Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 133 (2006) 841–862
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Bootstrapping the Box–Pierce Q test: A robust test of uncorrelatedness

Joel L. Horowitz[a], I.N. Lobato[b], John C. Nankervis[c], N.E. Savin[d,*]

[a]Department of Economics, Northwestern University, Evanston IL 60208-2600, USA
[b]Centro de Investigaciön Econömica and Departamento de Economïa Instituto Tecnolögico Autönomo de México, Av Camino Sta Teresa 930, Mexico DR 10700, Mexico
[c]Department of Accounting, Finance and Management, University of Essex, Colchester CO4 3SQ, UK
[d]Department of Economics, University of Iowa, Iowa City IA 52242, USA

Available online 22 August 2005

Abstract

This paper describes a test of the null hypothesis that the first K autocorrelations of a covariance stationary time series are zero in the presence of statistical dependence. The test is based on the Box–Pierce Q statistic with bootstrap-based P-values. The bootstrap is implemented using a double blocks-of-blocks procedure with prewhitening. The finite sample performance of the bootstrap Q test is investigated by simulation. In our experiments, the performance is satisfactory for samples of $n = 500$. At this sample size, the differences between the empirical and nominal rejection probabilities are essentially eliminated.

JEL classification: C4

Keywords: Serial correlation tests; Box–Pierce Q; Blocks of blocks bootstrap; Adjusted P-values; Double bootstrap

*Corresponding author. Tel.: +9 319 335 0855; fax: +1 319 335 1956.
E-mail addresses: joel-horowitz@northwestern.edu (J.L. Horowitz), ilobato@itam.mx (I.N. Lobato), jcnankervis@essex.ac.uk (J.C. Nankervis), gene-savin@uiowa.edu (N.E. Savin).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.06.014

1. Introduction

The Box–Pierce (1970) Q_K statistic is commonly used to test the null hypothesis that the first K autocorrelations of a covariance stationary time series are zero. The Q_K statistic is asymptotically distributed as chi-square with K degrees of freedom when the null is true and the observations are independently and identically distributed. If the null hypothesis is true but the time series is statistically dependent, the Q_K test can produce seriously misleading inferences when the critical value or P-value is obtained from the chi-square distribution. Time series models that generate uncorrelated but statistically dependent observations have been widely used in economics and finance. The GARCH model for stock returns is a leading example. In this paper, a block bootstrap procedure is used to estimate the distribution of the Q_K statistic when the data are uncorrelated but dependent. The paper presents the results of a Monte Carlo investigation of the numerical performance of this bootstrap procedure.

The block bootstrap is a procedure for generating bootstrap samples from time series when a parametric model is not available. The blocking procedure consists of dividing the data into blocks and sampling the blocks randomly with replacement. Under mild regularity conditions, the block bootstrap provides a first-order approximation to the distribution of test statistics. In other words, the block bootstrap produces the right asymptotic distribution whereas the chi-square approximation does not in the setting that we consider. Romano and Thombs (1996) have proposed using the block bootstrap to make robust inferences about the individual autocorrelation coefficients in the presence of statistical dependence.

When a test statistic is asymptotically pivotal, the block bootstrap provides approximations that are more accurate than the approximations of first-order asymptotic theory under certain regularity conditions (Hall et al., 1995; Hall and Horowitz, 1996; Andrews, 2002). However, the Q_K statistic is not asymptotically pivotal in the presence of statistical dependence. Hence, there is no reason for supposing that the block bootstrap provides a higher-order approximation to the distribution of the Q_K statistic.

In this paper, the Q_K test statistic with block bootstrap-based P-values is used to test the null hypothesis that the first K autocorrelations are zero. The Q_K statistic is not a studentized statistic. Studentization requires a heteroskedastic and autocorrelation consistent (HAC) estimator of the covariance matrix of the correlation coefficients. We use Q_K instead of a studentized test statistic because of computational considerations. In addition, the HAC estimator can be very imprecise as well as difficult to compute. The imprecision of the HAC estimator may decrease the power of a test based on a studentized statistic.

The blocking method employed here is the blocks-of-blocks (BOB) bootstrap with overlapping blocks (Künsch, 1989). See Davison and Hinkley (1997) for further explanation of this method. Hall et al. (1995) and Lahiri (1999) show that the moving block bootstrap based-estimator is superior to the non-overlapping block bootstrap (Carlstein, 1986) and that both are superior to the estimator based on the stationary bootstrap of Politis and Romano (1994). Our Monte Carlo results confirm that the BOB bootstrap is rather insensitive to the choice of block length.

Beran (1988) gives conditions under which iterating the bootstrap can produce further reductions in the difference between the true and nominal probability of rejecting a correct null hypothesis (hereinafter error in the rejection probability or ERP) when the data are a random sample and the statistic is asymptotically pivotal. Specifically, bootstrap iteration increases the rate at which the ERP converges to zero. This does not happen with the block bootstrap. Nonetheless, Monte Carlo evidence indicates that iterating the block bootstrap can reduce the finite-sample ERP of a test and the finite-sample difference between the true and nominal coverage probabilities of a confidence interval (error in the coverage probability or ECP). See, for example, Politis et al. (1997) and Romano and Wolf (2000).

Künsch (1989) and Politis et al. (1997) have proved the consistency of the single block bootstrap; the consistency of the double block bootstrap is proved in Appendix A. However, there is at present no theoretical explanation of the ability of the iterated block bootstrap to reduce finite-sample ERPs and ECPs. One possible explanation is that block bootstrap iteration reduces the constants that multiply the rates of convergence of the ERP and ECP. Another possibility is that block bootstrap iteration reduces the sizes of higher-order terms in asymptotic expansions of ERPs and ECPs. Regardless of the underlying cause, the simulation evidence that block bootstrap iteration reduces the finite-sample ERP of a test motivates us to carry out experiments with the iterated BOB bootstrap.

This paper investigates the numerical performance of the Q_K test when the P-value is obtained using the single block-of-blocks or SBOB bootstrap and the double blocks-of blocks or DBOB bootstrap. We refer to the Q_K tests that use SBOB and DBOB bootstrap P-values as SBOB and DBOB bootstrap tests, respectively. In the Monte Carlo experiments analyzed, the data are generated by stochastic processes that are martingale difference sequences (MDSs) as well as by non-MDS processes. The MDS examples use a one-dependent process considered by Romano and Thombs (1996), a Gaussian GARCH model, and a non-Gaussian GARCH model. The motivation for entertaining non-MDS processes is the growing evidence that the MDS assumption is too restrictive for financial data; see El Babsiri and Zakoian (2001). The non-MDS examples use a nonlinear moving average model and a bilinear model.

For the examples used in the experiments, the DBOB bootstrap reduces the ERP to nearly zero with sample sizes of 500 or more. Moreover, the DBOB bootstrap achieves lower ERPs than does the single BOB bootstrap. Although we have no theoretical explanation for these results, we note that they add to existing Monte Carlo evidence that iterating the block bootstrap reduces ERPs and ECPs. The development of a theoretical explanation for this phenomenon may be a worthwhile topic for future research.

Finally, the performance of the Q_K test is compared to that of other tests of uncorrelatedness that can be viewed as extensions of the Q_K test. These tests are the

Q_K^* test (Diebold, 1986; Lo and MacKinlay, 1989; Lobato et al., 2001), the GP_K test (Guo and Phillips, 1998), and the $\tilde{Q}_K$ test (Lobato et al., 2002). The Q_K^* and GP_K tests are designed for time series generated by MDS processes. The Q_K^* test assumes that the asymptotic covariance matrix of the sample autocorrelations is diagonal. The GP_K test does not make the diagonality assumption and hence is more general than the Q_K^* test. The $\tilde{Q}_K$ test is asymptotically valid for both MDS and non-MDS processes, and, hence, is a natural competitor to the Q_K test with bootstrap-based P-values. For expositional purposes, we refer to the Q_K^*, GP_K and $\tilde{Q}_K$ tests as robust tests.

The Markov conditional bootstrap (Horowitz, 2003) is an alternative to the block bootstrap when the process is a Markov process or can be approximated by one with sufficient accuracy. However, this procedure is not the focus of the paper. Andrews and Ploberger (1996) and Goncalves and Kilian (2003) have proposed tests of the null that a time series is iid. These tests have been developed for specific classes of data generating processes. These tests are likely to be more powerful than the DBOB Q_K test when their restrictive assumptions are satisfied. On the other hand, if the assumptions do not hold, then these tests are not valid. These tests are also not examined here. The aim of our paper is to investigate the performance of a test procedure that is consistent under very general conditions.

