cemmap
centre for microdata methods and practice

A discrete model for bootstrap iteration

Russell Davidson

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP38/15

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

A Discrete Model for Bootstrap Iteration

by

Russell Davidson

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

AMSE–GREQAM
Centre de la Vieille Charité
2 rue de la Charité
13236 Marseille cedex 02, France

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

Abstract

In an attempt to free bootstrap theory from the shackles of asymptotic considerations, this paper studies the possibility of justifying, or validating, the bootstrap, not by letting the sample size tend to infinity, but by considering the sequence of bootstrap P values obtained by iterating the bootstrap. The main idea of the paper is that, if this sequence converges to a random variable that follows the uniform $\mathrm{U}(0,1)$ distribution, then the bootstrap is valid. The idea is studied by making the model under test discrete and finite, so that it is characterised by a finite three-dimensional array of probabilities. This device, when available, renders bootstrap iteration to any desired order feasible. It is used for studying a unit-root test for a process driven by a stationary MA(1) process, where it is known that the unit-root test, even when bootstrapped, becomes quite unreliable when the MA(1) parameter is in the vicinity of -1. Iteration of the bootstrap P value to convergence achieves reliable inference except for a parameter value very close to -1. The paper then endeavours to see these specific results in a wider context, and tries to cast new light on where bootstrap theory may be going.

Keywords: Bootstrap, bootstrap iteration

JEL codes: C10, C12, C15

This research was supported by the Canada Research Chair program (Chair in Economics, McGill University), and by grants from the Fonds Québécois de Recherche sur la Société et la Culture. I am grateful for discussions with, and comments from, Andrew Chesher, James Davidson, and Joel Horowitz. The usual disclaimer applies.

May 2015

1. Introduction

The statistical method called the bootstrap relies in no way on any asymptotic considerations, but current bootstrap theory relies on them. I find this an extremely undesirable state of affairs, if for no other reason than that the choice of an asymptotic construction is inevitably somewhat arbitrary. This paper tries to take a first step towards remedying the sitution. Asymptotic arguments rely on sequences of random variables that converge, in probability or in distribution, to some desirable limit. Since bootstrap inference is exact only in rare instances, any justification of it must also rely on some sequence of random variables with a desirable limit.

The approach of this paper involves the convergence or otherwise of a sequence of bootstrap P values for testing the correct specification of an econometric (or statistical) model, obtained by iterating the bootstrap. Computing resources are nowadays such that the first iterate, called the double bootstrap, is in many cases feasible; when it is, it is presumed that it can provide more reliable statistical inference than either conventional asymptotics or the single uniterated bootstrap.

It is certainly tempting to suppose that the only barrier to still more reliable inference *via* bootstrap iteration is computational infeasibility. This paper attempts to show that bootstrap iteration can indeed improve reliability of inference. In order to do so, a procedure of discretisation is used, by means of which the model under test is described by a finite three-dimensional array of probabilities.

In the next Section, I introduce the definitions and notation needed to formulate the problem. The notation is somewhat unconventional, and makes use of the idea that a test statistic for a model, as also a bootstrap data-generating process (DGP), can be represented mathematically by a stochastic process, with index set the set of DGPs contained in the model. In Section 3, the discretisation procedure is laid out, and the mechanics of bootstrap iteration developed. The notion of the *bootstrap discrepancy* is briefly considered in Section 4, and some simple conclusions drawn. Then, in Section 5, a specific model is treated, and analysed numerically. The model under test is characterised by a unit-root process, obtained by cumulating a stationary MA(1) process, and the specification is tested by an augmented Dickey-Fuller (ADF) test. The model is completely parametric, with one single parameter, namely the MA(1) parameter. It is well known that, when this parameter is close to -1, the ADF test becomes thoroughly unreliable. A parametric bootstrap suggests itself as a way of improving reliability. The numerical study shows that it does so except for a parameter in a very small neighbourhood of -1, if the sequence of iterated P values is followed until (approximate) convergence. This occurs for the 44^{th} bootstrap iteration, an order of iteration completely inconceivable in normal circumstances. The results obtained for this example are discussed in a more general context in Section 6, and some concluding remarks are offered in Section 7.

2. Concepts and Notations

We treat a test statistic as a stochastic process with as index set the set of DGPs in a model $\mathbb{M}$. If we denote such a statistic by τ, then we have

$$\tau \ : \ \mathbb{M} \times \Omega \to \mathbb{R},$$

where $(\Omega, \mathcal{F}, P)$ is a suitable probability space, which can, for 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 $\omega \in \Omega$. For the moment, we suppose that, under any DGP we may consider, the distribution of τ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$. The statistic τ is used to test the hypothesis that the true DGP μ belongs to a model $\mathbb{M}_0$, which is a subset of $\mathbb{M}$.

For notational convenience, we suppose that the range of τ is the $[0, 1]$ interval rather than the whole real line, and that the statistic takes the form of an approximate P value, which thus leads to rejection when the statistic is too small. Let $R_0 : [0, 1] \to [0, 1]$ be the cumulative distribution function (CDF) of τ under any DGP $\mu \in \mathbb{M}$:

$$R_0(\alpha, \mu) = P\big\{\omega \in \Omega \,|\, \tau(\mu, \omega) \leq \alpha\big\}.$$

We also define the quantile function as the inverse of R_0 for given μ:

$$R_0\big(Q_0(\alpha, \mu), \mu\big) = \alpha = Q_0\big(R_0(\alpha, \mu), \mu\big).$$

For $\mu \in \mathbb{M}_0$, the random variable $R_0\big(\tau(\mu, \omega), \mu\big)$ is distributed as $\mathrm{U}(0, 1)$. This property is what allows Monte Carlo tests to give exact inference when τ is pivotal with respect to $\mathbb{M}_0$.

If τ is not pivotal, exact inference is no longer possible. 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, which we denote by b. Let a realisation of τ be denoted as t. Then the bootstrap statistic that follows the $\mathrm{U}(0, 1)$ distribution approximately is $R_0(t, b)$. We make the definition

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

where the ω_j^* are independent. Each ω_j^* can be thought of as a set of those random numbers needed to generate a realisation of the statistic. Then, as $B \to \infty$, $\hat{R}_0(\alpha, \mu)$ tends almost surely to $R_0(\alpha, \mu)$. Accordingly, we estimate the bootstrap statistic by $\hat{R}_0(t, b)$.

Just as t is $\tau(\mu, \omega)$ where neither the true DGP μ nor the realisation ω is observed, so also the bootstrap DGP b can be expressed as $\beta(\mu, \omega)$, for the same unobserved μ and ω as for t. We have

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

where, although the model $\mathbb{M}$ on the left-hand side may be a superset of $\mathbb{M}_0$, we insist that the $\mathbb{M}_0$ on the right-hand side is the null model under test (the first Golden Rule of bootstrapping; see, for instance, Davidson (2007)).

With this definition, the bootstrap statistic $R_0(t, b)$ is a realisation of a random variable $p_1(\mu, \omega)$, where the new function $p_1 : \mathbb{M} \times \Omega \to [0, 1]$ is defined as follows:

$$p_1(\mu, \omega) \equiv R_0\big(\tau(\mu, \omega), \beta(\mu, \omega)\big).$$

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 $\mathrm{U}(0, 1)$. But, as with $R_0\big(\tau(\mu, \omega), \mu\big)$, which is also distributed as $\mathrm{U}(0, 1)$, this fact is not enough to allow exact inference, because the actual μ that generates the data is unknown outside the context of a simulation experiment.

