hat{p}_2^f(t,b)$ as the proportion of the t_j^* smaller than $\hat{Q}^{1*}$.
8. Compute $\hat{Q}^{1**} \equiv \hat{Q}^1(\hat{p}_2^f(t,b),b)$ as the order statistic of the t_j^{1*} of rank $\lceil B\hat{p}_2^f(t,b)\rceil$.
9. Compute $\hat{R}^{2*} \equiv \hat{R}^2(\hat{Q}^1(\hat{p}_2^f(t,b),b),b)$ as the proportion of the t_j^{2*} smaller than $\hat{Q}^{1**}$.
10. Compute $\hat{Q}^{1***} \equiv \hat{Q}^1(\hat{R}^2(\hat{Q}^1(\hat{p}_2^f(t,b),b),b),b)$ as the order statistic of the t_j^{1*} of rank $\lceil r\hat{R}^{2*}\rceil$.
11. Compute $\hat{p}_3^f(t,b)$ as the proportion of the t_j^* smaller than $\hat{Q}^{1***}$.

Remark:

It can be hoped that, if the FDB improves on the single bootstrap, the FTB should improve on the FDB to a similar extent, since the approximations used in deriving the theoretical FTB P value (26) hold with the same accuracy as those used to obtain p_2^f. That this is so in circumstances in which the FDB works reasonably well emerges from the simulations reported in Section 7.

6. Fast Higher-Order Bootstraps

The ideas that lead to the FDB and FTB P values can obviously be extended to higher orders. For the FDB, we approximate the distribution of the first-level bootstrap P value $p_1(\mu,\omega)$, and evaluate it at the computed first-level P value $p_1(t,b)$ and the bootstrap DGP b. For the FTB, we approximate the distribution of the FDB P value $p_2^f(\mu,\omega)$ and evaluate it at the computed FDB P value $p_2^f(t,b)$ and b. For a fast quadruple bootstrap, we wish to approximate the distribution of the FTB P value $p_3^f(\mu,\omega)$ and evaluate it at the computed FTB P value $p_3^f(t,b)$ and b. And so on.

The approximate CDFs R_1^f and R_2^f are given explicitly by (18) and (25). We define higher-order approximate CDFs and fast higher-order bootstrap P values recursively, as follows:

$$R_k^f(x,\mu) \approx \mathrm{E}[\mathrm{I}(p_k^f(\mu,\omega) < x)], \text{ and} \tag{28}$$

$$p_{k+1}^f(\mu,\omega) = R_k^f(p_k^f(\mu,\omega), b(\mu,\omega)), \tag{29}$$

where the exact nature of the approximation in (28) above will now be made explicit.

For the following discussion, the notation t stands for $\tau(\mu,\omega)$, and b stands for $\beta(\mu,\omega)$. We need not distinguish between the random variables and their realisations. Thus $p_2^f(t,b)$ means $p_2^f(\mu,\omega)$, and so forth. By (18), the approximate CDF $R_1^f(x,\mu)$ of $p_1(t,b)$ is $R_0(Q^1(x,\mu),\mu))$. By (25), $R_2^f(x,\mu)$ is $R_0(Q^1(R^2(Q^1(x,\mu),\mu),\mu),\mu)$. As we will see, the pattern for R_k^f is the composition of 2^k functions with R and Q alternating. In order to see how to determine what these functions are, we consider explicitly the case of the fast quadruple bootstrap, which will let us describe the iterative procedure necessary for the explicit expression of R_k^f for general k.

The explicit expression of $p_3^f(t,b)$ is given by (26) and (27), and for present purposes it can be written as

$$p_3^f(t,b) = R_0(Q^1(R^2(Q^1(p_2^f(t,b),b),b),b),b). \tag{30}$$

In order to approximate its distribution, we define the random variable

$$\tau^3(\mu,\omega_1,\omega_2,\omega_3,\omega_4) = \tau(\beta(\beta(\beta(\mu,\omega_1),\omega_2),\omega_3),\omega_4).$$

Its CDF is readily seen to be

$$R^3(x,\mu) = \mathrm{E}[R^2(x,\beta(\mu,\omega))].$$

The corresponding quantile function, $Q^3(x,\mu)$ is such that

$$\mathrm{E}[Q^2(x,\beta(\mu,\omega))] = Q^3(x,\mu)$$

with error small of second order. By extension, it is obvious how to define τ^k, R^k and Q^k.

The CDF of $p_3^f(t,b)$ evaluated at x is $\mathrm{E}[\mathrm{I}(p_3^f(t,b) < x)]$. By use of (30), we see that the inequality in the indicator here can be expressed as

$$p_2^f(t,b) < R^1(Q^2(R^1(Q_0(x,b),b),b),b). \tag{31}$$

The probability that this inequality is satisfied under μ, conditional on b, can be approximated by the approximate CDF R_2^f of $p_2^f(t,b)$ evaluated at the right-hand side of (31) and μ. By (25), this is

$$R_0(Q^1(R^2(Q^1(R^1(Q^2(R^1(Q_0(x,b),b),b),b),\mu),\mu),\mu),\mu). \tag{32}$$

Then $R_3^f(x,\mu)$ is the unconditional expectation of this expression.

An argument by now familiar shows that this unconditional expectation can be approximated by

$$\begin{aligned}&\mathrm{E}[R_0(Q^1(R^2(Q^1(R^1(Q^2(R^1(Q^1(x,\mu),b),b),b),\mu),\mu),\mu),\mu)] = \dots\\ \dots &= R_0(Q^1(R^2(Q^1((R^2(Q^3(R^2(Q^1(x,\mu),\mu),\mu),\mu),\mu),\mu),\mu),\mu),\mu).\end{aligned} \tag{33}$$

If we compare the sequence of functions in the expression (33) of $R_3^f(x,\mu)$ and the expression (25) of $R_2^f(x,\mu)$, we see that there are exactly twice as many in the former as in the latter. This arises because the inequality (31) has the four functions of p_2^f, and they are then the final four in (32), preceded by the four functions of R_2^f. We may observe that in (32) the composition of the final four functions is the inverse of the composition of the first four. When we get to (33), the indices of the final four functions have all been raised by 1.

By the definition (29), the fourth-level P value $p_4^f(t,b)$ is $R_3^f(p_3^f(t,b),b)$. It is not hard to check that, in the explicit expression of p_4^f, we have, first, the eight functions in (33), followed by the inverse of their composition with indices raised by 1, that is, the composition of the sixteen functions that we write in sequence as follows:

$$R_0Q^1R^2Q^1R^2Q^3R^2Q^1R^2Q^3R^4Q^3R^2Q^3R^2Q^1.$$

Although the way in which we have arrived at this sequence of functions is easy enough to describe, the explicit structure seems not to be expressible in closed form other than by actually working it out.

The following algorithm provides an equivalent but slightly easier way to derive the sequence of functions in R_{k+1}^f or p_{k+1}^f from the sequence in R_k^f and p_k^f.

Algorithm FkB:

1. Divide the sequence S_k of the 2^k functions in R_k^f into two sequences A_k and B_k, of length 2^{k-1} each, such that $S_k = A_kB_k$.
2. Obtain the sequence C_k of functions the composition of which is the inverse of the composition of the functions in the sequence B_k.
3. Obtain a new sequence D_k by incrementing the indices of the elements of the sequence C_k by 1.
4. The sequence S_{k+1} used to define R_{k+1}^f and p_{k+1}^f is $A_kB_kD_kB_k$, of length 2^{k+1}.

7. Illustrations

In this section, we present the results of various simulation experiments designed to see to what extent the fast double and triple bootstraps can improve the reliability of inference. The first experiments, which deal with a test for a unit root, make use of a parametric bootstrap that makes the distributions of bootstrap statistics absolutely continuous. The other two sets of experiments, one dealing with a test for an ARCH effect, the other a test for serial correlation of the disturbances in a regression model, use resampling bootstraps, which lead to bootstrap statistics with discrete distributions, in violation of Assumption 1. Since resampling is far and away the most frequently used form of bootstrapping in empirical work, it is highly desirable to see if the fast iterated bootstraps suffer noticeably from the discrete distributions induced by resampling.

Testing for a unit root

There are well-known difficulties in testing for a unit root in a series obtained by summing a stationary series that is an MA(1) process with a parameter θ close to -1. Unless special precautions are taken, augmented Dickey-Fuller (ADF) tests can suffer from severe overrejection under the null hypothesis of a unit root, on account of the near cancellation of the unit root by the MA component in the driving stationary series. We may cite Schwert (1989) and Perron and Ng (1996) in this regard.

Over the last decade or so, various bootstrap techniques have been proposed as a means of improving size distortions in the unit-root testing literature. One bootstrap technique designed to deal with autoregressive models is the so called sieve bootstrap first proposed by Bühlmann (1997). Simulation evidence demonstrates that this bootstrap approach has certain appeal in reducing size distortions. Here it is appropriate to cite Psaradakis (2001), Chang and Park (2002), and Palm, Smeekes, and Urbain (2008) as evidence of this fact. These papers show the ability of bootstrapped ADF statistics to outperform their asymptotic counterparts. However, despite the ability of the sieve bootstrap to reduce size distortions in certain cases, the gain is really unimpressive in the case considered in our simulations, in which the MA component has a parameter close to -1. In this regard, Richard (2009) applies several variations of the FDB to show that size distortions can be significantly reduced by imposing certain linear restrictions on the truncation parameter of the bootstrap ADF regression and the truncation parameter of the bootstrap sieve.