The remainder of the paper is organized as follows. Section 2 describes the Q_K test with SBOB and DBOB bootstrap-based P-values. Section 3 reports the empirical rejection probabilities of the Q_K test with SBOB and DBOB bootstrap-based P-values when the null is true for MDS examples and non-MDS examples. The empirical rejection probabilities of the Q_K^*, GP_K and $\tilde{Q}_K$ tests based on asymptotic P-values are also reported. Section 4 compares the empirical power of the Q_K test based on DBOB bootstrap-based P- values with the empirical power of the $\tilde{Q}_K$ test based on asymptotic P-values. An empirical example is presented in Section 5. Concluding comments are in Section 6. The proof that the proposed SBOB and DBOB bootstrap procedures provides a consistent estimator of the distribution of Q_K statistic under the null hypothesis is given in Appendix A. The stopping rules used to reduce the computation time for the double bootstrap tests are summarized in Appendix B.

2. Bootstrap test

The null hypothesis can be tested by comparing the Q_K statistic to a bootstrap-based critical value, or what is equivalent, by comparing a bootstrap-based P-value to α, the nominal probability of making a type I error. For this purpose, we use the SBOB and DBOB bootstrap with prewhitening to calculate the P-values. In the Monte Carlo experiments we compare the performance of the SBOB and DBOB bootstrap tests. The first objective of this section is to describe the calculation of bootstrap P-values for the Q_K test using the SBOB and DBOB bootstrap. The second objective is to describe the prewhitening procedure employed.

2.1. Preliminaries

Let $y_1, \ldots, y_n$, denote a real-valued strictly stationary time series with mean μ. Define the lag-j autocovariance by $\gamma(j) = E(y_t - \mu)(y_{t+j} - \mu)$ and the lag-j autocorrelation by $\rho(j) = \gamma(j)/\gamma(0)$. Define the sample mean, sample variance and sample autocovariance by $m = \sum_{t=1}^{n} y_t/n$, $c(0) = \sum_{t=1}^{n}(y_t - m)^2/n$ and $c(j) = \sum_{t=1}^{n-j}(y_t - m)(y_{t+j} - m)/n$. Then the usual estimator of $\rho(j)$ is $r(j) = c(j)/c(0)$. Under general weak dependence conditions, the vector $n^{1/2}r = n^{1/2}[r(1), \ldots, r(K)]'$ is asymptotically normally distributed with asymptotic covariance matrix V, where the ijth element of V is given in Hannan and Heyde (1972) and Romano and Thombs (1996). If V is known, H_K: $\rho = [\rho(1), \ldots, \rho(K)]' = 0$ can be tested using a test statistic of the form $nr'V^{-1}r$, which asymptotically is chi-square distributed with K degrees of freedom when H_K is true. In practice, V is unknown. A feasible test can be obtained either by replacing V by a known matrix or by estimating V.

The Box–Pierce Q_K statistic (Box and Pierce, 1970) replaces V with the identity matrix. The Q_K^* test replaces V with an estimator that is consistent under the null for MDS processes where the asymptotic covariance matrix of the sample autocorrelations is diagonal, and the GP_K test replaces V with an estimator that is consistent under the null for MDS processes. The $\tilde{Q}_K$ test replaces V with an estimator that is consistent under the null for both MDS and non-MDS processes; for details, see Lobato et al. (2002).

In this paper, H_K is tested using the P-value of the Q_K statistic. Each sample of n observations $y_1, \ldots, y_n$ produces a specific value of Q_K, say τ. For any fixed number z, let $S(z) = P(Q_K > z|\mathrm{H}_K)$. The P-value associated with τ is $p = S(\tau)$. The exact symmetric test of H_K rejects at level α if $p \equiv S(\tau) < \alpha\cdot$ The P-value can be calculated from some predetermined distribution or estimated by the bootstrap. We now show how to obtain an estimate of the P-value using the SBOB bootstrap and DBOB bootstrap.

2.2. Single bootstrap

In order to implement the BOB bootstrap, we define a new $(K+1) \times (n-K)$ data matrix as $(Y_1, Y_2, \ldots, Y_{n-K})$ where $Y_i = (y_i, y_{i+1}, \ldots, y_{i+K})' = (y_i^1, y_i^2, \ldots, y_i^{K+1})'$. Ignoring prewhitening, which reduces the number of observations from n to $n-K$, the bootstrap sample is obtained by resampling blocks from the $K+1$ dimensional series and creating a sample of length n from the blocks. Denote the block size by b and let $h = n/b$. Let B_i be a $(K+1) \times b$ matrix given by $B_i = Y_i, \ldots, Y_{i+b-1}$, where $i = 1, \ldots, q$, and $q = n - b - K + 1$. The SBOB bootstrap test is obtained by using the following algorithm:

1. Sample randomly with replacement h times from the set $\{B_1, \ldots B_q\}$. This produces a set of blocks $B_1^*, \ldots, B_h^*$. These blocks are then laid end-to-end to form a new time series matrix of order $(K+1) \times n$, which is the bootstrap sample and is denoted by $Y^* = (Y_1^*, \ldots, Y_n^*)$, where $Y_i^* = (y_i^{1^*}, y_i^{2^*}, \ldots, y_i^{(K+1)^*})'$ is a bootstrap replicate of Y_i.

2. Using the bootstrap sample, calculate the statistic $Q_K^S = n\sum_{k=1}^{K}[r^*(k) - r_b(k)]^2$ where $r_b(k)$ is defined below

$$r^*(k) = \sum_{t=1}^{n}\left(y_t^{1^*} - \bar{y}^{1^*}\right)\left(y_t^{(k+1)^*} - \bar{y}^{(k+1)^*}\right) \Bigg/ \left[\sum_{t=1}^{n}\left(y_t^{1^*} - \bar{y}^{1^*}\right)^2 \sum_{t=1}^{n}\left(y_t^{(k+1)^*} - \bar{y}^{(k+1)^*}\right)^2\right]^{1/2}$$

and $\bar{y}^{j*} = \sum_{t=1}^{n} y_t^{j*}/n$.

3. Repeat steps 1 and 2 M_1 times.

Due to the use of overlapping blocks, some observations receive more weight than others in the set $\{B_1,\ldots,B_q\}$. As a result, the Q_K^S statistic defined above is centered using the estimator

$$r_b(k) = \sum_{t=1}^{n-K} w_t(y_t^1 - \bar{y}^1)\left(y_t^{(k+1)} - \bar{y}^{(k+1)}\right) \Bigg/ \left[\sum_{t=1}^{n-K} w_t(y_t^1 - \bar{y}^1)^2 \sum_{t=1}^{n-K} w_t\left(y_t^{(k+1)} - \bar{y}^{(k+1)}\right)^2\right]^{1/2}$$

with $\bar{y}^k = \sum_{t=1}^{n-K} w_t y_t^k$, where

$$\begin{aligned} w_t &= t/b(n-K-b+1), \quad t = 1,\ldots,b-1 \\ &= 1/(n-K-b+1), \quad t = b,\ldots,n-K-b+1 \\ &= (n-K+1-t)/b(n-K-b+1), \quad t = n-K-b+2,\ldots,n-K. \end{aligned}$$

The empirical distribution of the M_1 values of Q_K^S is the bootstrap estimate of the distribution of Q_K based on the single bootstrap. The SBOB bootstrap P-value, denoted by p_K^*, is an estimate of p where $p_K^* = \#(Q_K^S > Q_K)/M_1$ is the number of Q_K^S greater than Q_K divided by M_1. Given a nominal level of α, the SBOB bootstrap test of H_K rejects if $p_K^* < \alpha\cdot$

The bootstrap test based on p_K^* has rejection probability α if $P(p_K^* < \alpha/\mathrm{H}_K) = \alpha$, that is, if the distribution of p_K^* is uniform on [0, 1]. If the distribution is not uniform, there will exist some β such that $P(p_K^* < \beta/\mathrm{H}_K) = F_{p*}(\beta) = \alpha$. The unknown β is the inverse of F_{p^*} evaluated at α, $\beta = F_{p*}^{-1}(\alpha)$. This suggests that given an estimate of F_{p^*}, we can obtain an estimate of β and hence the error in the P-value. The double bootstrap can be used to estimate F_{p*} and therefore $\beta\cdot$

2.3. Double bootstrap

A double bootstrap sample is obtained by resampling blocks from a bootstrap sample $Y_1^*, \ldots, Y_n^*$ and creating a new sample of length n from these blocks. Again, let the block size be b, where $n = hb$. Let B_i^* be the block of b consecutive observations starting with Y_i^*; that is, $B_i^* = Y_i^*, \ldots, Y_{i+b-1}^*$, where $i = 1, \ldots, q$ and $q = n - b - K + 1$. The DBOB bootstrap test is described by the following algorithm:

Do steps (1) and (2) above.