However, the bootstrap principle can again be applied, and the unknown μ replaced by the estimate $b = \beta(\mu, \omega)$, which is directly observed, or at least can be calculated from observed data. This leads to the double bootstrap, of which the P value, for realisations t and b, can be written as

$$\hat{R}_1\big(\hat{R}_0(t, b), b\big),$$

where $\hat{R}_1$ is defined analogously to (1) as

$$\hat{R}_1(\alpha, \mu) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{I}\big(p_1(\mu, \omega_i^*) < \alpha\big).$$

Under the assumption that the true μ belongs to the null model $\mathbb{M}_0$, this is an estimate of the probability mass in the distribution of the single bootstrap statistic to the left of the estimate $\hat{R}_0(t, b)$ of $R_0(t, b)$. As both B and N tend to infinity, the double bootstrap P value tends almost surely to $R_1\big(R_0(t, b), b\big)$. Expressed as a random variable, this limiting P value is

$$p_2(\mu, \omega) \equiv R_1\big(R_0(\tau(\mu, \omega), \beta(\mu, \omega)), \beta(\mu, \omega)\big)$$

If we write the right-hand side above as $R_1\big(p_1(\mu, \omega), \beta(\mu, \omega)\big)$, the analogy with $p_1(\mu, \omega)$ is complete. This demonstrates that the double bootstrap, by estimating the probability mass to the left of the single bootstrap P value, effectively bootstraps the single bootstrap P value.

From that observation, it is clear that we can define iterated bootstraps as follows. For $r = 0, 1, 2, \ldots$, we define

$$\begin{aligned} R_r(\alpha, \mu) &= P\big\{\omega \in \Omega \,|\, p_r(\mu, \omega) \leq \alpha\big\}, \\ p_{r+1}(\mu, \omega) &= R_r\big(p_r(\mu, \omega), \beta(\mu, \omega)\big), \end{aligned} \tag{2}$$

where we initialise the recurrence by the definition $p_0(\mu, \omega) = \tau(\mu, \omega)$. Thus $p_{r+1}(\mu, \omega)$ is the bootstrap P value obtained by bootstrapping the r^{th} order P value $p_r(\mu, \omega)$. It estimates the probability mass in the distribution of the r^{th} order P value to the left of its realisation.

In order for bootstrap iteration to be useful, it is necessary for the sequence $\{p_r(\mu, \omega)\}$ of iterated bootstrap P values to converge as $r \to \infty$, and for the limit of the sequence to be distributed as $\mathrm{U}(0, 1)$ for all $\mu \in \mathbb{M}_0$. Note that it is possible to have convergence to a distribution quite different from $\mathrm{U}(0, 1)$. For instance, with a resampling bootstrap, since with high probability each resample does not contain some of the observations of the original sample, repeated iteration leads to iterated bootstrap DGPs that have only one observation of the original sample, repeated as many times as the original sample has observations. This is one reason for our having limited attention until now to absolutely continuous distributions.

If inference based on iterated P values is to have the usual properties, then the sequences should converge, preferably almost surely, or at least in probability, to some random variable distributed as $\mathrm{U}(0, 1)$. Perhaps the limiting variable should be $R_0(\tau(\mu, \omega), \mu)$, which is indeed distributed as $\mathrm{U}(0, 1)$, but perhaps not.

This research project aims at sorting this out, and finding sufficient conditions for convergence. Necessary conditions would be even better, if we achieve sufficiency. It would then be possible to free the bootstrap from asymptotic theory in order to say what bootstrap procedures are valid and what not. The new criterion will be convergence of the iterated bootstrap P value to a known distribution.

3. Making Things Discrete

It's not obvious where to start in the study of convergence. But it is sometimes helpful, at least for intuition, if a discrete model is considered. This means that it is useless to aim for the continuous $\mathrm{U}(0, 1)$ distribution, but that need not be harmful. Various wild bootstraps exist, and although they are in many ways similar to resampling bootstraps, they do not suffer from the problem that the bootstrap DGP becomes degenerate if the bootstrap is iterated. The P value remains discrete, however. In the case of a wild bootstrap that assigns random signs to n residuals, there are at most 2^n possible bootstrap samples, and so at most 2^n possible bootstrap statistics. But if the bootstrap statistic is distributed uniformly over these 2^n values, that is quite enough for reliable inference, even although one cannot achieve exact inference for a test at any arbitrary level in $[0, 1]$.

In what follows, I assume that the statistic, in approximate P value form, can take on only the values i/n, $i = 0, 1, \ldots, n$, where n is not necessarily the sample size. Further, I assume that there are only m possible bootstrap DGPs. Thus I can let the outcome space Ω consist of just $m(n + 1)$ points, labelled by two coordinates (i, j), $i = 0, 1, \ldots, n$, $j = 1, \ldots, m$. Golden Rule 1 requires the bootstrap DGP to satisfy the null hypothesis, and so it makes sense to let the null contain exactly the m DGPs already considered. Under the null, then, μ is represented by k, say, with $k = 1, \ldots, m$.

We can then write

$$\tau\big(k,(i,j)\big) = i/n, \qquad \beta\big(k,(i,j)\big) = j,$$

where the DGP μ is represented by k, and the outcome ω by (i,j). The model is then completely characterised by the probabilities p_{kij}, $k,j = 1,\ldots,m$, $i = 0,1,\ldots,n$, where

$$p_{kij} = P\big[\tau\big(k,(i,j)\big) = i/n \text{ and } \beta\big(k,(i,j)\big) = j\big]. \tag{3}$$

We have, for all $k = 1,\ldots,m$, that

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

and we make the definitions

$$a_{kij} = \sum_{l=0}^{i-1} p_{klj} \qquad \text{and} \qquad A_{ki} = \sum_{j=1}^{m} a_{kij} \quad i = 0,\ldots,n+1. \tag{4}$$

Thus a_{kij} is the probability under DGP k that τ is less than i/n and that $b = j$, while A_{ki} is the marginal probability under k that $\tau < i/n$. Thus we may write

$$R_0(\alpha,k) = A_{k,\lfloor \alpha n\rfloor+1}.$$

Note that $a_{k0j} = A_{k0} = 0$ for all $k,j = 1,\ldots,m$. Further, $A_{k(n+1)} = 1$ and $a_{k(n+1)j}$ is the marginal probability under k that $\beta = j$, for all $k = 1,\ldots,m$.

While the probabilities p_{kij} depend on the DGP k, the *realisations* $\tau\big(k,(i,j)\big)$ and $\beta\big(k,(i,j)\big)$ do not. This is analogous to the fact in the continuous case that, once the statistic $\tau(\mu,\omega)$ and the bootstrap DGP $\beta(\mu,\omega)$ are realised as t and b respectively, everything that one can compute, all iterated bootstrap P values or bootstrap DGPs, can be expressed as functions of t and b alone, without reference to the DGP μ that generated them.

Consider next the bootstrap P value under DGP k with realisation (i,j). It is the probability mass under the bootstrap DGP j of a value of τ less than i. It follows that

$$p_1\big(k,(i,j)\big) = A_{ji}, \tag{5}$$

where k does not appear on the right-hand side. Analogously, in the continuous case, the realised bootstrap P value is $R_0(t,b)$, without mention of the underlying μ.