In Davidson (2010), it is shown that, under the assumption that the MA(1) process has Gaussian innovations, parametric FDB tests can significantly reduce the size distortion of an ADF test. It is argued that bootstrapping the FDB should reduce the distortions still further. This is of course very computationally demanding. In what follows, we show that comparable results may be obtained far more economically with the fast triple bootstrap.

The model studied in this section may be summarised as follows:

$$y_t = \rho y_{t-1} + v_t \tag{34}$$

$$v_t = u_t + \theta u_{t-1}, \qquad u_t \sim \text{NID}(0, \sigma^2), \qquad t = 1, \dots, n. \tag{35}$$

The observed series is y_t, and the null hypothesis of a unit root sets $\rho = 1$. Under that hypothesis, $v_t = \Delta y_t$, where Δ is the first-difference operator. We may write (35) in vector notation using the lag operator L, as follows:

$$\boldsymbol{v} = (1 + \theta \text{L})\boldsymbol{u}, \quad \text{or} \quad \boldsymbol{u} = \boldsymbol{v} - R(\text{L})\boldsymbol{v},$$

where we define $R(\text{L}) = \theta(1 + \theta\text{L})^{-1}\text{L}$. Davidson (2010) demonstrates that θ may be estimated by a two-stage nonlinear least squares regression using the model:

$$\boldsymbol{v} = \phi \boldsymbol{e}_1 + R(\text{L})(\boldsymbol{v} - \phi \boldsymbol{e}_1) + \boldsymbol{u}, \tag{36}$$

where $\boldsymbol{e}_1$ is the vector with first element equal to 1 and all others zero. In the first stage, ϕ is set to zero and a preliminary estimate of θ is obtained. Then an estimate of ϕ is given by $\boldsymbol{s}^\top(1 - R(\text{L}))\boldsymbol{u}/\boldsymbol{s}^\top\boldsymbol{s}$, where $\boldsymbol{s} = (1 - R(\text{L}))\boldsymbol{e}_1$. Finally, this estimate of ϕ is used as a known constant in (36), which is re-estimated to obtain the second-stage estimate of θ.

Testing for a unit root in (34) proceeds by computing an ADF statistic using the ADF regression:

$$\Delta y_t = \beta_0 + \beta_1 y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \text{residual}. \tag{37}$$

The statistic is the standard t-statistic for the null hypothesis $\beta_1 = 0$ when (37) is estimated by ordinary least squares. The ADF statistic, although easy to compute, has a non-standard asymptotic distribution which is that of a functional of Brownian motion that depends on no nuisance parameters. The ADF statistic is thus an asymptotic pivot.

A suitable bootstrap DGP can be expressed as follows:

$$\boldsymbol{u}^* \sim \mathrm{N}(\mathbf{0}, \mathbf{I}), \quad \boldsymbol{v}^* = (1 + \hat{\theta}\mathrm{L})\boldsymbol{u}^*, \quad \boldsymbol{y}^* = (1 - \mathrm{L})^{-1}\boldsymbol{u}^*, \tag{38}$$

where $\hat{\theta}$ is obtained by the procedure outlined above, with $\boldsymbol{v} = \Delta\boldsymbol{y} = (1 - \mathrm{L})\boldsymbol{y}$. Note that, since the ADF statistic is scale-invariant, we can set $\sigma = 1$. Thus the bootstrap DGP is completely determined by one single parameter, the estimate $\hat{\theta}$.

In the simulation experiments we now discuss, the data were generated by the model defined by (34) and (35), with $\rho = 1$ and for various values of θ close to -1. For any given estimate $\hat{\theta}$, the bootstrap DGP (38) was used. Second-level bootstrap DGPs were formulated by generating bootstrap data with $\hat{\theta}$ in (35), and then obtaining the estimate of θ for these data. This estimate, which we denote $\hat{\theta}^*$, then characterises a second-level bootstrap DGP, which can subsequently be used to generate second-level bootstrap data, used to obtain an estimate $\hat{\theta}^{**}$, which characterises a third-level bootstrap DGP. This experimental design is the same as the one used in Davidson (2010).

Figures 1 and 2 below show the errors in rejection probability (ERP), that is, the difference between the experimentally observed rejection rate and the nominal level α, of the ordinary single bootstrap, the FDB, and the FTB for varying values of the MA(1) parameter θ, for sample sizes $n = 50$ and 100, and for all nominal levels $0 \leq \alpha \leq 1$. Although for practical purposes one need not consider values of α greater than 0.1, using all possible values allows a clearer view of how the distributions of the various bootstrap P values differ from the ideal U(0,1) distribution. All graphs in Figure 1 were computed with $N = 25{,}000$ replications, $B = 4{,}999$ bootstrap repetitions, and ADF truncation parameter $p = 12$. The time required to complete the simulations using a computer cluster with 54 nodes with 4 CPUs each was roughly 16 hours and 22 hours for sample sizes $n = 50$ and $n = 100$ respectively.

We see that the FTB P values tend to have lower ERPs than their FDB and single bootstrap (BS) counterparts. As expected, the size correction is not very impressive for $\theta = -0.99$, but as θ increases away from $\theta = -1$ improvements can be drastic. What is interesting however is that when θ is very close to -1 as in Figure 1, the FTB ERP is slightly worse than the FDB ERP, which in turn is slightly worse than the single bootstrap ERP in the region of the nominal size which is most relevant to practitioners, namely $0 \leq \alpha \leq 0.10$.

An unexpected and somewhat strange feature of the results shown in the figures is that the distortions for $n = 50$ are often smaller than those for $n = 100$. This is just another manifestation of the failure of inference in the close vicinity of $\theta = -1$.

Figures 3 and 4 are like Figures 1 and 2, but they omit the case of the single bootstrap, and include results for the standard double and triple bootstraps. They show that the behaviour of the fast iterated bootstraps is very similar to that of the standard iterated bootstraps. The latter are slightly less distorted than the former, but the fast triple bootstrap usually outperforms the standard double bootstrap. Some caution is required in comparing the experimental results in Figures 3 and 4, because, for reasons of computational feasibility, the results for the standard iterated bootstraps were obtained with $N = 12{,}500$, $B = 99$, $B_1 = 119$, and $B_2 = 129$. in order to complete the simulations in a reasonable time frame. Even so, using all 216 CPUs of the cluster, it

took over 24 hours to complete a simulation for a single value of θ with a sample size $n = 50$. But, even allowing for non-negligible simulation randomness, the ordering of the various procedures is clearly established.

Before ending this section, we must remark that the model considered here is much too restrictive to have any practical interest. We have studied it as a computationally feasible test bed for the fast and standard iterated bootstraps, and to demonstrate that going as far as the fast triple bootstrap does indeed have the possibility of giving improved reliability over the single and fast double bootstraps.

This model was used in a quite different context in Davidson (2016a) to investigate standard bootstrap iteration by making use of a discretised version of the model. It is shown there that the sequence of iterated bootstrap P values does converge eventually, but not necessarily to a limit useful for inference, at least in the neighbourhood of $\theta = -1$.

A test for ARCH

In Davidson and MacKinnon (2007), one of the examples used to show how the FDB improves on the single bootstrap is a test for ARCH disturbances in a linear regression model. Since the seminal work of Engle (1982), it has been recognized that serial dependence in the variance of the disturbances of regression models using time-series data is a very common phenomenon. It is therefore usually advisable to test for the presence of such serial dependence.

Consider the linear regression model

$$\begin{aligned} y_t &= \boldsymbol{X}_t\boldsymbol{\beta} + u_t, \quad u_t = \sigma_t\varepsilon_t, \quad t = 1, \dots, n, \\ \sigma_t^2 &= \sigma^2 + \gamma u_{t-1}^2 + \delta\sigma_{t-1}^2, \quad \varepsilon_t \sim \text{IID}(0,1). \end{aligned} \tag{39}$$

The disturbances of this model follow the GARCH(1,1) process introduced by Bollerslev (1986). The easiest way to test the null hypothesis that the u_t are IID in the model (39) is to run the regression

$$\hat{u}_t^2 = b_0 + b_1\hat{u}_{t-1}^2 + \text{residual}, \tag{40}$$

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.

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. As mentioned above, this is in violation of Assumption 1. Resampling gives rise to a discrete distribution, although, for samples of reasonable size, it may be close enough to being continuous for the discreteness not to matter. It is of course of interest to see to what extent the theory of fast iterated bootstraps can be used effectively with resampling. Alternatively, the discrete distribution of the resampled objects can be smoothed. Davidson and MacKinnon tried this with the model (39), and found that it made a noticeable, but very slight, difference to the performance of both single and fast double bootstraps. In the experiments described in this section, we do not smooth.

The experimental design is copied from Davidson and MacKinnon (2007). In all cases, $\boldsymbol{X}_t$ consists of a constant and two independent, standard normal random variates, since changing the number of regressors has only a modest effect on the finite-sample behaviour of the tests. The sample size takes on the values 40, 80, and 160, a small subset of the set of sample sizes studied by Davidson

and MacKinnon. In order to have non-negligible ERPs, the ε_t are drawn from the χ^2_2 distribution, subsequently centred and rescaled to have variance 1. 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 as bootstrap DGP the following:

$$y_t^* = u_t^*, \quad u_t^* \sim \text{EDF}(y_t),$$

where the notation EDF (for "empirical distribution function") means simply that the bootstrap data are resampled from the original data. For iterated bootstraps, y_t^{**} is resampled from the y_t^*, and y_t^{***} is resampled from the y_t^{**}.