1′. For each single bootstrap sample, sample randomly with replacement h times from the set $\left\{B_1^*, \ldots, B_q^*\right\}$. This produces a set of blocks $B_1^{**}, \ldots, B_h^{**}$. As above, these blocks are then laid end-to-end to form a new time series of length n, which is the double bootstrap sample $Y^{**} = (Y_1^{**}, \ldots, Y_n^{**})$ where $Y_i^{**} = \left(y_i^{1**}, \ldots, y_i^{(K+1)**}\right)'$.

2′. From the double bootstrap sample, calculate the statistic

$$Q_K^D = n \sum_{k=1}^{K} \left[r^{**}(k) - r_b^*(k)\right]^2$$

where

$$r^{**}(k) = \sum_{t=1}^{n} \left(y_t^{1**} - \bar{y}^{1**}\right)\left(y_t^{(k+1)**} - \bar{y}^{(k+1)**}\right) \Bigg/ \left[\sum_{t=1}^{n} \left(y_t^{1**} - \bar{y}^{1**}\right)^2 \sum_{t=1}^{n} \left(y_t^{(k+1)**} - \bar{y}^{(k+1)**}\right)^2\right]^{1/2}$$

and $\bar{y}^{j**} = \sum_{t=1}^{n} y_t^{j**}/n$. Here $r_b^*(k)$ is computed by applying the procedure for obtaining $r_b(k)$ to the SBOB sample.

3′. Repeat Steps 1′ and 2′ M_2 times.

4′. Repeat Steps 1, 2 and 3′ M_1 times.

For each one of the M_1 single bootstrap samples, there are M_2 values of the test statistic Q_K^D. Hence, there are M_1 double bootstrap P-values, denoted by p_K^{**}, where $p_K^{**} = \#(Q_K^D > Q_K^S)/M_2$. The empirical distribution function of these M_1 P-values, denoted by F_{p**}, is used as an estimate of F_{p*}. So the estimate of β, β^*, is given by $\beta^* = F_{p**}^{-1}(\alpha)$. Accordingly, for a nominal rejection probability of α, the double BOB test of H_K rejects if $p_K^* < \beta^*$. That is, the DBOB bootstrap test rejects if $p_{Ka}^* = F_{p**}(p_k^*) < \alpha$ where p_{Ka}^* is what Davison and Hinkley (1997) call the *adjusted P-value*. The adjusted P-value is estimated by $\#[p_K^{**} \leq p_K^*]/M_1$; this formula is also given by Hinkley (1989).

Davison and Hinkley (1997) strongly recommend the use of adjusted P-values. Politis et al. (1997) use what they call calibrated confidence intervals to obtain the correct coverage probability for parameters of dependent processes. Double bootstrap tests are the hypothesis testing analogs of calibrated confidence intervals. The performance of adjusted P-values and calibrated confidence intervals are the motivation for using the double bootstrap test in our setting. As noted in the introduction, the refinement provided by prepivoting (double bootstrap) in the iid case is not available in the case of dependent data.

2.4. Prewhitening

We investigated the performance of the bootstrap with and without pre-whitening of the data series. The details of the prewhitening procedure are described below. The rationale for prewhitening is that it reduces the sample autocorrelations to asymptotically negligible levels, thereby making the sample satisfy H_K approximately. The results of our Monte Carlo experiments reveal that prewhitening usually either reduces the ERP of the DBOB test or leaves the error unchanged. However, in one experiment, the test of H_5 with the bilinear model described in Section 5, prewhitening increases the ERP. In Section 4 we present only the results of the experiments with prewhitening. The results of experiments without prewhitening are available upon request.

The prewhitening is carried out by regressing y_t on K lags, that is, by using an AR(K) regression, where K is the maximum number of autocorrelations to be tested. The residuals from the AR(K) regression are used as the pre-whitened series, which acts as the sample in bootstrap resampling. To calculate the SBOB bootstrap test with prewhitening we first fit an AR(K) to the original data to obtain the residuals, $e_t = y_t - \hat{y}_t$. From the residuals we calculate the sample autocorrelations. For SBOB, $(K+1) \times b$ blocks are resampled from $E = (E_{K+1}, \ldots, E_{n-2K})$ where $E_i = (e_i, \ldots, e_{i+K})'$ to obtain the $(K+1) \times n$ bootstrap sample $E^* = (E_1^*, \ldots, E_n^*)$. This bootstrap sample is used in place of Y^* in steps 1–3 in calculating the SBOB bootstrap test.

For DBOB we first prewhiten the SBOB sample by fitting an AR(K) to the $(K+1) \times n$ elements of E^* (or equivalently using appropriate weighting of repeated elements). In order to prewhiten DBOB we consider an augmented $(2K+1) \times n$ matrix E^* calculated as above. We then fit an AR(K) using E^* in which each of the elements in the lower $(K+1)$ rows is regressed on the K higher elements in the column. Each element e_t in the lower $(K+1)$ rows of E^* is replaced by $u_t = e_t - \hat{e}_t$ to form the prewhitened SBOB sample $U^* = (U_1^*, \ldots, U_n^*)$ from which the autocorrelations are calculated. For DBOB, $(K+1) \times b$ blocks are resampled from the matrix U^* to form the $(K+1) \times n$ matrix U^{**}. This bootstrap sample is used in place of Y^{**} in steps $1'$–$4'$ in calculating the DBOB bootstrap test.

3. Monte Carlo evidence

This section examines the performance of the SBOB and DBOB bootstrap tests with prewhitening in a set of Monte Carlo experiments. The examples used in the experiments include three MDS processes and two non-MDS processes. We first review the simulation evidence for the MDS examples.

3.1. MDS examples

The first MDS example is motivated by experiments conducted by Romano and Thombs (1996). This illustrates how the tests perform for a simple one-dependent MDS process where the asymptotic covariance matrix of the sample autocorrelations is diagonal under H_K. The second and third examples illustrate how the tests perform for a GARCH (1,1) model when the errors are normally distributed and when they are distributed as a centered chi-square variable with 3 degrees of freedom. Under H_K, the asymptotic covariance matrix of the sample autocorrelations is diagonal when the errors are normal and nondiagonal when the errors are chi-square (3).

The tests with SBOB and DBOB bootstrap-based P-values are calculated using $M_1 = 999$ and $M_2 = 249$ replications. However, for the double bootstrap tests, stopping rules are used in order to reduce the computation time. Due to these rules, the actual number of bootstrap replications required is reduced by up to a factor of 15. The stopping rules are briefly described in Appendix B. For more details, see Nankervis (2002).

The tables in this section report the empirical rejection probabilities of bootstrap tests of H_K: $\rho(1) = \cdots = \rho(K) = 0$, $K = 1, 5, 10$, for samples of $n = 500$. The empirical rejection probabilities for the bootstrap tests are calculated using 5000 replications. The results for the bootstrap tests are reported for three block lengths, $b = 4$, 10 and 20.

The empirical rejection probabilities are also reported for the Q_K, Q_K^*, GP_K and $\tilde{Q}_K$ tests based on asymptotic P-values. The empirical rejection probabilities of the asymptotic tests are calculated using 25,000 replications. The performance of the asymptotic tests provides a benchmark for measuring the improvement achieved by the bootstrap tests. The $\tilde{Q}_K$ test is implemented using the VARHAC procedure described in Lobato et al. (2002) with the lag length chosen by BIC. The maximum lag length is set at 10.

The random number generator used in the experiments was the very long period generator RANLUX with luxury level 3; see Hamilton and James (1997). Calculations were performed on a Silicon Graphics R10000 system and a 500 MHz PC using double precision Fortran 77.