In order to iterate the bootstrap, we need the function $R_1(\alpha,\mu)$, the CDF of p_1 under the DGP μ. In the discrete case, the function can be written as

$$R_1(\alpha,k) = P\big[p_1\big(k,(i,j)\big) < \alpha\big] = \sum_{i=0}^{n}\sum_{j=1}^{m} p_{kij}\mathrm{I}(A_{ji} < \alpha), \tag{6}$$

where I$(\cdot)$ is an indicator function. Now let $q_j^1(\alpha)$ be defined by

$$q_j^1(\alpha) = \max_{i=0,\ldots,n} \big\{ i \;\big|\; A_{ji} \le \alpha \big\}$$

In this discrete context, $q_j^1(\alpha)/n$ can be interpreted as an α-quantile of the distribution of the statistic τ with DGP j. In fact, for $\alpha = A_{ji}$, it is easy to see that $q_j^1(A_{ji}) = i$. Since the a_{jil} are increasing in i, so too are the A_{ji}, and we see that

$$A_{ji} \le \alpha \quad \text{for all } i \le q_j^1(\alpha) \text{ and } \quad A_{ji} > \alpha \quad \text{for all } i > q_j^1(\alpha).$$

Thus the event $\{A_{ji} < \alpha\}$ is equivalent to the inequality $i < q_j^1(\alpha)$.

This allows us to compute $R_1(\alpha, k)$. We find from (6) that

$$R_1(\alpha, k) = \sum_{j=1}^{m} \sum_{i=0}^{q_j^1(\alpha)-1} p_{kij} = \sum_{j=1}^{m} a_{kq_j^1(\alpha)j}. \tag{7}$$

We may think of the indices k (a DGP), i (a statistic), and j (a DGP) as forming a three-dimensional array, with k varying in the horizontal direction and i in the vertical direction of a flat two-dimensional surface, and j varying perpendicular to that surface. Fixing k in $R_1(\alpha, k)$ puts us on a sheet defined by the vertical i-direction and the perpendicular j-direction. The sum in the right-hand expression above takes us on a trip across this sheet, where, as we vary j, the i-coordinate traces out the values $q_j^1(\alpha)$.

The right-hand side of (7) can be seen to correspond precisely to the usual definition of R_1 in the continuous case. Since $a_{kq_j^1(\alpha)j}$ is the probability under k that j is the bootstrap DGP and that $\tau < q_j^1(\alpha)/n$, and since $q_j^1(\alpha)/n$ corresponds to $Q_0(\alpha, j)$, (7) says that

$$\begin{aligned} R_1(\alpha, k) &= \sum_j \Pr{}_k\big(\beta = j \text{ and } \tau < Q_0(\alpha, j)\big) \\ &= P\big[\tau(k,\omega) < Q_0(\alpha, \beta(k,\omega)\big] \\ &= P\big[R_0\big(\tau(k,\omega), \beta(k,\omega)\big) < \alpha\big] = P\big[p_1(k,\omega) < \alpha\big], \end{aligned}$$

which is indeed the probability that the bootstrap P value is less than α.

Next, we wish to compute the double bootstrap P value $p_2(\mu, \omega)$, or, in the notation of the discrete problem, $p_2\big(k, (i,j)\big)$. Now $p_2(\mu,\omega) = R_1\big(p_1(\mu,\omega), \beta(\mu,\omega)\big)$ in the general notation, and this last equation translates to

$$p_2\big(k, (i,j)\big) = R_1\big(A_{ji}, j\big) = \sum_{l=1}^{m} a_{jq_l^1(A_{ji})l} \tag{8}$$

where the first equality follows because $p_1\big(k, (i,j)\big) = A_{ji}$, and $\beta\big(k, (i,j)\big) = j$, and the second follows from the definition (7). Make the definition

$$A_{ji}^2 = \sum_{l=1}^{m} a_{jq_l^1(A_{ji})l}, \tag{9}$$

where the superscript 2 is an index, not an exponent. Then the double bootstrap P value $p_2\big(k,(i,j)\big) = A^2_{ji}$; compare with the relation (5).

In fact, the definition (9) corresponds to the definition of the double bootstrap P value in the continuous case. To see this, note that (9) tells us that

$$A^2_{ji} = \sum_{l=1}^{m} \Pr{}_j\big(\beta = l \text{ and } \tau < q^1_l(A_{ji})\big), \tag{10}$$

where j in the discrete case plays the role of μ in the continuous case. Then, since $q^1_l(\alpha)$ is $Q_0(\alpha, l)$, and A_{ji}, the bootstrap P value, can be expressed as $R_0(i/n, j)$, it follows that $q^1_l(A_{ji}) = Q_0\big(R_0(i/n, j), l\big)$. Then (10) becomes

$$\begin{aligned} A^2_{ji} &= P\big[\tau(j,\omega) < Q_0\big(R_0(i/n,j), \beta(j,\omega)\big)\big] \\ &= P\big[R_0\big(\tau(j,\omega), \beta(j,\omega)\big) < R_0(i/n, j)\big] \\ &= P\big[p_1(j,\omega) < R_0(i/n,j)\big] = R_1\big(R_0(i/n,j), j\big) \end{aligned} \tag{11}$$

Now i/n corresponds to $\tau(\mu,\omega)$, and j corresponds to $\beta(\mu,\omega)$, so that, expressed as a random variable, A^2_{ji} is written as

$$R_1\big(R_0(\tau(\mu,\omega),\beta(\mu,\omega)),\beta(\mu,\omega)\big) = R_1\big(p_1(\mu,\omega),\beta(\mu,\omega)\big),$$

as we wished to show.

The CDF of the random variable p_2 under DGP k is denoted by $R_2(\alpha, k)$, and it is defined as follows:

$$R_2(\alpha,k) = \sum_{i=0}^{n}\sum_{j=1}^{m} p_{kij}\mathrm{I}\Big(p_2\big(k,(i,j)\big) < \alpha\Big) = \sum_{i=0}^{n}\sum_{j=1}^{m} p_{kij}\mathrm{I}(A^2_{ji} < \alpha).$$

From their definition, the functions $q^1_j(\alpha)$ are increasing in α, and we saw that the A_{ji} are increasing in i. Consequently the A^2_{ji} are increasing in i. Now we can make the definition

$$q^2_j(\alpha) = \max_{i=0,\dots,n}\{i \mid A^2_{ji} \le \alpha\}. \tag{12}$$

Clearly, the q^2_j are increasing in α, and then

$$R_2(\alpha,k) = \sum_{j=1}^{m}\sum_{i=0}^{q^2_j(\alpha)-1} p_{kij} = \sum_{j=1}^{m} a_{kq^2_j(\alpha)j}. \tag{13}$$

This again takes us on a trip across the plane defined by k, where the i-coordinate for given j is now $q^2_j(\alpha)$.

Although $q_j^1(\alpha)/n$ is an α-quantile of the statistic τ, it is not the case that $q_j^2(\alpha)/n$ can be interpreted as an α-quantile. We saw in (11) that $A_{ji}^2 = R_1\big(R_0(i/n, j), j\big)$. The definition (12) implies that $q_j^2(\alpha)$ is the value of i such that $R_1\big(R_0(i/n, j), j\big) = \alpha$. Thus

$$q_j^2(\alpha)/n = Q_0\big(Q_1(\alpha, j), j\big), \tag{14}$$

where Q_1 is the quantile function inverse to R_1. Note that $q_j^2(A_{ji}^2) = i$.

In order to grasp the pattern of the iterations, we now look at the triple bootstrap P value, p_3. From (13), we have

$$p_3\big(k, (i, j)\big) = R_2\Big(p_2\big(k, (i, j)\big), j\Big) = R_2(A_{ji}^2, j) = \sum_{l=1}^{m} a_{jq_l^2(A_{ji}^2)l}.$$

We make the definitions

$$A_{ji}^3 = \sum_{l=1}^{m} a_{jq_l^2(A_{ji}^2)l}, \tag{15}$$

and

$$q_j^3(\alpha) = \max_{i=0,\ldots,n} \{i \mid A_{ji}^3 \le \alpha\}, \tag{16}$$

so that A_{ji}^3 is the realisation of the triple bootstrap P value for $\omega = (i, j)$. The distribution R_3 of the triple bootstrap P value is therefore

$$R_3(\alpha, k) = \sum_{i=0}^{n}\sum_{j=1}^{m} p_{kij}\mathrm{I}(A_{ji}^3 < \alpha) = \sum_{j=1}^{m}\sum_{i=0}^{q_j^3(\alpha)-1} p_{kij} = \sum_{j=1}^{m} a_{kq_j^3(\alpha)j},$$

which invites us to take yet another trip across the plane defined by k.