The experiments consisted of $N = 10{,}000$ replications with $B = 399$ bootstrap repetitions each. For each replication, P values were computed for the asymptotic test, the single bootstrap, the FDB, and the FTB. The results under the null are displayed in Figure 5. Observe that there is little difference in the ERPs for $n = 80$ and $n = 160$. The ordering of the four procedures is quite clear. The single bootstrap improves considerably on the asymptotic test based on the χ^2_1 distribution, the FDB again considerably on the single bootstrap, and the FTB has no significant distortion at all for $n = 80$ and $n = 160$, and very little for $n = 40$.

One might wonder if the excellent behaviour of the FTB is bought at the cost of diminished power. That this is not the case is demonstrated by the experiments the results of which are shown in Figure 6. The simulated data were generated by the DGP of model (39) with $\sigma^2 = 1$, $\gamma = 0.3$, and $\delta = 0.3$. These values were chosen in order to have some power for $n = 40$ and not to have so much power for $n = 160$ for an ordering of the results to be unclear. In this figure, the experimental rejection rate is plotted as a function of the *nominal* level of the test; see Horowitz and Savin (2000) and Davidson and MacKinnon (2006) for discussion of why this makes more sense than attempting any sort of "size adjustment".

The under-rejection of the asymptotic test under the null is carried over under the DGP studied here, to the extent that, for $n = 40$, the rejection rate of the test is smaller than the nominal level for some values of the latter. This is no longer true for the larger values of n. The three bootstrap tests have very similar rejection rates, always greater than that of the asymptotic test, and with the FDB and FTB having slightly greater power than the single bootstrap test.

A test for serial correlation

Another of the examples of the good performance of the FDB found in Davidson and MacKinnon (2007) is given by the Durbin-Godfrey test for serial correlation of the disturbances in a linear regression model. The test was introduced in Durbin (1970) and Godfrey (1978). The model that serves as the alternative hypothesis for the test is the linear regression model

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + \gamma y_{t-1} + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{IID}(0, \sigma^2), \quad t = 1, \dots, n, \tag{41}$$

where $\boldsymbol{X}_t$ is a $1 \times k$ vector of observations on exogenous variables. The null hypothesis is that $\rho = 0$. Let the OLS residuals from running regression (41) be denoted $\hat{u}_t$. Then the Durbin-Godfrey (DG) test statistic is the t statistic for $\hat{u}_{t-1}$ in a regression of y_t on $\boldsymbol{X}_t$, y_{t-1}, and $\hat{u}_{t-1}$. It is asymptotically distributed as $\text{N}(0, 1)$ under the null hypothesis. Since this test can either overreject or underreject in finite samples, it is natural to use the bootstrap in an effort to improve its finite-sample properties.

For the bootstrap DGP, from running regression (41), we obtain estimates $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}$, as well as the residuals $\hat{u}_t$. The semiparametric bootstrap DGP can be written as

$$y_t^* = \boldsymbol{X}_t\hat{\boldsymbol{\beta}} + \hat{\gamma} y_{t-1}^* + u_t^*, \tag{42}$$

where the u_t^* are obtained by resampling the residuals rescaled as $(n/(n-k-1))^{1/2}\hat{u}_t$. The initial value y_0^* is set equal to the actual pre-sample value y_0.

Our experimental design is similar to that in Davidson and MacKinnon (2007). We set $k = 6$, with the first regressor a constant, and the remaining five generated by independent, stationary AR(1) processes with normal innovations and parameter $\rho_x = -0.8$. The disturbances ε_t are normally distributed with $\sigma = 10$. We put $\boldsymbol{\beta} = \mathbf{0}$ and $\gamma = 0.75$. We look at sample sizes $n = 20$ and 40 under the null, and also at $n = 56$ in our study of power. These choices are sufficient for us to distinguish clearly the behaviour of the various testing procedures: asymptotic, single bootstrap, FDB, and FTB. As before, we used 10,000 replications each involving 399 bootstrap repetitions.

Figure 7 shows the ERPs under the null. For $n = 40$, we do not show the ERP of the asymptotic test, because it is so great that including it in the graph would make the ERPs of the other tests indistinguishable. All the bootstrap procedures are much less distorted than the asymptotic test, and they have, once again, the same ordering as before, with the FTB the least distorted. Even for as small a sample size as 40, its observed rejection rate is never different from the nominal level by more than 0.01.

Power is illustrated in Figure 8. The data for these experiments were generated with the autocorrelation parameter ρ in (41) equal to 0.5, $\sigma = 1$, and the parameter ρ_x used in generating the regressors equal to 0.8. These values were chosen on the basis of the results in Davidson and MacKinnon (2007), where they give rise to significant differences in the power of the single bootstrap and the FDB. For $n = 20$, there is no visible power at all; in fact there is some slight underrejection by all the tests. With $n = 40$, the asymptotic test rejects slightly more often than the bootstrap tests, and, for $n = 56$, all tests have very similar rejection rates. Once again, the superior performance of the fast iterated bootstrap tests does not entail reduced power.

8. Conclusions

The approximations that led to the fast double bootstrap have been extended not only to a fast triple bootstrap but to arbitrary levels of bootstrap iteration. Algorithms have been given for the implementation of the fast double and fast triple bootstraps, along with an algorithm for generating the expressions to be estimated by simulation for higher-level iterated bootstraps.

The approximations underlying the fast iterated bootstraps are justified only under two assumptions. The first is that the statistic being bootstrapped is approximately independent of the bootstrap DGP, for instance, if asymptotic independence holds, and the second is that all bootstrap DGPs generate statistics of which the distributions are absolutely continuous with respect to Lebesgue measure on the real line. Although the second assumption is not satisfied if a resampling bootstrap is used, the fast triple bootstrap is shown to suffer from less size distortion under the null than either the standard or the fast double bootstrap in a couple of simulation experiments. In the experiments with a unit root test, a parametric bootstrap is used, so that the second assumption is satisfied, but the first decidedly is not. Even so, the experiments show that the fast double and triple bootstraps suffer from only a little more distortion than their standard counterparts, with the same ranking of the various procedures as seen in the resampling examples. Further, the resampling examples demonstrate that improved behaviour under the null is not achieved at the cost of reduced power.

With regard to the assumption of approximate independence, it it does not hold, then the bootstrap can be expected to perform badly; see Davidson (2016b). Quite generally, then, the fast

iterated bootstraps improve reliability of inference when the bootstrap is already reeasonbly reliable, but whether they do so in other cases is undetermined in general.

The few experiments we report about the standard triple bootstrap required enormous computing resources. With current technology, and indeed with technology likely to be available in the near future, experiments on standard iterated bootstraps beyond the triple would consume computing power well beyond that accessible to academic researchers. The fast variants are much less computationally intensive, and their demands increase only linearly with the level of iteration, while those of the standard iterated bootstraps increase exponentially. For a single bootstrap P value, one has to evaluate $1 + B$ statistics and just one bootstrap DGP. For the level-k fast iterated bootstrap, one needs $1 + kB$ statistics, and $1 + (k - 1)B$ DGPs. But for the level-k standard iterated bootstrap, if for simplicity we do not use different numbers of bootstraps at each level, the number of statistics needed is $(B^{k+1} - 1)/(B - 1)$ and the number of DGPs is $(B^k - 1)/(B - 1)$.

Interesting theoretical questions remain, to do with the convergence or otherwise of the sequence of iterated bootstrap P values, fast and standard. It would be good to have conditions that guarantee convergence or non-convergence. Where convergence does occur, is the limiting distribution the uniform distribution on $[0, 1]$? With resampling, we know that it cannot be, since, with repeated resampling, eventually a stage is reached in which only one element is resampled. We hope to clarify these and other points in future work.

Appendix

Proof of Theorem PP:

Define the function $a : [0, 1] \times \mathbb{M}_0 \times \mathbb{M}_0$ by the formula

$$a(x, b, \mu) = \int_0^x f(t, b, \mu)\, \mathrm{d}t,$$

where $f(t, b, \mu)$ is defined in Assumption 4. The CDF $R_0(\cdot, \mu)$ of the statistic τ under the DGP μ is then defined as follows:

$$R_0(x, \mu) = \int_{\mathbb{M}_0} a(x, b, \mu)\, m(\mathrm{d}b).$$

The CDF $R_1(\cdot, \mu)$ of the bootstrap P value $p_1(\mu, \omega) = R_0\big(\tau(\mu, \omega), \beta(\mu, \omega)\big)$ is defined thus:

$$\begin{aligned}
R_1(x, \mu) &= P\big(R_0(\tau(\mu, \omega), \beta(\mu, \omega)) \le x\big) \\
&= P\big(\tau(\mu, \omega) \le Q_0(x, \beta(\mu, \omega))\big) \\
&= \int_0^1 \int_{\mathbb{M}_0} f(t, b, \mu)\, \mathrm{I}\big(t \le Q_0(x, b)\big)\, \mathrm{d}t\, m(\mathrm{d}b) \\
&= \int_{\mathbb{M}_0} m(\mathrm{d}b) \int_0^{Q_0(x,b)} f(t, b, \mu)\, \mathrm{d}t \\
&= \int_{\mathbb{M}_0} a(Q_0(x, b), b, \mu)\, m(\mathrm{d}b). \qquad (43)
\end{aligned}$$