Example 1. (One-dependent case). Let $y_t = z_t \bullet z_{t-1}$ where $\{z_t\}$ is a sequence of *iid* $\mathrm{N}(0,1)$ random variables and $\bullet$ denotes multiplication. The y_t process is uncorrelated with $\rho(k) = 0$ for all k, but not independent. For this process, $\gamma_0 = E(y_t - \mu)^2 = 1$, $E(y_t - \mu)^3/\gamma_0^{3/2} = 0$, $E(y_t - \mu)^4/\gamma_0^2 = 9$, and V is the identity matrix except that $v_{11} = 3$. Romano and Thombs (1996) generated a sample of $n = 1000$ for this sequence and applied the single moving block bootstrap using $M_1 = 200$ replications and a block length of $b = 40$.

The numerical results of the Monte Carlo experiments for the above diagonal MDS example are summarized in Table 1. The main features of the results for Example 1 are the following:

(i) The Q_K test based on asymptotic P-values over-rejects by a very large margin: the maximum absolute difference (MAD) between the empirical and nominal rejection probability (percent) is about 12 percent at the nominal 1 percent level and 23 percent at the 10 percent level.

(ii) The DBOB bootstrap eliminates the distortions in the rejection probabilities for the first two hypotheses: the MAD is about 0.4 percent at the nominal 1 percent level and 1.5 percent at 10 percent. For the third hypothesis the distortions are almost eliminated: the MAD is about 0.3 percent at the nominal 1 percent level and 1.7 percent at 10 percent.

(iii) The SBOB bootstrap substantially reduces the distortions in the empirical rejection probabilities for all hypotheses.

Table 1
Rejection probabilities (percent) of tests: diagonal MDS, $n = 500$

Tests	H_1			H_5			H_{10}		
	1	5	10	1	5	10	1	5	10
Diagonal One-Dependent Case									
Q_K	12.8	25.0	33.3	6.4	15.4	22.9	4.8	12.1	18.7
SBOB Q_K									
$b = 4$	2.0	7.0	12.4	1.4**	6.1	11.1	1.0*	4.8*	10.3*
$b = 10$	2.4	8.1	13.3	1.9	7.1	12.8	1.2*	5.8**	11.7
$b = 20$	2.9	8.9	14.3	1.9	7.6	14.4	1.4**	6.7	13.6
DBOB Q_K									
$b = 4$	1.3*	5.6*	10.8*	1.1*	5.3*	9.9*	0.7*	3.8	8.3
$b = 10$	1.0*	6.2	11.1**	1.4*	5.6*	11.3	0.9*	4.6*	9.7*
$b = 20$	1.1*	6.2	11.0**	1.3*	6.2	11.5	0.9*	5.3*	10.8*
Q_K^*	0.8**	4.9*	10.1*	0.9*	4.5	9.6**	0.9*	4.6**	9.6*
GP_K	0.8**	4.9*	10.1*	0.6	3.9	8.8	0.4	3.4	7.9
$\tilde{Q}_K$	0.6	4.1	9.3	0.6	4.1	8.9	0.6	3.9	8.6

Notes: The numbers 1, 5, and 10 below H_1, H_5, and H_{10} are nominal rejection probabilities. The number of replications for the Q_K test with BOB bootstrap-based P-values is 5000. The number of replications for the asymptotic Q_K, Q_{K^*}, GP_K and $\tilde{Q}_K$ tests is 25,000. One asterisk (two asterisks) indicates that an asymptotic 95% (99%) confidence interval for the ERP contains zero.

(iv) The SBOB and DBOB tests are roughly insensitive to the choice of the block length, which confirms the findings of Davison and Hinkley (1997) for the BOB bootstrap.

The asymptotic Q_K^* test tends to work satisfactorily for all three hypotheses. The asymptotic GP_K test tends to under-reject for all three hypotheses, especially for H_{10}. The asymptotic $\tilde{Q}_K$ test under-rejects for all three hypotheses: the MAD is 0.4 percent at the nominal 1 percent level and 1.4 percent at 10 percent. Note that the asymptotic confidence intervals for the rejection probabilities are tighter for the asymptotic tests than for the bootstrap tests because the performance of the asymptotic tests is investigated using 25,000 replications.

Example 2. (Gaussian GARCH case). Let $y_t = z_{t\bullet}\sigma_t$, $\{z_t\}$ is an *iid* N(0, 1) sequence and $\sigma_t^2 = \omega + \alpha_0 y_{t-1}^2 + \beta\sigma_{t-1}^2$, where α_0 and β are constants such that $\alpha_0 + \beta < 1$. This condition is needed to insure that y_t is covariance stationary. He and Teräsvirta (1999) show that the unconditional fourth moment of y_t exists for GARCH (1,1) models if and only if $\beta^2 + 2\alpha_0\beta v_2 + \alpha_0^2 v_4 < 1$ where $v_i = E|z_t|^i$. Estimates from stock return data suggest that $\alpha_0 + \beta$ is close to 1 with β also close to 1; for example, see Bera and Higgins (1997). We set $\omega = 0.001$, $\alpha_0 = 0.05$ and $\beta = 0.90\cdot$ With this parameter setting, the He and Teräsvirta (1999) condition for the existence of the fourth moment of y_t is satisfied. The y_t process is uncorrelated with $\rho(k) = 0$ for all k, but not independent. For this process, $\gamma_0 = E(y_t - \mu)^2 = 0.02$, $E(y_t - \mu)^3/\gamma_0^{3/2} = 0$, $E(y_t - \mu)^4/\gamma_0^2 = 3.16$ and V is diagonal where the diagonal elements follow the recursion $v_{jj} = (1 - \alpha_0 - \beta) + (\alpha_0 + \beta)v_{j-1,j-1}$ where $v_{11} = 1.16$. Lobato et al. (2002) have also used this example.

Example 3. (Chi-square (3) GARCH case). This GARCH (1, 1) model is the same as in Example 2 except that now z_t is a demeaned and standardized chi-square random variable with 3 degrees of freedom. The He and Teräsvirta (1999) condition is also satisfied when z_t is a chi-square (3) random variable. In this case (the skewness is an estimate), $\gamma_0 = E(y_t - \mu)^2 = 0.02$, $E(y_t - \mu)^3/\gamma_0^{3/2} = 1.72$, $E(y_t - \mu)^4/\gamma_0^2 = 8.3$ where V is no longer diagonal.

The numerical results for the GARCH (1, 1) models are summarized in the first and second panels of Table 2. The Q_K test based on asymptotic P-values over-rejects by a large margin. In Table 1 the largest over-rejections occurred for H_1 while in Table 2 they occurred for H_{10}. The DBOB bootstrap essentially eliminates the

Table 2
Rejection probabilities (percent) of tests: GARCH(1, 1) models, $n = 500$

Tests	H_1			H_5			H_{10}		
	1	5	10	1	5	10	1	5	10
GARCH(1, 1) with normal errors									
Q_K	1.5	6.5	12.2	1.9	7.5	13.9	2.2	8.0	14.4
SBOB Q_K									
$b = 4$	1.0*	5.7**	10.8*	0.8*	5.0*	10.3*	1.0*	4.9*	9.8*
$b = 10$	1.3*	6.3	11.6	1.0*	5.5*	11.4	1.0*	5.4*	11.0*
$b = 20$	1.8	7.0	12.6	1.2*	6.2	12.9	1.0*	5.8**	12.6
DBOB Q_K									
$b = 4$	0.9*	5.2*	10.3*	0.7*	4.6*	9.1**	0.9*	4.1	8.0
$b = 10$	1.0*	5.4*	10.3*	0.8*	5.0*	10.0*	0.8*	4.7*	9.6*
$b = 20$	1.2*	5.7**	10.9**	1.0*	5.3*	10.8*	0.9*	4.8*	10.1*
Q_K^*	0.9*	5.0*	9.9*	1.0*	4.9*	9.7*	1.0*	4.9*	9.5**
GP_K	0.9*	5.0*	9.9*	0.8	4.5	9.4	0.7	3.8	8.5
$\tilde{Q}_K$	0.8**	4.9*	9.9*	0.8**	4.5	9.5**	0.6	4.2	8.8
GARCH(1, 1) with chi-square(3) errors									
Q_K	2.4	8.6	14.8	4.2	12.6	19.9	5.5	14.3	22.1
SBOB Q_K									
$b = 4$	1.4	6.2	11.8	0.9*	5.7**	11.4	1.3*	5.1*	10.1*
$b = 10$	1.7	6.7	12.5	1.2*	6.7	13.2	1.1*	5.5*	11.7
$b = 20$	2.1	7.5	13.5	1.5	7.3	14.7	1.2*	6.3	13.3
DBOB Q_K									
$b = 4$	1.0*	5.4*	10.2*	0.5	4.5*	9.7*	0.8*	4.0	8.1
$b = 10$	1.0*	5.5*	10.7*	0.7*	5.2*	10.9**	0.8*	4.6*	9.3*
$b = 20$	1.2*	5.7**	11.6*	0.8*	5.2*	11.4	0.9*	4.6*	10.0*
Q_K^*	1.1*	5.1*	10.1*	1.1*	5.5	10.9	1.3	5.8	10.8
GP_K	1.1*	5.1*	10.1*	0.9*	4.9*	10.1*	0.7	4.4	9.3
$\tilde{Q}_K$	1.0*	5.0*	10.1*	1.1*	5.3*	10.8	0.9*	5.0*	10.3*

Notes: See Table 1.

distortions in the rejection probabilities when the null is true for all three hypotheses. The distortions are much reduced by the SBOB bootstrap.