To see how this compares with the continuous case, we can observe that (15) says that

$$A_{ji}^3 = \sum_{l=1}^{m} \Pr{}_j\big(\beta = l \text{ and } \tau < q_l^2(A_{ji}^2)/n\big).$$

Now from (14) we know that $q_l^2(A_{ji}^2)/n$ can be written as $Q_0\big(Q_1(A_{ji}^2, l), l\big)$, while from (11) $A_{ji}^2 = R_1\big(R_0(i/n, j), j\big)$. It follows that

$$q_l^2(A_{ji}^2)/n = Q_0\Big(Q_1\big(R_1\big(R_0(i/n, j), j\big), l\big), l\Big).$$

The inequality $\tau < q_l^2(A_{ji}^2)/n$ is therefore equivalent to

$$R_0(\tau, l) < Q_1\Big(R_1\big(R_0(i/n, j), j\big), l\Big) \quad \text{or} \quad R_1\big(R_0(\tau, l), l\big) < R_1\big(R_0(i/n, j), j\big),$$

and so

$$\begin{aligned} A^3_{ji} &= P\big[R_1\big(R_0(\tau(j,\omega),\beta(j,\omega),\beta(j,\omega)\big) < R_1\big(R_0(i/n,j),j\big)\big] \\ &= P\big[p_2(j,\omega) < R_1\big(R_0(i/n,j),j\big)\big] \\ &= R_2\Big(R_1\big(R_0(i/n,j),j\big),j\Big), \end{aligned} \tag{17}$$

which corresponds to $R_2\big(R_1(R_0(\tau(\mu,\omega),\beta(\mu,\omega)),\beta(\mu,\omega)),\beta(\mu,\omega)\big)$ in the continuous case.

From (16) and (17), we can see that $q^3_j(\alpha)$ is the value of i such that $R_2\big(R_1(R_0(i/n,j),j),j\big)$ is equal to α. Thus

$$q^3_j(\alpha)/n = Q_0\Big(Q_1\big(Q_2(\alpha,j),j\big),j\Big),$$

where Q_3 is inverse to R_3.

The pattern is now clear. If we initialise the recurrence that defines bootstrap iteration as follows:

$$p_0\big(k,(i,j)\big) = A^0_{ji} = i/n; \quad q^0_j(\alpha) = \lfloor n\alpha \rfloor \text{ independent of } j,$$

then at each step r of the iteration, we have

$$\begin{aligned} p_r\big(k,(i,j)\big) &= A^r_{ji}, \\ q^r_k(\alpha) &= \max_{i=1,\ldots,n}\{i \mid A^r_{ki} \le \alpha\}, \\ q^r_k(A^r_{ki}) &= i, \text{ and} \\ R_r(\alpha,k) &= \sum_{j=1}^m a_{kq^r_j(\alpha)j}. \end{aligned}$$

The recurrence is then implemented by a trip across the plane defined by k:

$$A^{r+1}_{ki} = \sum_{j=1}^m a_{kq^r_j(A^r_{ki})j} = R_r(A^r_{ki},k). \tag{18}$$

It is easy enough to show that these definitions correspond to those of the continuous case, given in (2).

Given the model, that is, all the p_{kij}, this defines a mapping of the $m \times n$ matrix $\boldsymbol{A}^r$, with typical element A^r_{ki}, into $\boldsymbol{A}^{r+1}$. Since $A^r_{k0} = 0$, we do not need to include these as elements of $\boldsymbol{A}^r$. Convergence of the bootstrap iterations is thus convergence of iterates of this (very nonlinear) mapping. Although (18) appears to defined a separate mapping for each element ki of the matrix, the mappings of individual elements are interdependent, since A^{r+1}_{ki} depends on the $a_{kq^r_j(A_ki)j}$ for all $j = 1,\ldots,m$.

The present setup allows for almost anything. What we need is to find some conditions like being close to pivotal and/or having the bootstrap DGP "close" to the actual DGP that would imply convergence.

4. Bootstrap Discrepancies

The bootstrap discrepancy is defined, for a given DGP μ, and a given significance level α, as the difference between the rejection probability of the bootstrap test for DGP μ and level α and α itself. It is therefore equal to $R_1(\alpha, \mu) - \alpha$. For iterated bootstraps, the definition is the same: at level r, the order-r discrepancy is $R_r(\alpha, \mu) - \alpha$. The aim in bootstrapping, and in iterated bootstrapping, is to minimise the bootstrap discrepancy.

There are two trivial cases in which the discrepancy is zero, and the iterated bootstrap P values coincide with the single bootstrap P value. The first case arises when the statistic τ is a pivot. In the discrete context then, the A_{ki} do not depend on k; recall that A_{ki} is the probability under k that $\tau < i/n$. We may express this by writing $A_{ki} = A_{\cdot i}$. This means that $q^1_j(\alpha)$ depends only on α, not on j; we write $q^1_j(\alpha) = q^1_\cdot(\alpha)$. We see that, in this case,

$$R_1(\alpha, k) = \sum_{j=1}^{m} a_{kq^1_\cdot(\alpha)j} = A_{\cdot q^1_\cdot(\alpha)}, \tag{19}$$

and this does not depend on k. Therefore the bootstrap P value is also a pivot, with the same distribution whatever the DGP k, and the probability that it is less than α is the probability that $\tau < q^1_\cdot(\alpha)/n$, which would be exactly equal to α without the discreteness of the problem, and is indeed equal to α for $\alpha = A_{\cdot i}$, since $q^1_\cdot(A_{\cdot i}) = i$, and the probability that $\tau < i/n$ is $A_{\cdot i} = \alpha$. For these values of α, therefore, the bootstrap discrepancy is zero.

For realisation (i, j), the bootstrap P value is $A_{\cdot i}$. The double bootstrap P value is, from (8), $A^2_{ji} = R_1(A_{\cdot i}, \cdot)$, and, from (19), this is $A_{\cdot q^1_\cdot(A_{\cdot i})}$. Now

$$q^1_\cdot(A_{\cdot i}) = \max_j \{ j \mid A_{\cdot j} \leq A_{\cdot i} \},$$

and the value of the right-hand side is clearly just i. Thus the double bootstrap P value is $A_{\cdot i}$, the same as the single bootstrap P value, and also therefore independent of j.

The other special case arises when the bootstrap DGP always coincides with the true DGP. If so, then, for all admissible k, i, and j, we have $a_{kij} = A_{ki}\delta_{kj}$. The factor of A_{ki} on the right-hand side of this equation is justified, because

$$A_{ki} \equiv \sum_{j=1}^{m} a_{kij} = A_{ki} \sum_{j=1}^{m} \delta_{kj} = A_{ki}.$$

From (7) we have

$$R_1(\alpha, k) = \sum_{j=1}^{m} a_{kq^1_j(\alpha)j} = \sum_{j=1}^{m} \delta_{kj} A_{kq^1_j(\alpha)} = A_{kq^1_k(\alpha)}.$$

In particular, for $\alpha = A_{ki}$, $i = 0, \ldots, n$, since $q_k^1(A_{ki}) = i$, $R_1(A_{ki}, k) = A_{ki}$, which implies that for these values of α, the bootstrap discrepancy is zero. The double bootstrap P value is A_{ki}^2, which is

$$A_{ki}^2 = \sum_{j=1}^{m} a_{kq_j^1(A_{ki})j} = A_{kq_k^1(A_{ki})} = A_{ki}.$$

Thus, as in the case of a pivotal statistic, so too here the bootstrap discrepancy is zero for $\alpha = A_{ki}$, and the double bootstrap P value is identical to the single bootstrap P value, as are all higher-order P values.