Now by the definition of ε, we have

$$|Q_0(x,\mu) - Q_0(x,b)| < \varepsilon,$$

where $b \equiv \beta(\mu,\omega)$. This inequality can be written as

$$Q_0(x,\mu) - \varepsilon < Q_0(x,b) < Q_0(x,\mu) + \varepsilon.$$

If we act on all three members of these inequalities by $a(\cdot,b,\mu)$ and integrate with respect to b, we get

$$\int_{\mathbb{M}_0} a(Q_0(x,\mu) - \varepsilon, b, \mu) < \int_{\mathbb{M}_0} a(Q_0(x,b), b, \mu) < \int_{\mathbb{M}_0} a(Q_0(x.\mu + \varepsilon), b, \mu). \tag{44}$$

By (43), the middle member here is $R_1(x,\mu)$, while the left- and right-most members are $R_0(Q_0(x,\mu) \pm \varepsilon, \mu)$. By the mean-value theorem,

$$R_0(Q_0(x,\mu) \pm \varepsilon, \mu) = R_0(Q_0(x,\mu),\mu) \pm \varepsilon R_0'(q_\pm,\mu) = x \pm \varepsilon R_0'(q_\pm,\mu),$$

where $q_\pm$ is a convex combination of $Q_0(x,\mu)$ and $Q_0(x,\mu) \pm \varepsilon$. By the definition of the quantity r in the statement of the theorem, $\varepsilon R_0'(q_\pm,\mu) < r\varepsilon$. Thus (44) says that

$$x - r\varepsilon < R_1(x,\mu) < x + r\varepsilon, \quad \text{or} \quad |R_1(x,\mu) - x| < r\varepsilon.$$

Consequently, for any $\mu,\nu \in \mathbb{M}_0$, $|R_1(x,\mu) - R_1(x,\nu)| < 2r\varepsilon$. It is easy to show that we also have $|Q_1(x,\mu) - Q_1(x,\nu)| < 2r\varepsilon$, so that $\bar{\lambda}_1 \le 2r\varepsilon$. Since the condition of the theorem says that $2r\varepsilon < \bar{\lambda}_0$, the theorem is proved. ■

Proof of Lemma FDB:

On letting $x = Q_0(x',\mu)$ in (16) for any $x' \in [0.1]$, we see that

$$|R_0(Q_0(x',\mu),\mu) - Q_0(x',\mu)| < \delta,$$

and, since $R_0(Q_0(x',\mu),\mu) = x'$, this becomes

$$|Q_0(x,\mu) - x| < \delta \quad \text{for } x \in [0,1] \text{ and } \mu \in \mathbb{M}_0. \tag{45}$$

Differentiating the identity $R_0(Q(x,\mu),\mu) = x$ with respect to x gives

$$R_0'(Q_0(x,\mu),x)Q_0'(x,\mu) = 1,$$

whence

$$|Q_0'(x,\mu) - 1| = \left| \frac{1}{R_0'(Q_0(x,\mu),\mu)} - 1 \right|. \tag{46}$$

Now, from the second inequality in (16), we have

$$\begin{aligned} 1 - \delta < R_0'(Q_0(x,\mu),\mu) < 1 + \delta, \text{ or} \\ \frac{1}{1+\delta} < \frac{1}{R_0'(Q_0(x,\mu),\mu)} < \frac{1}{1-\delta}, \text{ or} \\ -\frac{\delta}{1+\delta} < \frac{1}{R_0'(Q_0(x,\mu),\mu)} - 1 < \frac{\delta}{1-\delta}. \end{aligned} \tag{47}$$

Since by definition $\varepsilon = \delta/(1-\delta)$, (46) and (47) imply that $|Q_0'(x,\mu) - 1| < \varepsilon$. Then, as $\varepsilon > \delta$, we may conclude that all of the quantities $|R_0(x,\mu) - x|$, $|Q_0(x,\mu) - x|$, $|R_0'(x,\mu) - 1|$, and $|Q_0'(x,\mu) - 1|$ are less than ε.

From the definition (15) of R^1, we have

$$|R^1(x,\mu) - x| = |\mathrm{E}[R_0(x,\beta(\mu,\omega)) - x]| \le \mathrm{E}|R_0(x,\beta(\mu,\omega)) - x| < \delta.$$

By reasoning like that leading to (45), we find that $|Q^1(x,\mu) - x| < \delta$, and, similarly to (47), $|(R^1)'(x,\mu) - 1| < \varepsilon$ and $|(Q^1)'(x,\mu) - 1| < \varepsilon$.

By the mean-value theorem, we see that

$$R_0(Q^1(x,\mu),\beta(\mu,\omega)) - R_0(Q_0(x,\beta(\mu,\omega),\beta(\mu,\omega)) = R_0'(q,\beta(\mu,\omega))[Q^1(x,\mu) - Q_0(x,\beta(\mu,\omega))], \tag{48}$$

where q is a convex combination of $Q^1(x,\mu)$ and $Q_0(x,\beta(\mu,\omega))$. Now we may write

$$R_0'(q,\beta(\mu,\omega)) = 1 + \eta_1, \text{ with } |\eta_1| < \varepsilon, \text{ and}$$
$$Q^1(x,\mu) - Q_0(x,\beta(\mu,\omega)) = \eta_2, \text{ with } |\eta_2| < 2\varepsilon.$$

From the definition $R^1(Q^1(x,\mu),\mu) = x$, and the definition (16) of R^1, we have from (48) that

$$\begin{aligned} 0 &= \mathrm{E}[R_0(Q^1(x,\mu),\beta(\mu,\omega)) - x] \\ &= \mathrm{E}[R_0'(q,\beta(\mu,\omega))(Q^1(x,\mu) - Q_0(x,\beta(\mu,\omega))] = \mathrm{E}((1+\eta_1)\eta_2). \end{aligned}$$

Therefore,

$$\mathrm{E}(\eta_2) = \mathrm{E}[Q^1(x,\mu) - Q_0(x,\beta(\mu,\omega))] = -\mathrm{E}(\eta_1\eta_2), \text{ with } |\eta_1\eta_2| < 2\varepsilon^2. \tag{49}$$

Now, by use of the mean-value theorem once more, we see that

$$R_0(Q_0(x,\beta(\mu,\omega)),\mu) - R_0(Q^1(x,\mu),\mu) = R_0'(q_2,\mu)[Q_0(x,\beta(\mu,\omega)) - Q^1(x,\mu)],$$

where q_2 is another convex combination of $Q^1(x,\mu)$ and $Q_0(x,\beta(\mu,\omega))$. Hence

$$\mathrm{E}[R_0(Q_0(x,\beta(\mu,\omega)),\mu)] - R_0(Q^1(x,\mu),\mu) = \mathrm{E}[R_0'(q_2,\mu)[Q_0(x,\beta(\mu,\omega)) - Q^1(x,\mu)]].$$

If we write $R_0'(q_2,\mu) = 1 + \eta_3$, with $|\eta_3| < \varepsilon$, the above becomes

$$\left|\mathrm{E}[R_0(Q_0(x,\beta(\mu,\omega)),\mu)] - R_0(Q^1(x,\mu),\mu)\right| = \left|\mathrm{E}[Q_0(x,\beta(\mu,\omega)) - Q^1(x,\mu)] + \mathrm{E}(\eta_2\eta_3)\right| < 4\varepsilon^2.$$

■

Proof of Theorem FDB:

The result follows immediately from (13) and (17). ■

References

Beran, R. (1987). "Prepivoting to reduce level error of confidence sets", *Biometrika* **74**, 457-68.

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

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* **31**, 307–27.

Bühlmann, P., (1997). "Sieve bootstrap for time series", *Bernoulli* **3**, 123-48.

Chan, K. Y. F. and S. M. S. Lee (2001). "An exact iterated bootstrap algorithm for small-sample bias reduction", *Computational Statistics and Data Analysis* **36**, 1–13.

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

Chang, Y. and J. Y. Park, (2002). "A sieve bootstrap for the test of a unit root", *Journal of Time Series Analysis* **24**, 379-400.

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

Davidson, R. (2010). "Size distortion of bootstrap tests: Application to a unit root test", *Review of Economic Analysis* **2**, 169–193.

Davidson, R. (2015). "Computing, the bootstrap and economics", *Canadian Journal of Economics* **48**, 1195–1214.

Davidson, R. (2016a). "A Discrete Model for Bootstrap Iteration", *Journal of Econometrics*, forthcoming.

Davidson, R. (2016b). "Diagnostics for the Bootstrap and Fast Double Bootstrap", *Econometric Reviews*, forthcoming.

Davidson, R. and J. G. MacKinnon (2002). "Fast double bootstrap tests of nonnested linear regression models", *Econometric Reviews* **21**, 419-429.

Davidson, R. and J. G. MacKinnon (2006). "The power of asymptotic and bootstrap tests", *Journal of Econometrics* **133**, 421–441.

Davidson, R. and J. G. MacKinnon (2007). "Improving the reliability of bootstrap tests with the fast double bootstrap", *Computational Statistics and Data Analysis* **51**, 3259–3281.

Davidson, R. and J. G. MacKinnon (2010). "Wild Bootstrap Tests for IV Regression", *Journal of Business & Economic Statistics*, **28**, 128–144.

DiCiccio, T. J., M. A. Martin, and G. A. Young (1992). "Analytical approximations for iterated bootstrap confidence intervals", *Statistics and Computing* **2**, 161–71.