The asymptotic Q_K^* test works satisfactorily for all three hypotheses for the Gaussian GARCH model; it tends to over-reject somewhat for GARCH with chi-square (3) errors. The asymptotic GP_K test works satisfactorily for H_1 for Gaussian GARCH and for H_1 and H_5 for GARCH with chi-square (3) errors; otherwise, it tends to under-reject. The asymptotic $\tilde{Q}_K$ test works satisfactorily for H_1 and H_5 for Gaussian GARCH and for all three hypotheses for GARCH with chi-square (3) errors.

3.2. Non-MDS examples

The first uncorrelated non-MDS process is generated by a nonlinear moving average model, and the second is generated by a bilinear model. These nonlinear models are described in Tong (1990, pp.114–115) and also in Granger and Teräsvirta (1993). For these two examples, the asymptotic matrix of the sample autocorrelations is nondiagonal under the null.

Example 4. (Nonlinear moving average case). Let $y_t = z_{t-1} \bullet z_{t-2} \bullet (z_{t-2} + z_t + c)$ where $\{z_t\}$ is a sequence of *iid* N(0, 1) random variables and $c = 1.0$. The y_t process is uncorrelated with $\rho(k) = 0$ for all k, but not independent. For this process, $\gamma_0 = E(y_t - \mu)^2 = 5$, $E(y_t - \mu)^3/\gamma_0^{3/2} = 0$, $E(y_t - \mu)^4/\gamma_0^2 = 37.80$.

Example 5. (Bilinear case). Let $y_t = z_t + b \bullet z_{t-1} \bullet y_{t-2}$ where $\{z_t\}$ is a sequence of *iid* $\mathrm{N}(0, \sigma^2)$ random variables $b = 0.50$ and $\sigma^2 = 1.0$. The y_t process is uncorrelated with $\rho(k) = 0$ for all k, but not independent and is covariance stationary provided that $b^2 \sigma^2 < 1$. The fourth moment of this process exists if $3b^4\sigma^4 < 1$. For this process, the first four moments are $\mu = 0$, $\gamma_0 = E(y_t - \mu)^2 = \sigma^2/(1 - b^2\sigma^2) = 1.333$, $E(y_t - \mu)^3/\gamma_0^{3/2} = 0$, $E(y_t - \mu)^4/\gamma_0^2 = 3(1 - b^4\sigma^4)/(1 - 3b^4\sigma^4) = 3.462$.

Granger and Andersen (1978) give further details for this example. Bera and Higgins (1997) have fitted a bilinear model to stock return data.

Table 3 summarizes the numerical results for the two non-MDS examples. The main conclusion from Table 3 is that the DBOB bootstrap tends to substantially reduce the distortions in the rejection probabilities for both of the non-MDS examples, especially for H_5 and H_{10}. This is despite the fact that the nonlinear moving average model produces massive distortions in the rejection probabilities (percent) of the asymptotic Q_K test: the MAD is about 25 percent at the nominal 1 percent level and 36 percent at 10 percent. The distortions are considerably less for the bilinear model, but they are large nonetheless. The SBOB bootstrap substantially reduces the distortions in the rejection probabilities, but it tends to over-reject, more so for the nonlinear moving average example than for the bilinear example.

The asymptotic Q_K^* test tends to perform satisfactorily for the nonlinear moving average example, except for H_1. For this example, the asymptotic GP_K test tends to under-reject, except for H_1. Turning to the bilinear example, the asymptotic Q_K^* and

Table 3
Rejection probabilities (percent) of tests: non-MDS, $n = 500$

Tests	H_1			H_5			H_{10}		
	1	5	10	1	5	10	1	5	10
Nonlinear moving average case									
Q_K	25.7	38.4	46.3	19.1	30.4	38.6	14.6	24.6	31.8
SBOB Q_K									
$b = 4$	3.6	11.6	19.0	2.2	8.8	16.2	1.5	6.3	12.6
$b = 10$	4.3	12.2	19.6	2.4	9.7	17.7	1.4**	7.1	13.7
$b = 20$	4.6	12.9	20.6	2.9	10.5	18.7	1.7	8.0	15.3
DBOB Q_K									
$b = 4$	1.3*	7.0	13.1	0.9*	5.8**	11.3	0.7*	3.9	8.1
$b = 10$	1.3*	6.3	12.5	1.0*	6.3	12.1	0.9*	4.5*	8.9**
$b = 20$	1.6	7.3	12.9	1.1*	6.2	12.0	1.0*	4.7*	9.7*
Q_K^*	1.3	7.2	14.4	1.1*	5.4**	11.2	1.2**	5.0*	10.1*
GP_K	1.3	7.2	14.4	0.6	4.2	9.7*	0.7	3.6	8.3
$\tilde{Q}_K$	0.6	4.0	9.7*	1.2**	4.4	9.4	2.4	6.1	11.0
Bilinear case									
Q_K	5.7	14.2	21.6	6.2	15.5	23.7	4.6	12.3	19.8
SBOB Q_K									
$b = 4$	2.5	8.9	15.6	1.8	7.6	13.7	1.3*	6.0	11.5
$b = 10$	2.4	8.1	14.4	2.0	7.7	14.2	1.3*	6.1	12.5
$b = 20$	2.5	8.7	15.1	2.1	8.4	15.1	1.3*	6.6	13.8
DBOB Q_K									
$b = 4$	2.0	7.2	13.6	1.3*	6.2	11.4	0.9*	4.4*	9.3*
$b = 10$	1.4**	5.9	11.4	1.1*	5.9	11.5	0.9*	4.6*	9.9*
$b = 20$	1.2*	6.1	11.7	1.4**	6.1	11.9	0.8*	4.9*	10.6*
Q_K^*	2.1	7.7	14.0	2.0	7.4	13.4	1.7	6.6	12.1
GP_K	2.1	7.7	14.0	1.5	6.6	13.1	1.0*	5.1*	10.3*
$\tilde{Q}_K$	1.4	6.3	12.3	1.6	6.8	13.0	1.1*	5.4**	10.9

Notes: See Table 1.

GP_K tests tend to over-reject for H_1 and H_5. However, the GP_K test performs satisfactorily for H_{10}. The asymptotic $\tilde{Q}_K$ test performs satisfactorily for the nonlinear moving average example in some cases, but over-rejects the H_1 and H_5 hypotheses in the bilinear case.

4. Power experiments

In many of our experiments, the empirical rejection probabilities of the Q_K test with DBOB bootstrap-based P-values and the $\tilde{Q}_K$ test with asymptotic P-values were close to the nominal rejection probabilities. Hence, a comparison of the powers of these two tests is empirically relevant. For completeness, the powers of the Q_K^* and GP_K tests with asymptotic P-values are also compared to the powers of DBOB Q_K test.