Suppose that the recurrence (18) converges, in the sense that, for given k and i,

$$q_j^{r+1}(A_{ki}^{r+1}) = q_j^r(A_{ki}^r) \quad \text{for all } j = 1, \ldots, m. \tag{20}$$

Then $A_{ki}^{r+s} = A_{ki}^r$ and $q_j^{r+s}(A_{ki}^{r+s}) = q_j^r(A_{ki}^r)$ for all j and for all positive integers s. Then, since $R_r(A_{ki}^r, k) = A_{ki}^{r+1} = A_{ki}^r$, it follows that the bootstrap discrepancy is zero for significance level A_{ki}^r if k is the DGP and $\tau = i/n$. If (20) holds for all $i = 0, 1, \ldots, n$, the bootstrap discrepancy is zero unconditionally for DGP k for levels A_{ki}^r.

The discrepancy can also be defined for the raw statistic τ. If it is assumed that it is in approximate P value form, then the discrepancy at nominal level α is the probability that $\tau < \alpha$ minus α. This discrepancy, and the bootstrap discrepancy at all orders, for all the DGPs in the discrete model, can be found using the elements of the matrices $\boldsymbol{A}^r$, $r = 0, 1, 2, \ldots$. For the raw discrepancy, A_{ki} is the probability under DGP k that $\tau < i/n$. The information contained in these numbers can be best examined graphically, with a P value plot or a P value discrepancy plot; see Davidson and MacKinnon (1998). The former is the locus of the points $(i/n, A_{ki})$, $i = 0, \ldots, n$; the latter of the points $(i/n, A_{ki} - i/n)$. Exact inference corresponds to the former being the 45° line across the unit square, or the latter coinciding with the horizontal axis.

For the first-order bootstrap discrepancy, the P value plot is the locus of points (A_{ki}, A_{ki}^2), and the P value discrepancy plot of the points $(A_{ki}, A_{ki}^2 - A_{ki})$. This is because $A_{ki}^2 = R_1(A_{ki}, k)$ is the probability that the bootstrap P value is less than A_{ki}; see (8) and (9). In general, for the order-r bootstrap discrepancy, the P value plot is the locus of the points (A_{ki}^r, A_{ki}^{r+1}), and similarly for the P value discrepancy plot.

5. An Example

Whereas it is easy to make the set of P values discrete, for instance by choosing $n = 100$ and thus letting a P value take on values $0.00, 0.01, 0.02, \ldots, 0.99, 1.00$ only, there are not very many examples for which it is obvious how to make the null model $\mathbb{M}_0$ discrete. An exception to this, which we study in this section, is the model treated in Davidson (2010). The model can be expressed as

$$y_t = y_{t-1} + u_t, \quad u_t = v_t + \theta v_{t-1}, \quad t = 1, \ldots, n \tag{21}$$

where the v_t are Gaussian white noise with variance 1, so that y_t is a unit-root process, obtained by cumulating the MA(1) process u_t with MA parameter θ. Since (21) takes the form of a recurrence relation, it must be initialised. In order to make u_t stationary, v_0 is set to a standard normal variable, independent of the v_t, $t = 1, 2, \ldots, n$.

The model is tested by a unit-root test, for which the test statistic is the τ_c version of the augmented Dickey-Fuller (ADF) test, proposed by Dickey and Fuller (1979), and justified asymptotically under much less restrictive assumptions by Said and Dickey (1984) and Phillips and Perron (1988). The statistics are computed using the ADF testing regression

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

where p is a lag truncation parameter, usually chosen in a data-determined way on the basis of some information criterion. When this regression is run by ordinary least squares, the τ_c statistic is the conventional t statistic for the hypothesis that $\beta_1 = 0$. Under the null hypothesis that the series y_t has a unit root, this statistic has a well-known but nonstandard asymptotic distribution.

With all of these assumptions, the model (21) is purely parametric, with only one parameter, namely θ. It is therefore easy to make the model discrete, by choosing a grid of values for θ. Since the model is parametric, the bootstrap is also parametric, and is characterised by a value of θ. In our discrete model, these values are restricted to the chosen grid. The parameter θ has to be estimated under the null hypothesis, as discussed in Davidson (2007) – the second "Golden Rule of Bootstrapping". In his earlier study, Davidson used a nonlinear least-squares (NLS) procedure that gave results comparable to those on maximum likelihood.

Testing for a unit root in a series obtained by summing a stationary MA(1) process with a parameter close to -1 leads to serious size distortions under the null, on account of the near cancellation of the unit root by the MA component in the driving stationary series u_t. Davidson (2010) found that, for a sample size of 100, the best results in terms of minimising size distortion under the null were obtained with a lag truncation parameter of $p = 12$. However, since we wish to study the effect of bootstrap iteration, it is very desirable to consider a setup in which the size distortion is non-negligible, and so the neighbourhood of $\theta = -1$ is suitable in this regard. Further, as θ approaches -1 from above, the distortion changes sign from negative to positive.

In Figure 1 are plotted the 0.01, 0.05, and 0.10 quantiles of the distribution of the asymptotic P value obtained for the raw τ_c statistic by use of the program described in MacKinnon (1994) and MacKinnon (1996), plotted as a function of θ from -0.4 to -0.99.

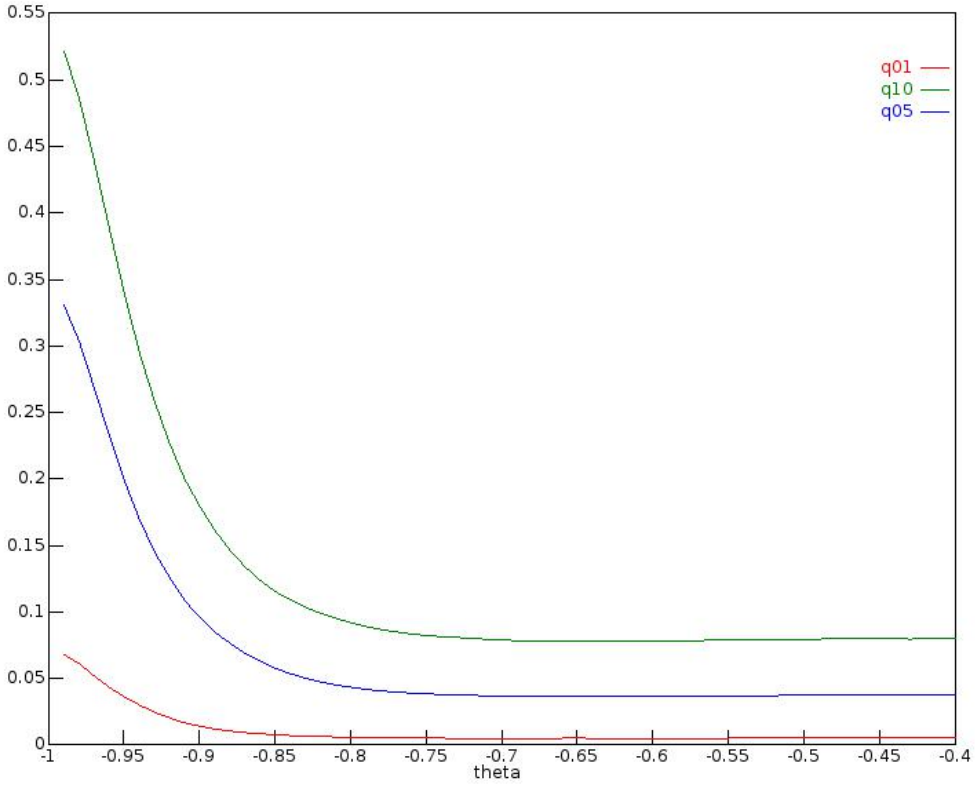

Figure 1

It can be see that, for values of θ from -0.4 to around -0.7, the curves are nearly flat, with P values close to nominal. Thus the distribution does not change much over this range. From -0.7 to -0.99, however, the distribution is very sensitive indeed to the value of θ, and this implies difficulties for the bootstrap if data are generated by a DGP in this region. The greatest size distortions are to be expected around $\theta = -0.90$, since it is in that neighbourhood that the curves bend most. Closer to $\theta = -1$, although the curves are very far from flat, they follow more or less straight lines. According to the analysis in Davidson and MacKinnon (1999), this fact helps to reduce distortion more than in regions in which the graphs are curved.