Dufour, J.-M., and L. Khalaf (2001). "Monte Carlo test methods in econometrics", Ch. 23 in *A Companion to Econometric Theory*, ed. B. Baltagi, Oxford, Blackwell Publishers, pp. 494–519.

Durbin, J. (1970). "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables", *Econometrica* **38**, 410–421.

Efron, B. (1979). "Bootstrap methods: Another look at the jackknife", *Annals of Statistics* **7**, 1–26.

Engle, R. F. (1982). "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica* **50**, 987–1007.

Giacomini, R., D. N. Politis, and H. White (2013). "A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators", *Econometric Theory* **29**, 567–89.

Godfrey, L. G. (1978). "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables", *Econometrica* **46**, 1293–1301.

Hall, P. (1986). "On the bootstrap and confidence intervals", *Annals of Statistics* **14**, 1431-52.

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

Hall, P. and M. A. Martin (1988). "On bootstrap resampling and iteration", *Biometrika* **75**, 661-71.

Ho, Y. H. S. and S. M. S. Lee (2005). "Iterated smoothed bootstrap confidence intervals for population quantiles", *Annals of Statistics* **33**, 437-62.

Horowitz, J. L. and N. E. Savin (2000). "Empirically relevant critical values for hypothesis tests", *Journal of Econometrics* **95**, 375389.

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

Lee, S. M. S. and G. A. Young (1995). "Asymptotic iterated bootstrap confidence intervals", *Annals of Statistics* **23**, 1301-30.

Lee, S. M. S. and G. A. Young (2003). "Prepivoting by weighted bootstrap iteration", *Biometrika* **90**, 393-410.

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. (2009). "Fast iterated bootstrap mean bias correction", working paper in the Proceedings of 2009 NZESG.

Palm, F. C., S. Smeekes, and J.-P. Urbain (2008). "Bootstrap unit root tests: comparison and extensions", *Journal of Time Series Analysis* **29**, 371-401.

Perron, P. and S. Ng (1996). "Useful modifications to unit root tests with dependent errors and their local asymptotic properties", *Review of Economic Studies* **63**, 435–65.

Psaradakis, Z. (2001). "Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors", *Journal of Time Series Analysis* **22**, 577-594.

Richard, P. (2009). "Modified fast double sieve bootstraps for ADF tests", *Computational Statistics & Data Analysis* **53**, 4490-4499.

Schilling, R. L. and L. Partzsch (2012). *Brownian Motion: An introduction to stochastic processes*, De Gruyter, Berlin/Boston.

Schwert, G. W. (1989). "Testing for unit roots: a Monte Carlo investigation", *Journal of Business and Economic Statistics* **7**, 147–59.

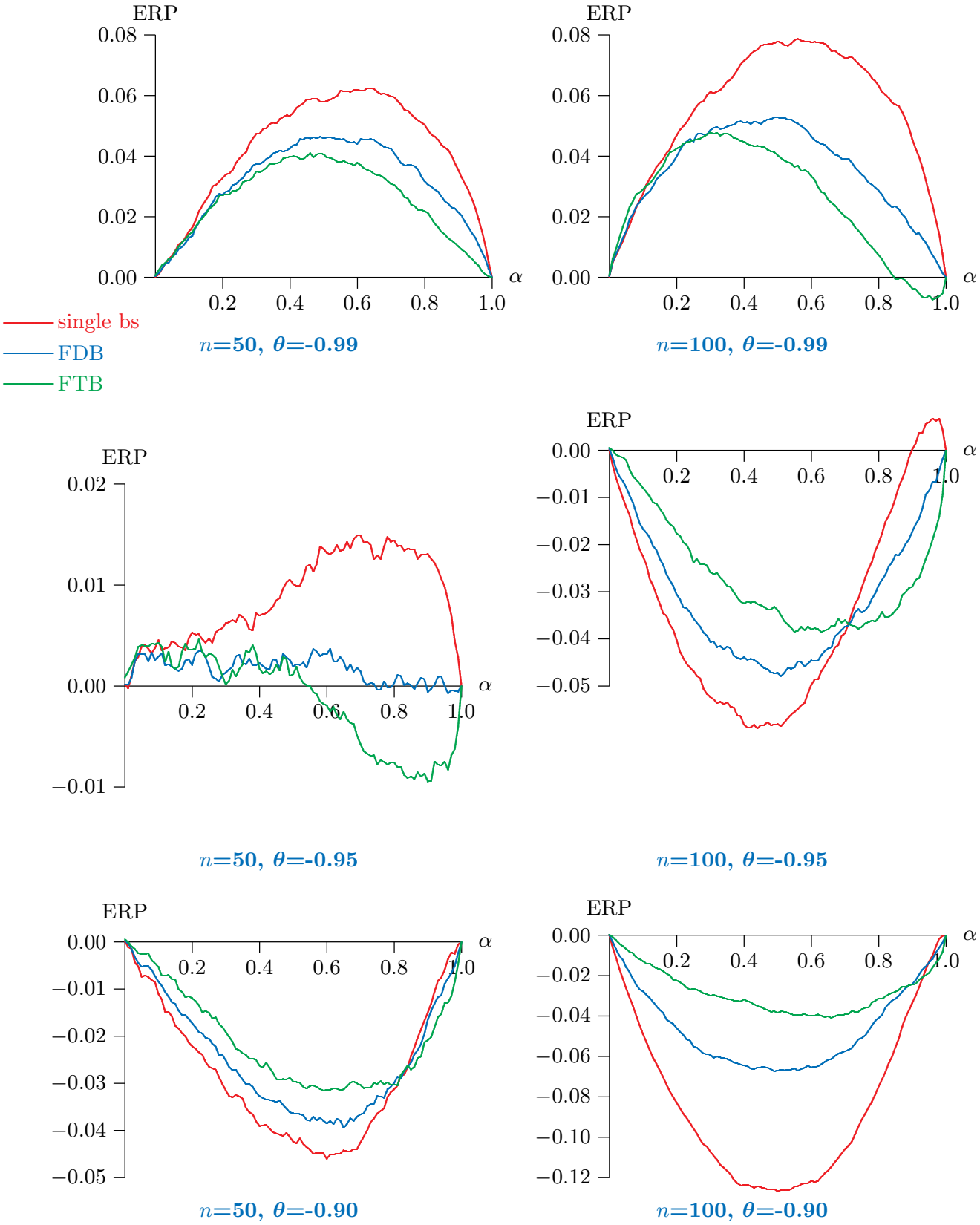
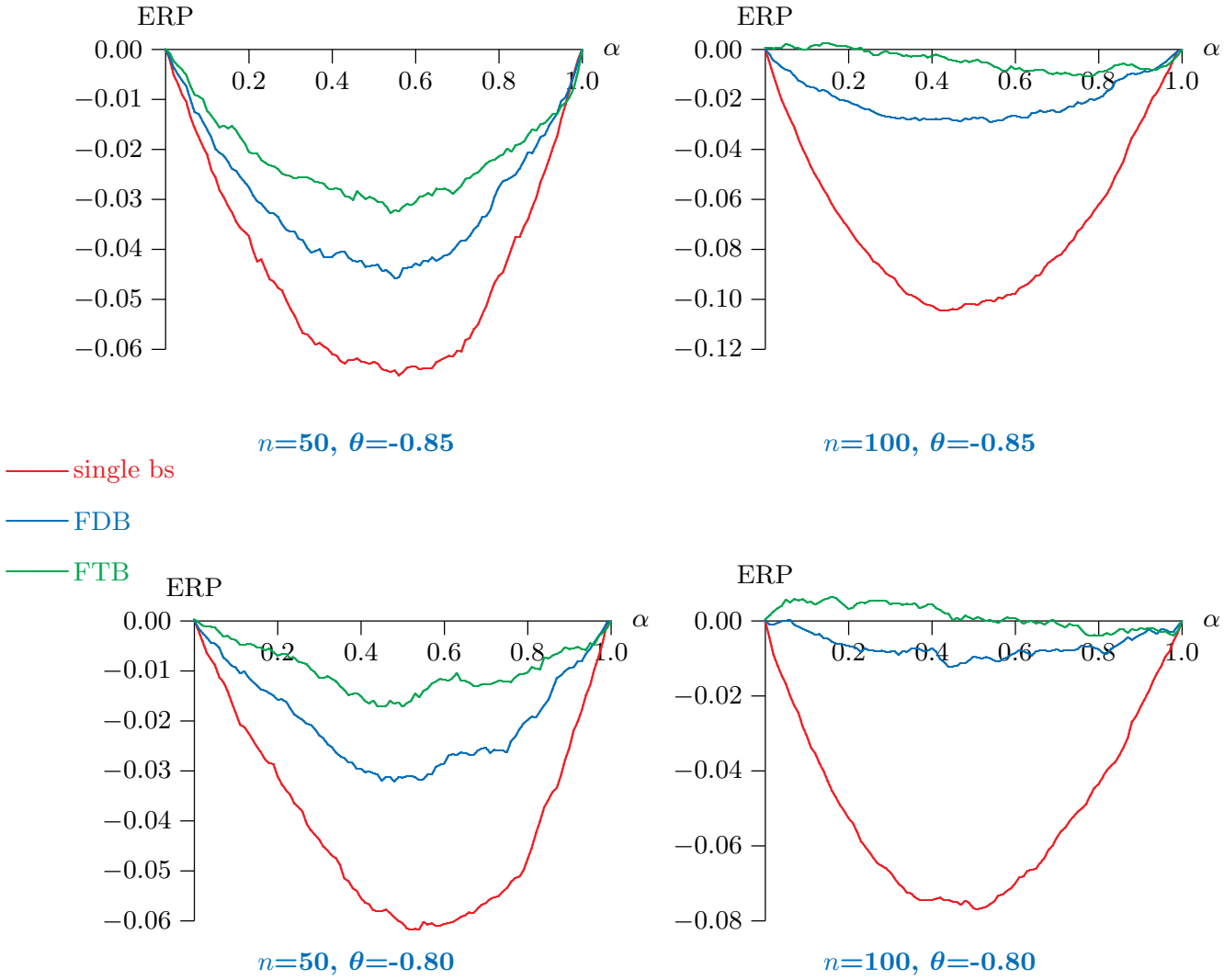