Table 4
Powers (percent) for the 0.05 nominal $\tilde{Q}_K$, Q_K^*, GP_K and double bootstrap Q_K tests, $n = 500$

		One-dependent case			Nonlinear moving average case		
		H_1	H_5	H_5	H_1	H_5	H_5
ρ	Tests	Alternative					
		$\rho_1 = \rho$	$\rho_1 = \rho,\ \rho_2 = \ldots = \rho_5 = 0$	$\rho_1 = \ldots = \rho_4 = 0,\ \rho_5 = \rho$	$\rho_1 = \rho$	$\rho_1 = \rho,\ \rho_2 = \ldots = \rho_5 = 0$	$\rho_1 = \ldots = \rho_4 = 0,\ \rho_5 = \rho$
$\rho = 0.0$	$\tilde{Q}_K$	4.1	4.1	4.1	4.0	4.4	4.4
	Q_K^*	4.9*	4.5	4.5	7.2	5.4**	5.4**
	GP_K	4.9*	3.9	3.9	7.2	4.2	4.2
	DBOB Q_K						
	$b = 4$	5.8**	5.3*	5.3*	6.8	5.8**	5.8**
	$b = 10$	5.8**	5.6*	5.6*	6.5	6.3	6.3
	$b = 20$	6.0	6.2	6.2	6.4	6.2	6.2
$\rho = 0.05$	$\tilde{Q}_K$	8.5	6.0	9.8	4.0	4.2	11.4
	Q_K^*	9.5	6.3	10.6	7.7	5.5	13.6
	GP_K	9.5	5.4	9.9	7.7	4.8	13.1
	DBOB Q_K						
	$b = 4$	10.6	9.0	8.5	6.3	6.7	7.3
	$b = 10$	10.4	9.9	9.1	5.9	7.1	8.3
	$b = 20$	10.3	9.6	9.1	6.2	6.8	8.4

$\rho = 0.1$	$\tilde{Q}_K$	23.3	12.7	32.0	10.6	6.7	30.2
	Q_K^*	25.0	13.1	33.6	16.7	9.0	38.0
	$\mathrm{GP_K}$	25.0	11.8	33.8	16.7	8.5	39.8
	DBOB Q_K						
	$b = 4$	26.4	23.1	21.7	14.0	13.9	15.5
	$b = 10$	26.2	24.0	23.3	13.2	14.0	16.5
	$b = 20$	25.8	23.4	22.8	13.4	13.5	16.8
$\rho = 0.2$	$\tilde{Q}_K$	74.0	50.7	87.2	44.7	24.4	66.0
	Q_K^*	76.1	48.5	91.8	52.3	27.1	80.2
	$\mathrm{GP_K}$	76.1	47.6	92.5	52.3	29.1	81.9
	DBOB Q_K						
	$b = 4$	75.1	70.2	80.2	51.2	47.7	60.0
	$b = 10$	73.5	70.4	80.4	47.6	47.9	59.2
	$b = 20$	72.3	68.6	77.9	45.9	46.2	56.3

Notes: The number of replications for the Q_K test with BOB bootstrap-based P-values is 5000. The number of replications for the $\tilde{Q}_K$, Q_K^*, GP_K tests is 25,000. The data generation process used for testing H_1 is a first-order moving average process, and the processes used for testing H_5 are a first-order moving average process and a special fifth-order moving average process.

The $\tilde{Q}_K$ test with asymptotic critical values is the natural competitor to the Q_K test with bootstrap critical values. Hence, it is of interest to compare the asymptotic local powers of these two tests. The Q_K test is asymptotically distributed as the weighted sum of independent noncentral chi-square variables, each with one degree of freedom, and the $\tilde{Q}_K$ test is asymptotically distributed as a noncentral chi-square with K degrees of freedom. From this it follows that both tests of H_1 have asymptotically the same power. This is because the Q_1 statistic differs from the $\tilde{Q}_1$ statistic only by a factor of proportionality. By contrast, when testing the null that two or more correlations are zero, neither test dominates the other in terms of local power. Examples can be constructed in which either test has higher local power than the other.

We also computed empirical powers of the tests by conducting Monte Carlo experiments. For a test of H_1, the data generation process is given by the moving average process $w_t = y_t + \theta y_{t-1}$ where y_t is generated by the uncorrelated processes used in Examples 1 and 4 above. The values of θ are chosen so that $\rho(1)$ takes the values 0.05, 0.1 and 0.2. The same setup is used for a test of H_5 where $\rho(1) = \rho,\ \rho(2) = \cdots = \rho(5) = 0$. For a test of H_5 where $\rho(1) = \cdots = \rho(4) = 0, \rho(5) = \rho$ we use $w_t = y_t + \theta y_{t-5}$ where again y_t is generated by the uncorrelated processes in Examples 1 and 4 above and where the values of θ are chosen so that $\rho(5)$ takes the values 0.05, 0.1 and 0.2. The experiments were carried out using 5000 replications with $M_1 = 999$ and $M_2 = 249$ and three block lengths, $b = 4$, 10 and 20.

Table 4 reports the results of experiments where 0.05 is the nominal rejection probability of the tests. Note that for reasons explained by Horowitz and Savin (2000), the P-values of the $\tilde{Q}_K$, Q_K^*, GP_K tests and the DBOB Q_K test are not corrected to be exactly 0.05 under the null. The main features of the results are the following:

(i) As expected from the discussion in the second paragraph of this section, the empirical powers of the DBOB Q_K test are generally similar to those of asymptotic $\tilde{Q}_K$ test for H_1; the empirical powers of the DBOB Q_K test dominate those of the $\tilde{Q}_K$ test when testing H_5 when the $\rho(1)$ is nonzero and $\rho(2) = \cdots = \rho(5) = 0$, and the $\tilde{Q}_K$ test dominates when the $\rho(1) = \cdots = \rho(4) = 0$, and the $\rho(5)$ is nonzero.
(ii) The powers of the asymptotic Q_K^* and GP_K tests are similar to those of the asymptotic $\tilde{Q}_K$ test.
(iii) The powers of the DBOB Q_K test are insensitive to the block length under the alternative as well as under the null hypothesis.

5. Empirical example

This section reports the empirical P-values of tests of H_5 for the example in Lobato et al. (2001). This example uses daily currency returns for the pound sterling. The currency returns are calculated using the noon (New York time) buying rates for the pound sterling (USD/GBP) from the H.10 Federal Reserve Statistical Release. The results are presented for the period January 1993–December 1996 ($n = 1005$). The period runs from the first trading day in January to the last trading day in December.

The empirical P-values for the Q_K, Q_K^*, GP_K and $\tilde{Q}_K$ tests are 0.026, 0.135, 0.130, and 0.104, and the empirical P-values for the DBOB (SBOB) Q_K tests for $b = 4$, 10 and 20 are 0.23 (0.18), 0.24 (0.16), 0.12 (0.10). The results show that the Q_K test rejects H_5 when the nominal rejection probability is 0.05; by contrast, this null hypothesis is accepted by the Q_K^*, GP_K and $\tilde{Q}_K$ tests and by the DBOB Q_K tests.

6. Discussion

The starting point for this study is the proposal by Romano and Thombs (1996) to use the bootstrap to make inferences about the individual autocorrelation coefficients. In this paper, the null hypothesis of uncorrelatedness is tested using the Q_K test with bootstrap-based P-values. The bootstrap was implemented using both a single and a double blocks-of-blocks (BOB) procedure with prewhitening. Monte Carlo experiments were conducted to investigate the true rejection probability of the Q_K test with BOB bootstrap-based P-values. The examples used in the experiments included three MDS processes and two non-MDS processes.

The main Monte Carlo findings for experiments under the null hypothesis are threefold. First, for samples of size 500, there were large distortions in the empirical rejection probabilities when the Q_K test was based on asymptotic P-values. Second, for martingale difference sequences, the double BOB bootstrap essentially eliminates the distortions in the empirical rejection probabilities that are present when the Q_K test is based on the asymptotic P-values. For non-martingale difference sequences, the double BOB bootstrap does not entirely eliminate the distortions, but the distortions are much reduced. Third, the results tend to be robust to the choice of the block length.

We conducted a Monte Carlo investigation of the Q_K^*, GP_K and $\tilde{Q}_K$ tests for uncorrelatedness. These first two tests are designed for the case where the time series is generated by a MDS process, and the last is asymptotically valid for both MDS and non-MDS processes. Roughly speaking, the Q_K^*, GP_K and $\tilde{Q}_K$ tests performed similarly when the null hypothesis is true. Finally, we investigated the power of the DBOB Q_K test against the power of the $\tilde{Q}_K$ test. Among the tests considered here, these are the only two that do not make the MDS assumption. The empirical powers of the Q_K test with bootstrap P-values were similar to those of the $\tilde{Q}_K$ test with asymptotic P-values when testing H_1. This is not surprising because the two tests are asymptotically equivalent. Neither test was superior to the other when testing H_5.