The discrete model is as follows:

- The number of possible test statistics, $n + 1$, is given by $n = 100$, so that the possible values are, as above, those of the grid $0.00, 0.01, 0.02, \ldots, 0.99, 1.00$.
- The τ_c test statistic is converted into an asymptotic P value by use of the program described in MacKinnon (1994) and MacKinnon (1996). The P value is rounded to the closest point on the grid defined above.
- The discrete model $\mathbb{M}_0$ is defined, with $m = 60$ points, for the DGPs (21) with parameters θ on the grid $-0.40, -00.41, -0.42, \ldots, -0.98, -0.99$. The value of $\theta = -1$ must of course be excluded, as belonging to the alternative hypothesis by which y_t is an I(0) process rather than to the null of an I(1) process.

- The parameter θ which defines the bootstrap DGP is estimated by the NLS procedure mentioned above. Values greater than -0.40 are rounded down to this value; values between -0.99 and -1 are rounded up to -0.99; intermediate values are rounded to the nearest point on the grid.

For the iteration procedure defined in Section 3 to be implemented, it is necessary to estimate by simulation the $60 \times 101 \times 60 = 363{,}600$ probabilities p_{kij} defined in (3). To this end, 60 separate simulation experiments were undertaken, for each of the 60 DGPs of the discrete model, on a machine with 64 cores. There were 100,000 replications for each of the 60 experiments. When all of the p_{kij} had been estimated in this way, the matrix $\boldsymbol{A}$ defined in (4) was constructed, and then the successive matrices $\boldsymbol{A}^r$ computed using (18). Convergence of the recurrence was achieved up to the third decimal place of the probabilities after 43 iterations. Since one iteration of the bootstrap gives what we call the double bootstrap, it is the 44-tuple bootstrap which can be identified with the infinitely iterated bootstrap. As expected on the basis of the evidence in Figure 1, convergence was slowest for the DGPs with θ around -0.95.

Selected results

The iterative procedure yields so many results, many of them rather uninteresting, that I limit myself here to a selection of more interesting things that emerge from the computations.

For values of θ where the distribution of the P value varies comparatively little, convergence is quite rapid. For $\theta = -0.50$, adequate convergence was achieved after only 5 iterations; for $\theta = -0.60$, after 6 iterations; for $\theta = -0.70$, 11 iterations were needed. Once the sensitive region is entered, 16 iterations were needed for $\theta = -0.80$ and $\theta = -0.85$, 22 iterations for $\theta = -0.90$, 26 for $\theta = -0.94$, but only 17 for $\theta = -0.99$. The 43 iterations were needed for a somewhat tighter criterion of convergence applied to all values of θ.

In Figures 2, 3, and 4, are shown results similar to some results in Davidson (2010). For $\theta = -0.90$, -0.95, and -0.99, P value discrepancy plots are given for bootstrap iterations, starting with the ordinary (single) bootstrap, through the quintuple bootstrap. For $\theta = -0.90$, the discrepancy plots appear to converge monotonically to the horizontal axis, implying that the quintuple bootstrap and higher iterations give exact inference up to the errors induced by the discretisation and the simulation errors in estimating the probabilities p_{kij}. However, for $\theta = -0.95$ things look rather different. The results are plainly contaminated by more noise than for $\theta = -0.90$, and it is not clear whether there really is convergence. For $\theta = -0.99$, the switch from under- to over-rejection is evident. Here, it is possible to believe that the dicrepancy plots are converging to the horizontal axis, but that has not happened even for the quintuple bootstrap.