Figure 1: ERPs of unit root test



n=50, θ=-0.85 n=100, θ=-0.85

n=50, θ=-0.80 n=100, θ=-0.80

Figure 2: ERPs of unit root test

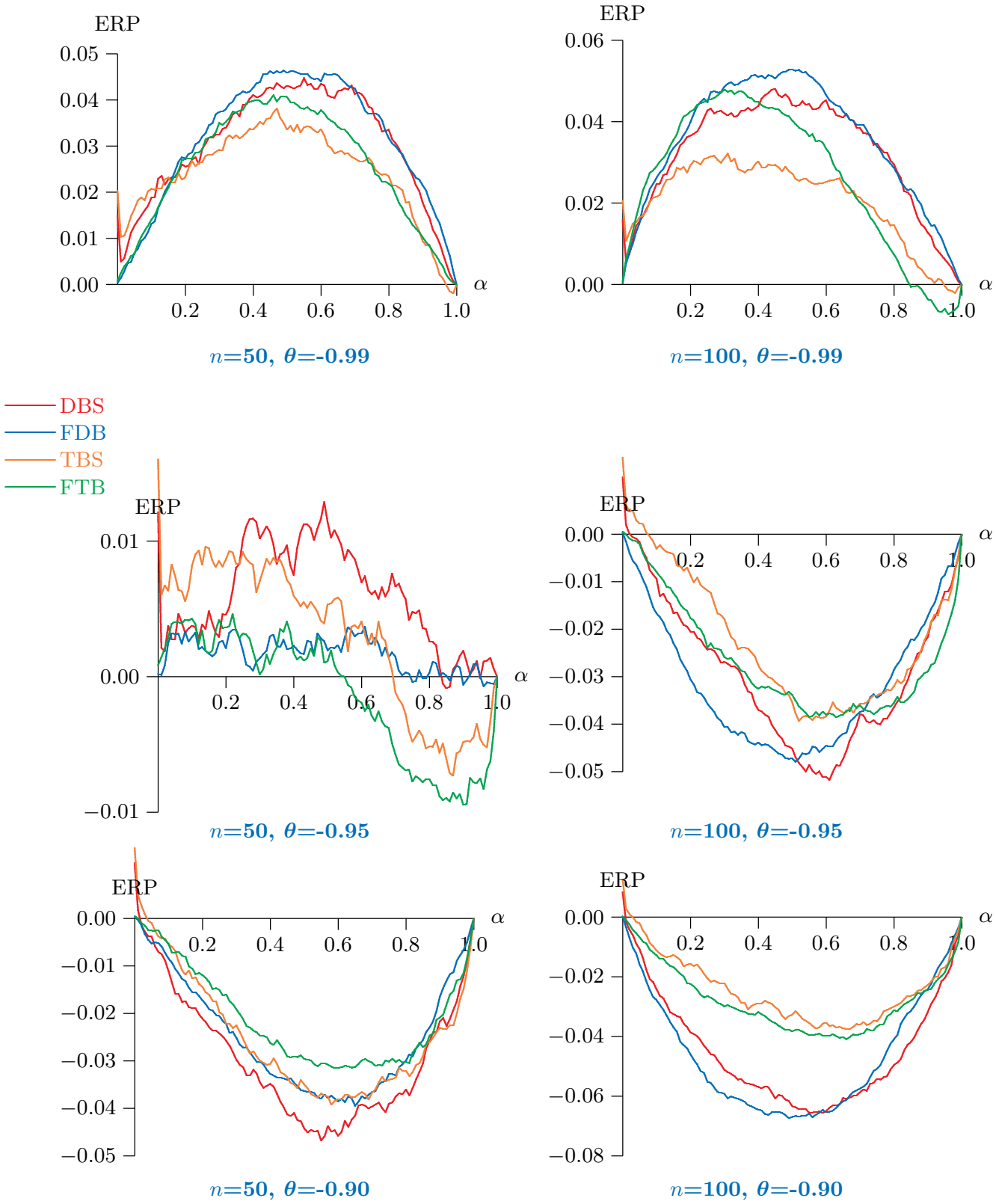
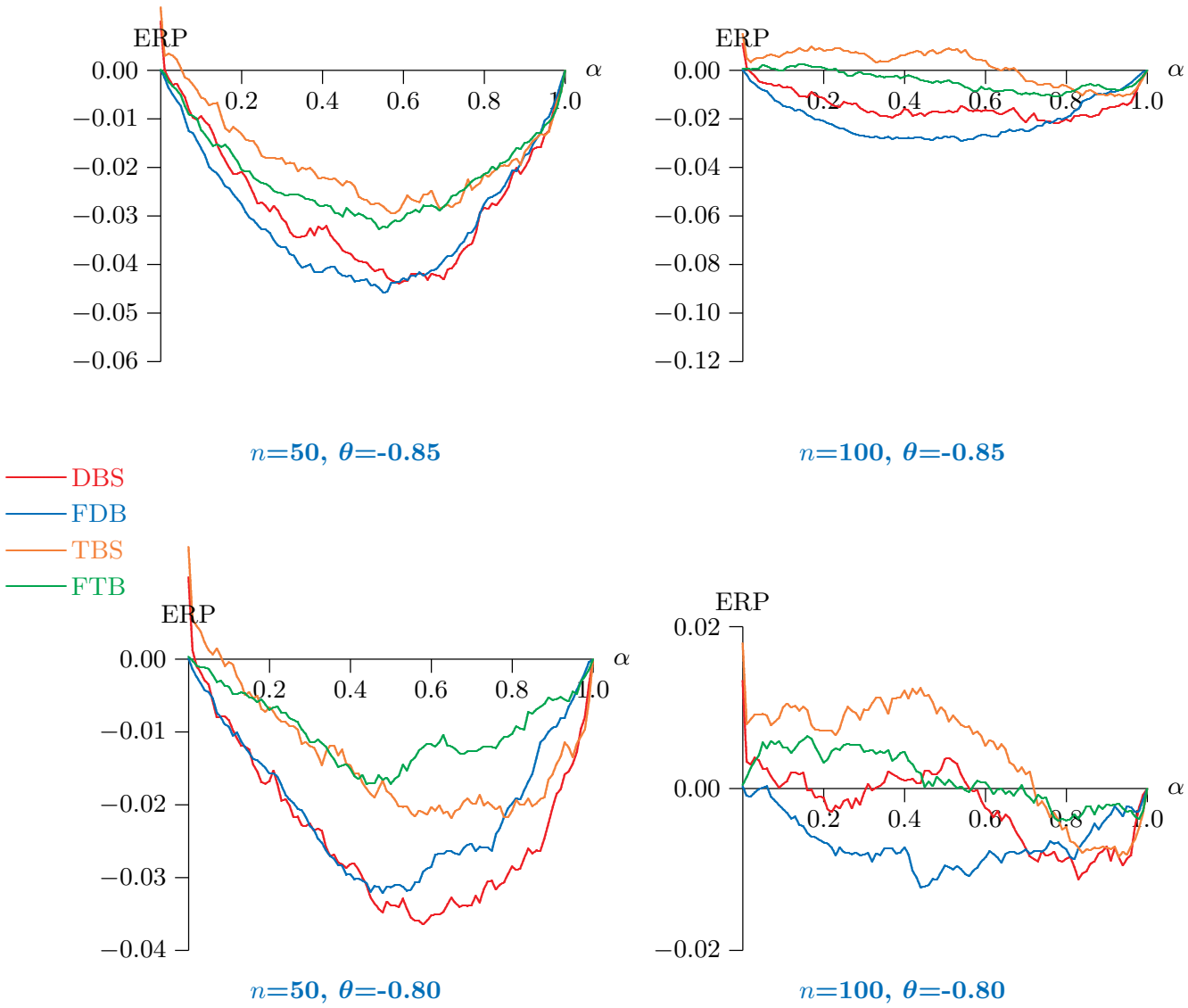