In conclusion, the DBOB Q_K test and the $\tilde{Q}_K$ test appear to provide satisfactory finite sample performance when the data generation process is stationary but is not assumed to be linear or a martingale difference sequence.

Acknowledgements

We thank Paul Weller, the seminar participants at Monash University and participants at the Cowles Foundation Conference "New Developments in Time Series Econometrics" for helpful suggestions and comments. The research of Joel L. Horowitz was supported in part by NSF Grant SES-9910925. I. N. Lobato acknowledges financial support from the Asociación Mexicana de Cultura and CONACYT Grant no. 41893-S. John Nankervis gratefully acknowledges financial support from the ESRC through Research Grant no. R000222581.

Appendix A. Consistency of double bootstrap procedure

The proof of consistency begins by considering a function of sample moments of a general vector time series, $\{X_t\}$. The results are then specialized to the Box–Pierce statistic.

A.1. Assumptions

Let $\{X_t : -\infty < t < \infty\} \subset \mathbb{R}^K$ be a time series. Let ℓ_1 and ℓ_2, respectively, be the lengths of the blocks in the first and second bootstrap stages. Let $\{X_t : t = 1, ..., T\}$ be a realization of $\{X_t\}$. Make the following assumptions:

A1. $\{X_t\}$ is strictly stationary and strongly mixing.
A2. For some $\delta > 0$ and $\Delta < \infty$: (a) $|X_t|^{2+2\delta} \leqslant \Delta$ for all t. (b) The mixing parameter $\alpha(\cdot)$ satisfies $\sum_{t=1}^{\infty}(t+1)^2\alpha(t)^{\delta/(4+\delta)} \leqslant \Delta$.
A3. As $T \to \infty$, $\ell_1 \to \infty$, $\ell_2 \to \infty$, $\ell_1/T \to 0$, and $\ell_2/\ell_1 \to 0$.

Our assumptions are not necessarily the weakest possible.

A.2. Preliminaries

Let $\mu = \mathrm{E}(X_1)$ and $\bar{X} = T^{-1}\sum_{t=1}^{T}(X_t - \mu)$. For $j = 1, 2, \ldots$ define the $K \times K$ matrix $\gamma_j = E(X_1 - \mu)(X_{1+j} - \mu)'$. Then under A1–A2

$$\lim_{T\to\infty} TE(\bar{X}\bar{X}') = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j \equiv \Omega.$$

Moreover, it follows from a central limit theorem (e.g. Theorem A.1 of Politis et al., 1997, hereinafter PRW) that $T^{1/2}\bar{X} \to^{\mathrm{d}} \mathrm{N}(0, \Omega)$.

The bth block $(b = 1, ..., T - \ell_1 + 1)$ of the first-stage bootstrap is $W_{1b} = (X_b, X_{b+1}, ..., X_{b+\ell_1-1})$. To avoid notational complications that provide no insight, assume that T/ℓ_1 is an integer. Let $\{X_t^* : t = 1, \ldots, T\}$ be a first-stage bootstrap sample that is obtained by sampling the blocks W_{1b} randomly with replacement T/ℓ_1 times and laying the sampled blocks end to end. Let $\boldsymbol{P}^*$ denote the probability measure induced by this sampling process. Define

$$m^* = T^{-1}\sum_{t=1}^{T}(X_t^* - \bar{X}).$$

Let $\Phi_K(\cdot, \Omega)$ denote the CDF of the K-variate normal distribution $\mathrm{N}(0, \Omega)$. The first-stage bootstrap estimator of the CDF of $T^{1/2}m^*$ is consistent if for every $\varepsilon > 0$

$$\lim_{T\to\infty} P\left[\sup_{x\in\mathbb{R}^K} |P^*(T^{1/2}m^* < x) - \Phi_k(x, \Omega)| > \varepsilon\right] = 0. \tag{1}$$

Now consider the second-stage bootstrap. The bth block $(b = 1, ..., T - \ell_2 + 1)$ is $W_{2b} = (X_b^*, X_{b+1}^*, ..., X_{b+\ell_2-1}^*)$. Avoid unimportant notational complications by assuming that T/ℓ_2 is an integer. Let $\{X_t^{**} : t = 1, \ldots, T\}$ be a second-stage bootstrap sample that is obtained by sampling the blocks W_{2b} randomly with replacement T/ℓ_2 times and laying the sampled blocks end to end. Let $\boldsymbol{P}^{**}$ denote the probability measure induced by this sampling process. Define

$$m^{**} = T^{-1}\sum_{t=1}^{T}\left(X_t^{**} - \bar{X}^*\right),$$

where $\bar{X}^* = T^{-1}\sum_{t=1}^{T} X_t^*$.

Now let $G : \mathbb{R}^K \to \mathbb{R}^K$ be a continuously differentiable function. Let $\tau \equiv TG(\bar{X})'G(\bar{X})$ be a test statistic with a non-degenerate limiting distribution. Assume that the null hypothesis, H_0, is rejected when τ is large. Set $\tau^* = TG(m^*)'G(m^*)$ and $\tau^{**} = TG(m^{**})'G(m^{**})$. Then the double-bootstrap test based on τ rejects a correct H_0 with asymptotic probability α if (i) the single bootstrap is consistent and (ii) for any $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$

$$\lim_{T\to\infty} \boldsymbol{P}\left\{\boldsymbol{P}^*\left[\sup_{x\ \in\mathbb{R}^K} |\boldsymbol{P}^{**}(T^{1/2}m^{**} \leqslant x) - \boldsymbol{P}^*(T^{1/2}m^* \leqslant x)|\ > \varepsilon_1\right] > \varepsilon_2\right\} = 0. \tag{2}$$

A.3. First-stage bootstrap

This section shows that (1) holds under assumptions A1–A3. Define

$$\bar{Z}_b^* = \ell_1^{-1/2}\sum_{t=1}^{\ell_1-1}(X_{b+t}^* - \bar{X}); \quad b = 1, \ldots, T - \ell_1 + 1$$

Then

$$T^{1/2}m^* = (\ell_1/T)^{1/2}\sum_{b=1}^{T/\ell_1}\bar{Z}_b^*.$$

By the Cramér–Wold device, it suffices to show that for any constant K-vector λ with $\lambda'\lambda = 1$, $T^{1/2}\lambda' m^* \to^{\mathrm{d}} \mathrm{N}(0, \lambda'\Omega\lambda)$ with probability approaching 1 as $T \to \infty$. The variables $\bar{Z}_b^*$ are *iid* and, by Theorem A.1 of PRW, their CDF converges in probability to that of the $\mathrm{N}(0, \lambda'\Omega\lambda)$ distribution. Therefore, by Proposition B.1 of PRW, consistency of the first-stage bootstrap follows if $V^* \equiv Var^*\left(T^{1/2}m^*\right) \rightarrow^{\mathrm{p}} \lambda'\Omega\lambda$. But

$$V^* = \frac{1}{T - \ell_1 + 1} \sum_{b=1}^{T-\ell_1+1} (\lambda'\bar{Z}_b)^2 - \left(\frac{1}{T - \ell_1 + 1} \sum_{b=1}^{T-\ell_1+1} \lambda'\bar{Z}_b \right)^2, \tag{3}$$

where

$$\bar{Z}_b = \ell_1^{-1/2} \sum_{t=0}^{\ell_1 - 1} (X_{b+t} - \bar{X}); \quad b = 1, 2, ..., T - \ell_1 + 1.$$

It follows from Propositions 3.34 and 3.44 of White (2001) that the first term on the right-hand side of (3) converges almost surely to $\lambda'\Omega\lambda$ and the second term converges almost surely to 0. This establishes consistency of the first-stage bootstrap.