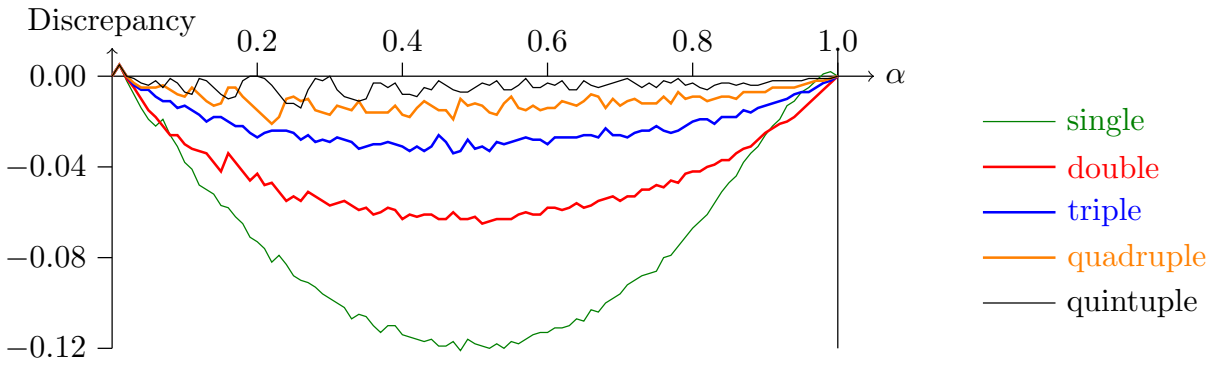

Figure 2: $n = 100$, θ = -0.90

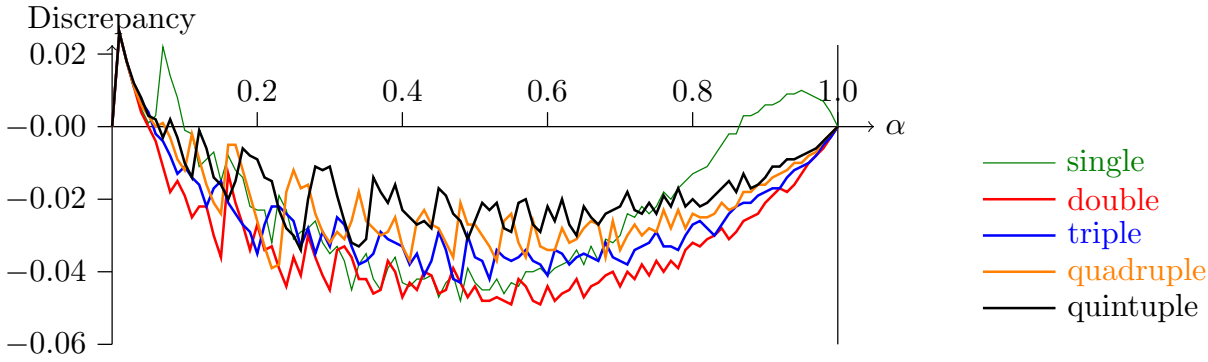

Figure 3: $n = 100$, θ = -0.95

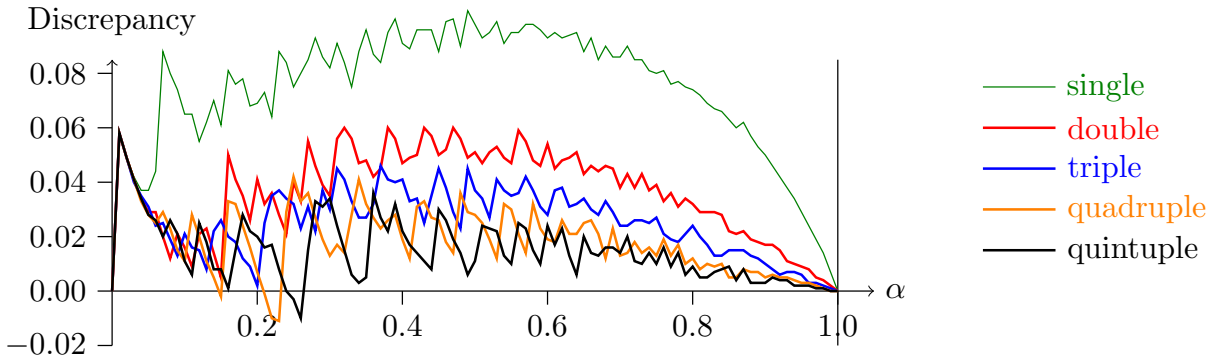

Figure 4: $n = 100$, θ = -0.99

It can be remarked here that, for values of θ greater than -0.90, convergence to the horizontal axis is essentially complete by the quadruple or quintuple bootstrap. One example is presented in Figure 5, for $\theta = -0.70$. Notice the scale of the vertical axis. Evidently, the distortion even of the single bootstrap is not very great.

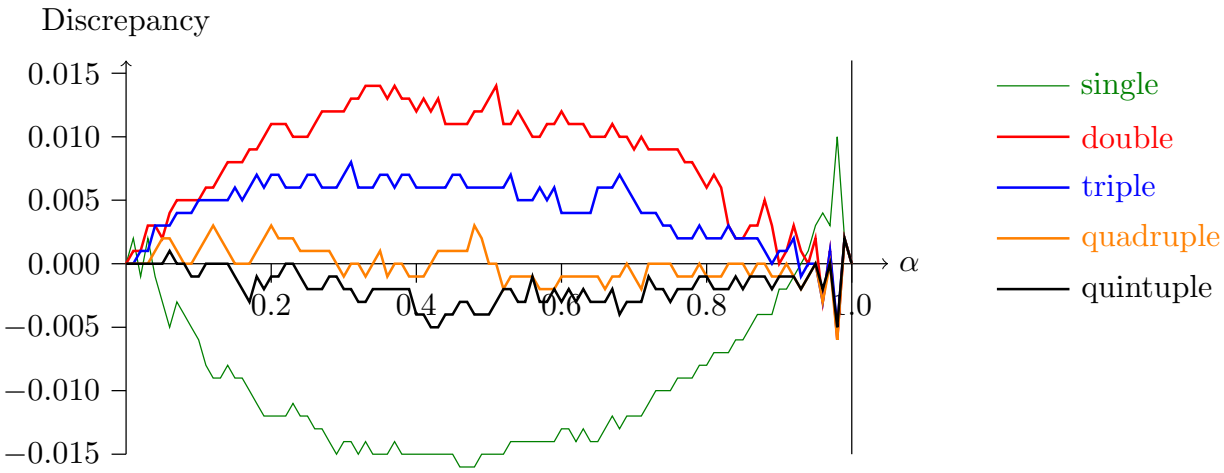

Figure 5: $n = 100$, θ = -0.70

After the recurrence has converged, we can look at the discrepancy plot for the infinitely iterated, or 44-tuple, bootstrap. In Figure 6, these discrepancy plots are shown for $\theta = -0.70$, -0.90, -0.95, and -0.99. Because the recurrence converged, the discrepancy in all cases must be exactly equal to zero for the nominal levels in the relevant rows of the matrix we may write as $\boldsymbol{A}^{\infty}$. All but the first few of these levels are in the neighbourhood of 1. Indeed, all of the curves cross the horizontal axis several times, and are almost coincident with it for values of α close to 1. The fact that the discrepancies are not zero everywhere is of course a consequence of the discrete nature of the model. For $\theta = -0.95$ and -0.99, there is a peak near $\alpha = 0$, for the same reason. However, even for $\theta = -0.90$, and more so for $\theta = -0.70$, the discrepancy is everywhere very small indeed.

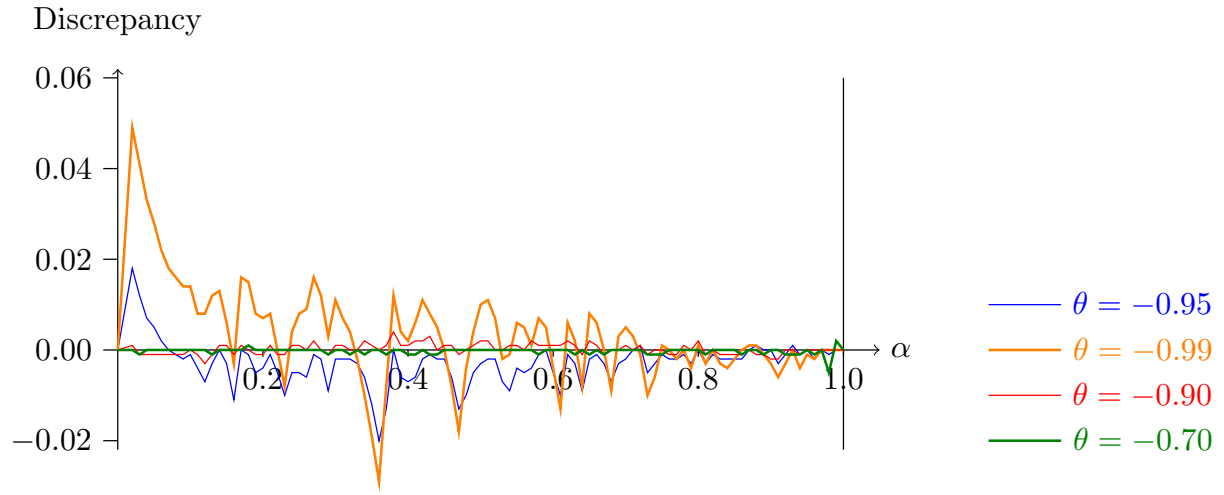

Figure 6: $n = 100$, fully iterated bootstraps

6. Discussion

The study in this paper raises a number of issues. One obvious one is to what extent the use of a discrete model is feasible in practice, with the sort of bootstrap commonly

used, rather than one governed by a single scalar parameter. Two parameters are probably feasible, but with any more one swiftly runs afoul of the curse of dimensionality.

It is clear that the most important advantage of this discrete approach is that it eliminates conventional bootstrapping based on a simulation experiment. In exchange, it is necessary to conduct the probably costly simulation experiment needed in order to estimate the p_{kij} of (3). However, once that experiment is carried out, it serves as a fixed overhead for arbitrary levels of bootstrap iteration. It thus becomes feasible to examine the convergence or otherwise of the sequence of iterated bootstrap P values.

In the discrete case, convergence of the sequence of iterated bootstrap P values is probably guaranteed. It remains to be seen whether this can be proved, and, if so, under what regularity conditions. However, this need not imply that exact inference is possible. In the example of Section 5, we saw that, with $\theta = -0.99$, the P value discrepancy after convergence is exactly zero only for a specific set of significance levels. Away from the neighbourhood of $\theta = -1$, we saw that this hardly mattered, as those values were spread almost uniformly over the $[0, 1]$ interval, while, for $\theta = -0.99$, they were almost all close to 1.