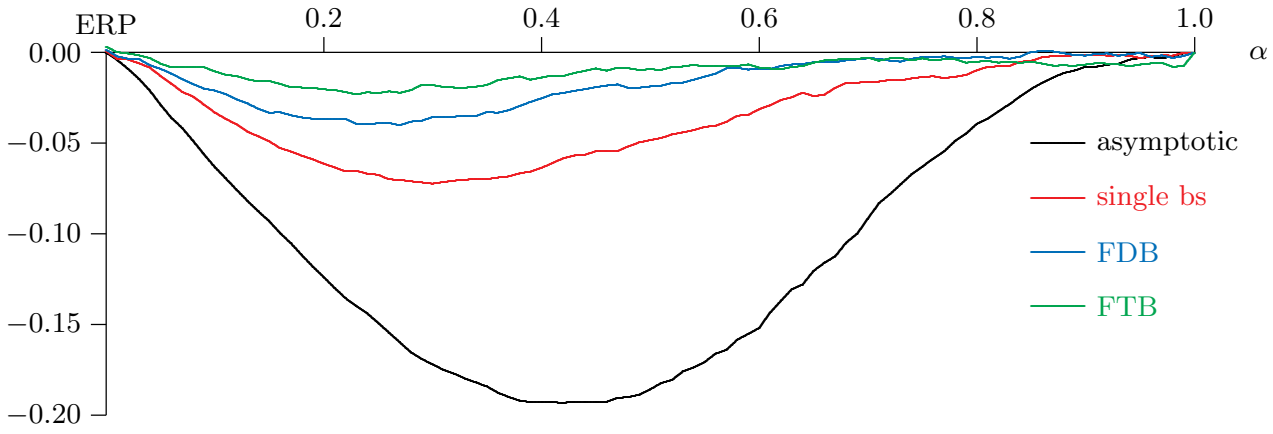
Figure 3: ERPs of unit root test



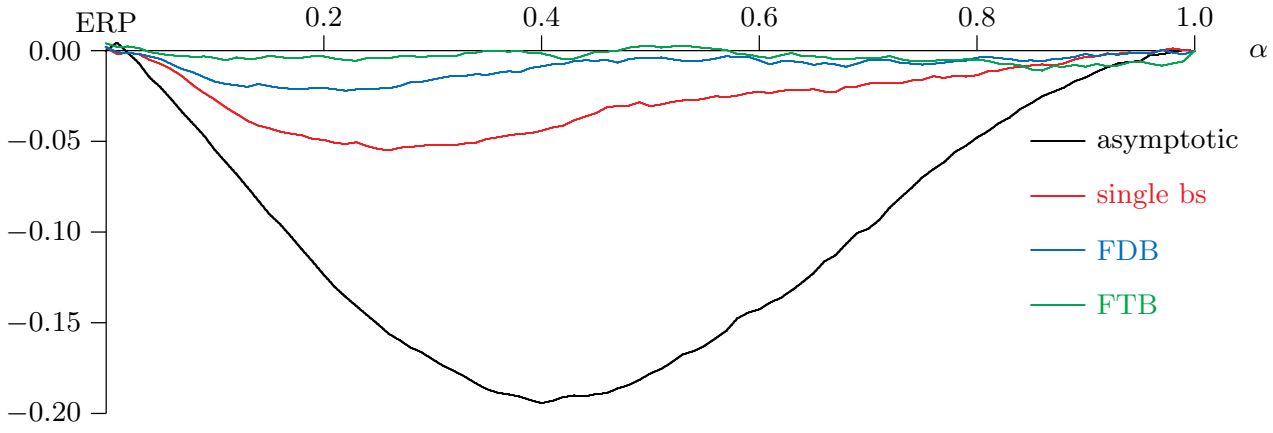
n=50, θ=-0.85 n=100, θ=-0.85

n=50, θ=-0.80 n=100, θ=-0.80

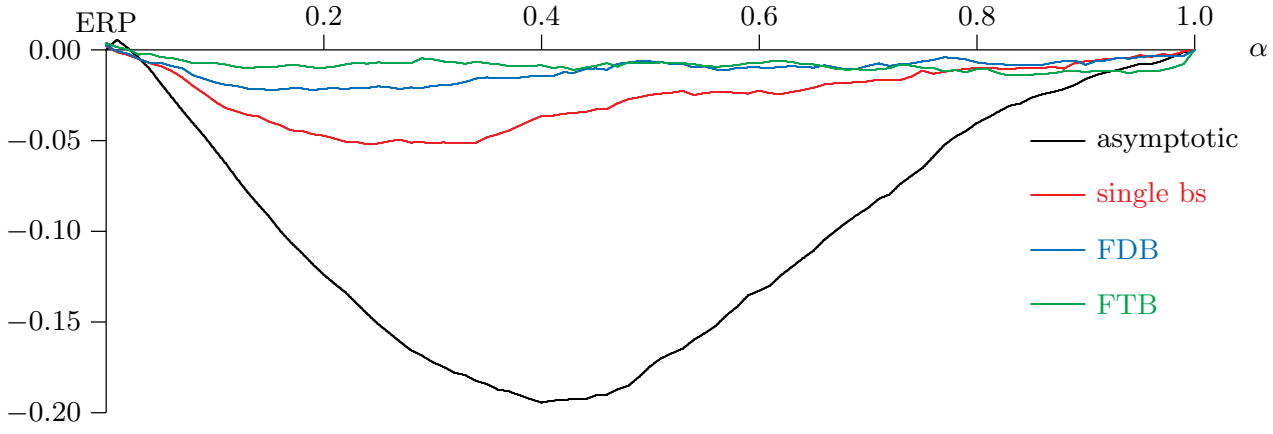
Figure 4: ERPs of unit root test



***n*=40**



***n*=80**



***n*=160**

Figure 5: ERPs of test for ARCH