A.4. Second-stage bootstrap

This section shows that (2) holds under assumptions A1–A3. Define

$$\bar{Z}_b^{**} = \ell_2^{-1/2} \sum_{t=1}^{\ell_2 - 1} (X_{b+t}^{**} - \bar{X}^*), \quad b = 1, \ldots, T - \ell_2 + 1.$$

Then

$$T^{1/2}\lambda' m^{**} = (\ell_2/T)^{1/2} \sum_{b=1}^{T/\ell_2} \lambda'\bar{Z}_b^{**}.$$

Because $\ell_2/\ell_1 \to 0$, the probability that a second-stage block overlaps a boundary between two first-stage blocks approaches 0 as $T \to \infty$. Therefore, the $\lambda'\bar{Z}_b^{**}$ can be treated asymptotically as *iid* random variables that are completely contained within first-stage blocks. Accordingly, it follows from Theorem A.1 of PRW that the CDF of $\lambda'\bar{Z}_b^{**}$ converges in $\boldsymbol{P}^*$ probability to that of $\mathrm{N}(0, \lambda'\Omega\lambda)$ except, possibly, if the first-stage sample belongs to a set whose $\boldsymbol{P}$ probability converges to 0 as $T \to \infty$. Moreover, arguments like those applied to (3) show that $Var^{**}\left(T^{1/2}m^{**}\right) \rightarrow^{p*} \lambda'\Omega\lambda$ except, possibly, if the first-stage sample belongs to a set whose $\boldsymbol{P}$ probability converges to 0 as $T \to \infty$. This establishes (2).

A.5. The Box– Pierce statistic

To connect the foregoing results with the Box–Pierce statistic, define

$$X_t = (Y_t, Y_t^2, Y_t Y_{t+1}, ..., Y_t Y_{t+K})' \tag{4}$$

Let $\bar{X}_j$ denote the jth component of $\bar{X}$ ($j = 1, \ldots, K+2$). For $j = 1, \ldots, K$, define $g_j(\bar{X}) = (\bar{X}_{j+2} - \bar{X}_1^2)/(\bar{X}_2 - \bar{X}_1^2)$,

$$G(\bar{X}) = [g_1(\bar{X}), ..., g_k(\bar{X})]' \tag{5}$$

and $\tau = TG(\bar{X})'G(\bar{X})$. Then τ is a version of the Box–Pierce statistic and is asymptotically equivalent to Q_K. Define τ^* and τ^{**} as before with X_t and G as in (4) and (5). Because the edge effects that are corrected by w_t are asymptotically negligible, τ^* is $\boldsymbol{P}^*$-asymptotically equivalent to Q_K^S, and τ^{**} is $\boldsymbol{P}^{**}$-asymptotically equivalent to Q_K^D. Therefore, consistency of the Q_K^D test follows from the results of Section 4.

Appendix B. Stopping rules

The stopping rules are implemented by first doing the M_1 single bootstrap calculations, saving all single bootstrap samples. The single bootstrap P-value, p_K^* is then calculated. We then do a maximum of M_1 sets of double bootstrap replications where each set corresponds to one of the M_1 single bootstrap samples. In each of these sets we do a maximum of M_2 double bootstrap replications. *Stopping Rule* 1: If $p_K^* = 1$ then $p_{Ka}^* = 1$ and no double bootstrap calculations are required. *Stopping Rule* 2: The adjusted P-value is calculated as $\#(p_{K^{**}} \leqslant p_{K^*})/M_1 = \#(\sum_{i=1}^{M_2} I(Q_{Ki}^D > Q_K^S)/M_2 \leqslant p_{K^*})/M_1$ where I(·) is the indicator function equaling 1 if the expression inside the brackets is true and zero otherwise. We stop after m_2 replications if $\sum_{i=1}^{m_2} I(Q_{Ki}^D > Q_K^S)$ either exceeds $M_2 p_K^*$ or cannot exceed $M_2 p_K^*$ in the remaining $M_2 - m_2$ double bootstrap replications for each single bootstrap sample. *Stopping Rule* 3: For a maximum nominal level of 0.1, we stop doing double bootstrap replications if p_{Ka}^* must exceed 0.1; i.e. stop after m_1 sets of double bootstrap replications if $\sum_{i=1}^{m_1} I(p_{Ki^{**}} \leqslant p_{K^*})$ exceeds $0.1\, M_1$. The effectiveness of Stopping Rule 3 is enhanced by exploiting the negative correlation between p_K^{**} and Q_K^S and doing the calculations for the sets of double bootstrap replications in an order corresponding to decreasing size of Q_K^S. The combined effect of these rules is that we require only from $M_1 M_2/15$ to $M_1 M_2/11$ double bootstrap replications in our experiments under both the null and alternative hypotheses.

References

Andrews, D.W.K., 2002. Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. Econometrica 70, 119–162.

Andrews, D.W.K., Ploberger, W., 1996. Testing for serial correlation against an ARMA (1, 1) process. Journal of the American Statistical Association 91, 1331–1342.

Bera, A.K., Higgins, M.L., 1997. ARCH and bilinearity as competing models for nonlinear dependence. Journal of Business and Economic Statistics 15, 43–51.

Beran, R., 1988. Prepivoting test statistics: a bootstrap view of asymptotic refinements. Journal of the American Statistical Association 83, 687–697.

Box, G.E.P., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. Journal of the American Statistical Association 65, 1509–1526.

Carlstein, E., 1986. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. Annals of Statistics 14, 1171–1179.
Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge University Press, Cambridge.
Diebold, F.X., 1986. Testing for serial correlation in the presence of heteroskedasticity. Proceedings of the American Statistical Association, Business and Economics Statistics Section, pp. 323-328.
El Babsiri, M., Zakoian, J.-M., 2001. Contemporaneous asymmetry in GARCH processes. Journal of Econometrics 101, 257–294.
Goncalves, S., Kilian, L., 2003. Bootstrapping autogressions with conditional heteroskedasticity of unknown form. Unpublished, University of Montreal.
Granger, C.W.J., Andersen, A.P., 1978. An Introduction to Bilinear Time Series Models. Vanenhoek and Ruprecht, Gottingen.
Granger, C.W.J., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. University of Oxford Press, Oxford.
Guo, B.B., Phillips, P.C.B., 1998. Testing for autocorrelation and unit roots in the presence of conditional heteroskedasticity of unknown form. Unpublished manuscript, Yale University, Cowles Foundation for Research in Economics.
Hall, P., Horowitz, J.L., 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64, 891–916.
Hall, P., Horowitz, J.L., Jing, B., 1995. On blocking rules for the bootstrap with dependent data. Biometrika 82, 561–574.
Hamilton, K.G., James, F., 1997. Acceleration of RANLUX. Computer Physics Communications 101, 241–248.
Hannan, E.J., Heyde, C.C., 1972. On limit theorems for quadratic functions of discrete time series. Annals of Mathematical Statistics 43, 2058–2066.
He, C., Teräsvirta, T., 1999. Properties of moments of a family of GARCH processes. Journal of Econometrics 92, 173–192.
Hinkley, D.V., 1989. Bootstrap significance tests, Proceedings of the 47th Session of the International Statistical Institute, vol. VIII, pp. 65-74.
Horowitz, J.L., 2003. Bootstrap methods for Markov processes. Econometrica 71, 1042–1092.
Horowitz, J.L., Savin, N.E., 2000. Empirically relevant critical values for hypotheses tests: a bootstrap approach. Journal of Econometrics 95, 375–389.
Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Annals of Statistics 17, 1217–1241.
Lahiri, S.N., 1999. Theoretical comparisons of block bootstrap methods. Annals of Statistics 27, 386–404.
Lo, A.W., MacKinlay, A.C., 1989. The size and power of the variance ratio test in finite samples: a Monte Carlo investigation. Journal of Econometrics 40, 203–238.
Lobato, I., Nankervis, J.C., Savin, N.E., 2001. Testing for autocorrelation using a modified Box–Pierce Q test. International Economic Review 42, 187–205.
Lobato, I., Nankervis, J.C., Savin, N.E., 2002. Testing for zero autocorrelation in the presence of statistical dependence. Econometric Theory, 730–743.
Nankervis, J.C., 2002. Stopping rules for double bootstrap tests. Unpublished manuscript, Department of Economics, University of Surrey.
Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.
Politis, D.N., Romano, J.P., Wolf, M., 1997. Subsampling for heteroskedastic time Series. Journal of Econometrics 81, 281–318.
Romano, J.L., Thombs, L.A., 1996. Inference for autocorrelations under weak Assumptions. Journal of American Statistical Association 91, 590–600.
Romano, J.P., Wolf, M., 2000. Improved nonparametric confidence intervals in time series regressions, Technical Report No. 2000-39, Department of Statistics. Stanford University.
Tong, H., 1990. Nonlinear Time Series. Oxford University Press, Oxford.
White, H., 2001. Asymptotic Theory for Econometricians, second ed. Academic Press, New York.