It is quite possible to concoct pathological examples in which, after convergence, the only levels capable of giving exact inference are zero and one; hardly a desirable state of affairs. One such example is given by a model with only two DGPs, and a statistic the distribution of which under one of these strictly stochastically dominates the distribution under the other. If the bootstrap DGP is always, deterministically, just one of the two, then, under that one, as we saw in Section 4, convergence is immediate and allows exact inference. Under the other DGP, numbered 2 say, however, convergence is eventually achieved, with all of the A_{2i} equal to 0 or 1, except for the first or the last, depending on the direction of the dominance.

What we saw for $\theta = -0.99$ is on the way to being an example of this, but not quite. This brings out another point, which is that, for one and the same model, for some DGPs we may have convergence, in the discrete case, of the sort we observed for θ greater than -0.90, while, for others, it may be more like what we saw with $\theta = -0.99$. In the continuous case, convergence to a uniform limit may occur for some DGPs in the model, but not for others. If there is such convergence for all DGPs in the null model, then the infinitely iterated bootstrap P value is a pivot with respect to that model.

The most frequently used sort of bootstrap is a resampling bootstrap. This implies that the bootstrap distribution is discrete, although it is usually presumed that the underlying distribution or distributions are continuous. It is in any case impossible to iterate a resampling bootstrap indefinitely, because each resample has fewer separate objects than the sample from which the resample is drawn. If one iterates a resampling bootstrap, eventually there will be only one element of the original sample left, which one presumably being randomly selected with equal probabilities for all the elements. "Convergence" in such a case would be meaningless.

If we abstract from the simulation noise in the estimation of the p_{kij}, the discrete model is quite nonrandom. We are, in effect, working simultaneously with every point in the outcome space. Convergence, therefore, is to be understood in the ordinary sense of convergence of a sequence of real numbers. In the continuous case, of course, we have to speak of stochastic convergence, which may perhaps be almost sure, or in probability. If this discrete approach were to be used with real data, it would be necessary to use these data to compute realisations of the quantity being bootstrapped and of the bootstrap DGP, and then to discretise them according to the plan of discretisation in use. If the realised quantity is indexed by i and the realised bootstrap DGP by j, then for the r-tuple bootstrap, the bootstrap P value is A^r_{ji}.

Many questions remain regarding the numerical stability of the discrete model. In the example in Section 5, I chose to declare convergence when the sum of the absolute values of the differences between the elements of the matrices $\boldsymbol{A}^r$ and $\boldsymbol{A}^{r+1}$ was less than 0.1. Until the iterations were stopped, it was observed that this sum decreased monotonically with each new iteration. This is suggestive of numerical stability, but is not definitive, of course. It would also be useful to examine to what extent simulation error in the preliminary estimation of the p_{kij} is propagated through the iterations; I plan to study this in future work. Another point that will bear future investigation is to what extent coarseness or fineness of the discretisation matters, both for the statistics (the P values) and for the DGPs of the model. If relatively coarse discretisation yields satisfactory results, this would be of enormous importance for any practical use of this approach.

It would be immensely useful to find ways of discretising the set of bootstrap DGPs used in situations that are not purely parametric. While it is easy enough to replace the use of a discrete empirical distribution for resampling by a continuous version, thus avoiding the problem inherent to iterating a conventional resampling bootstrap, it is not obvious how to make discrete the set of bootstrap DGPs that would be obtained in this way. I conjecture that, when bootstrapping an approximately pivotal statistic, it may be possible to cover the set of bootstrap DGPs rather coarsely and still achieve satisfactory results. How best to do so remains to be seen.

The double bootstrap was introduced by Beran in two papers, Beran (1987) and Beran (1988), in which he refers to "pre-pivoting", meaning making some quantity more close to being pivotal for a model by bootstrapping it, and then bootstrapping the result. This interpretation clearly applies to higher orders of bootstrap iteration. In some sense, the iterative procedure, if it converges appropriately, serves to project the original statistic into a space of pivotal statistics. It would be desirable to formalise this intuition. It is also necessary to see to what extent this "projection" may adversely affect the power of a test. Of course, power is not uniquely defined when a non-pivotal statistic is used, but, as remarked previously, if an iterated bootstrap P value follows the uniform $\mathrm{U}(0,1)$ distribution, it is by definition a pivot.

In discussing bootstrap "validity", it is conventional to make use of an appropriate asymptotic construction in order to show that the limiting distribution of the quantity considered is the same as the limiting distribution of its bootstrap counterpart. This is of course a

very weak requirement. A somewhat better justification for the bootstrap comes from any refinements that can be demonstrated by an asymptotic argument, as in Hall's (1992) book, where he uses Edgeworth expansion. But the bootstrap is not intrinsically an asymptotic procedure; rather, current bootstrap *theory* relies on asymptotic arguments. It seems to me that convergence of the sequence of iterated bootstrap P values to the uniform distribution is a much richer and more satisfactory means of justifying or validating the bootstrap. No asymptotic argument is involved, so that the potential arbitrariness of the choice of an asymptotic construction is avoided. To the extent that the approach of this paper can be made operational for problems of interest, the approach carries its validity along with it.

Further, the new proposed criterion for validity is by no means equivalent to asymptotic validity. An example of this is when a regression model, the disturbances of which are not necessarily Gaussian, is bootstrapped using a bootstrap DGP that imposes Gaussianity. Under very weak conditions on the asymptotic construction, this bootstrap is asymptotically valid. But it certainly is not, by the criterion of convergence of iterated P values to $\mathrm{U}(0,1)$, for any DGPs in the model the disturbances of which are in fact not Gaussian.

7. Concluding Remarks

In this paper, I have tried to take a step towards realising my ambition of freeing bootstrap theory from the use of asymptotic methods. The main idea is that the bootstrap can be justified – or not, as the case may be – by the convergence of the iterations of the bootstrap. This idea is developed here by making it tractable to study the bootstrap to any order of iteration, by means of a discretisation procedure that makes the model under test represented by a finite three-dimensional array of probabilities.

Much work remains to be done if the approach of the paper is to be useful, either in practice, or in the further development of bootstrap theory. One question seems particularly urgent: can we find sufficient conditions for convergence of the sequence of iterated P values to the uniform $\mathrm{U}(0,1)$ distribution, conditions that can be checked without excessive trouble in situations of practical interest? This, and other related questions, will be pursued in future work.

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**, 687–97.

Davidson, R. (2007). "Bootstrapping Econometric Models", *Quantile*, **3**, 13–36 (in Russian). English version available from the author's website:
http://russell-davidson.arts.mcgill.ca/articles/

Davidson, R. (2010). "Size Distortion of Bootstrap Tests: Application to a Unit Root Test", *Review of Economic Analysis*, **2**, 169–93.

Davidson, R. and J. G. MacKinnon (1998). "Graphical Methods for Investigating the Size and Power of Hypothesis Tests," *The Manchester School*, **66**, 1-26.

Davidson, R. and J. G. MacKinnon (1999). "The Size Distortion of Bootstrap Tests," *Econometric Theory*, **15**, 361-76.

Dickey, D. A., and W. A. Fuller (1979). "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, **74**, 427–31.

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

MacKinnon, J. G. (1994). "Approximate asymptotic distribution functions for unit root and cointegration tests," *Journal of Business and Economic Statistics*, **12**, 167–76.

MacKinnon, J. G. (1996). "Numerical distribution functions for unit root and cointegration tests", *Journal of Applied Econometrics*, **11**, 601–18.

Phillips, P. C. B., and P. Perron (1988). "Testing for a unit root in time series regression," *Biometrika*, **75**, 335–46.

Said, E. S., and D. A. Dickey (1984). "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, **71**, 599–607.