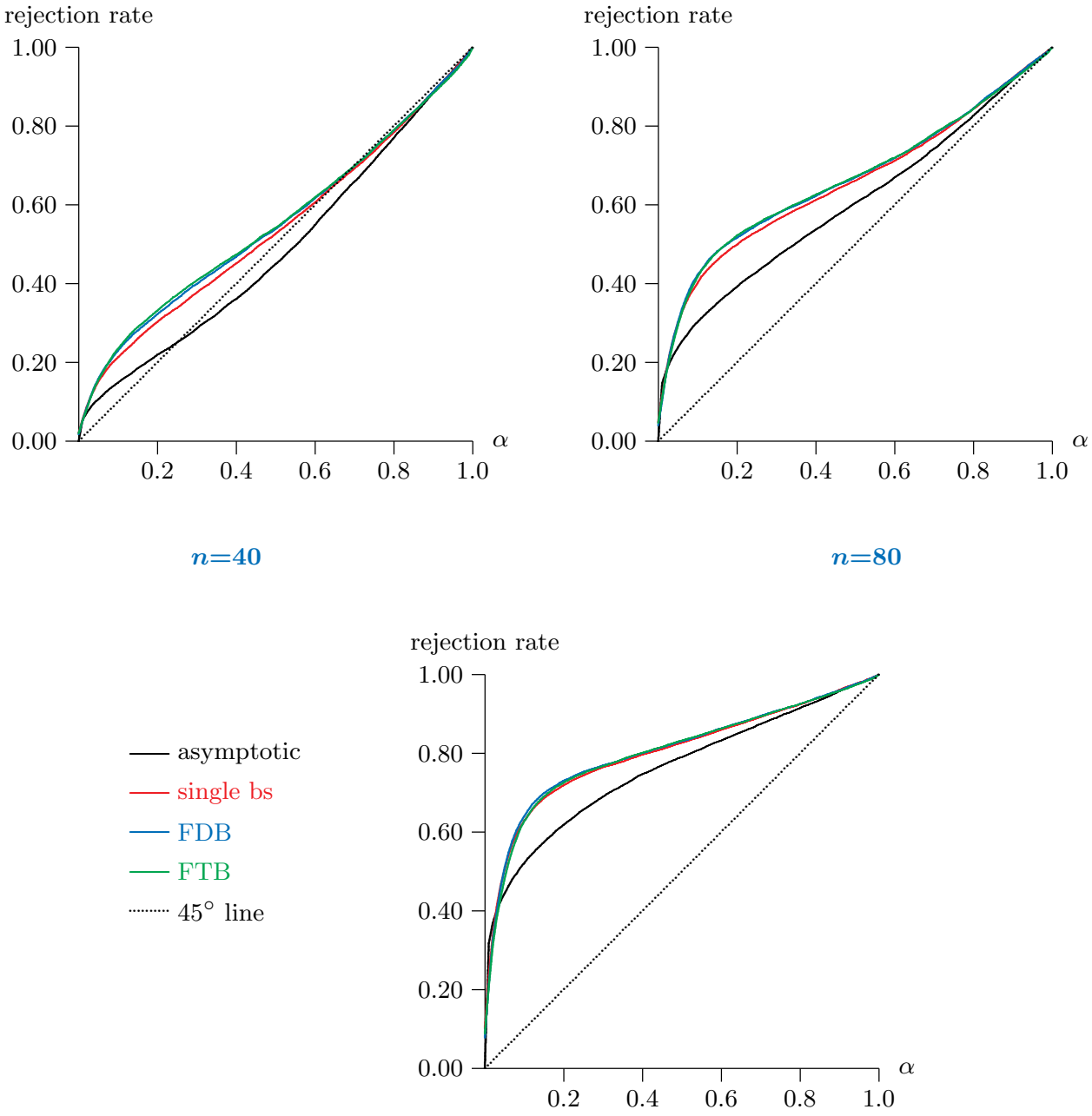
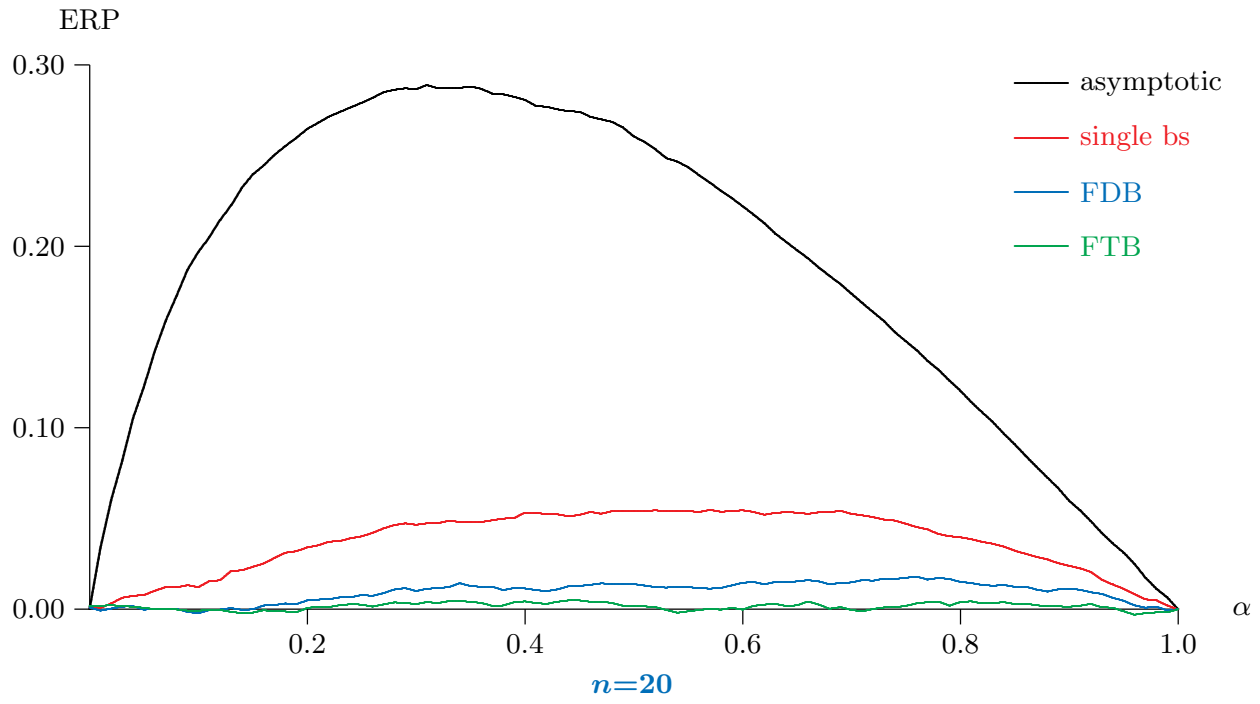
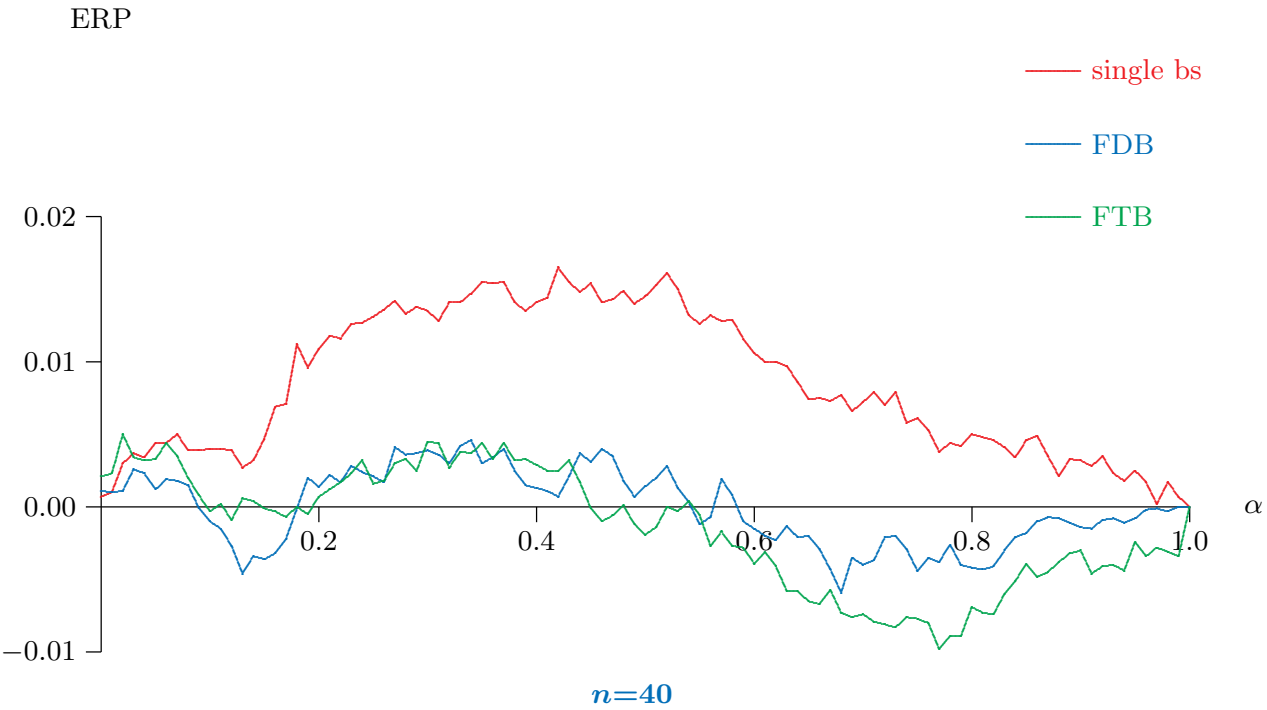


Figure 6: ARCH test rejection rate, α=1, γ=δ=0.3



***n*=20**



***n*=40**

Figure 7: ERPs of Durbin-Godfrey test

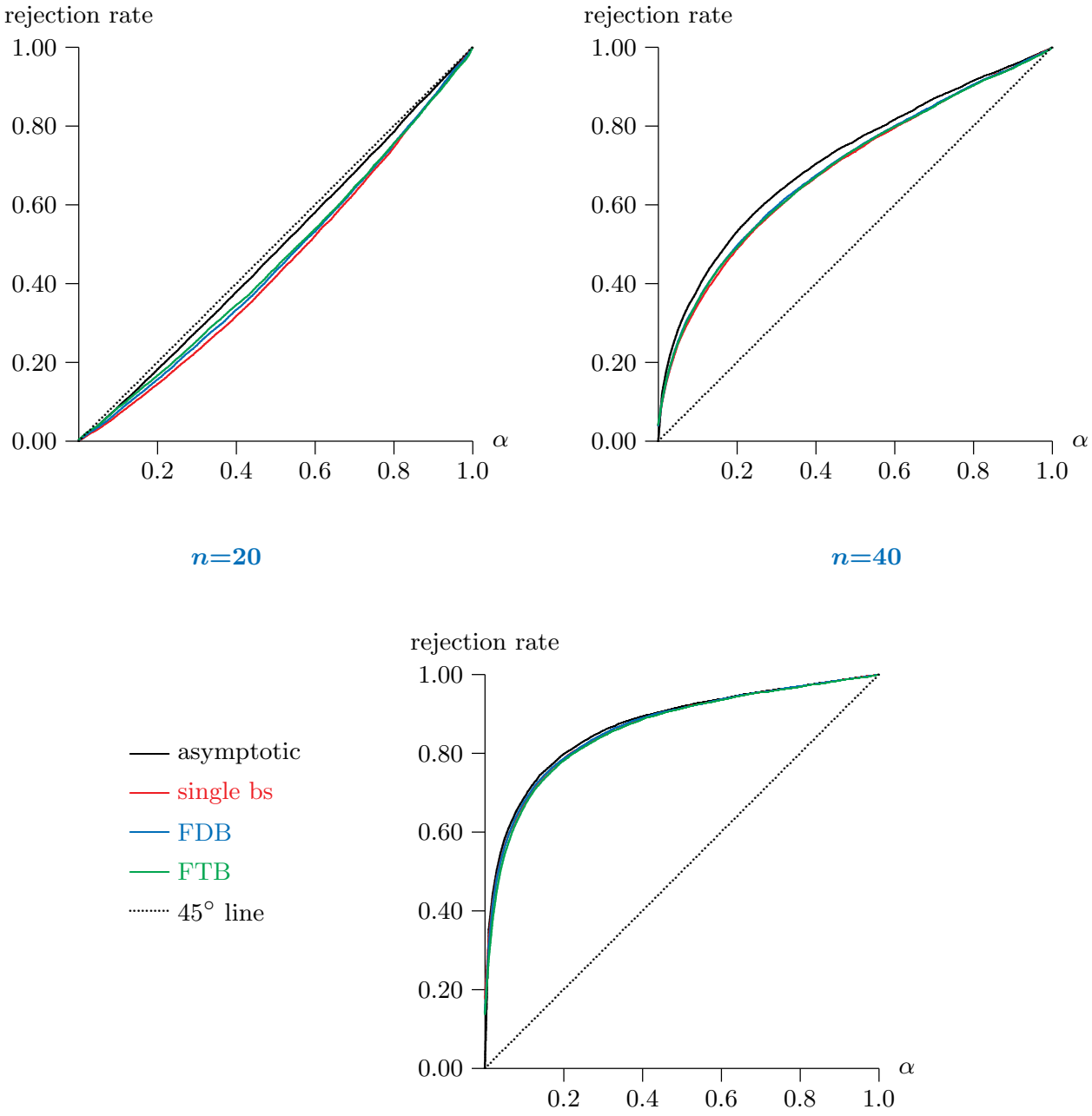


Figure 8: Durbin-Godfrey test rejection rate, ρ=0.5