Bootstrapping the Box-Pierce Q test: A Robust Test of Uncorrelatedness

Joel L. Horowitz

Department of Economics, Northwestern University, Evanston, IL 60208-2600

I. N. Lobato

Centro de Investigación Económica, ITAM, México 10700.

John C. Nankervis

Department of Economics, University of Surrey, Guildford, GU2 5XH U.K.

N. E. Savin

Department of Economics, University of Iowa, Iowa City, IA 52242.

November 25, 2001

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 distortions in the rejection probabilities are essentially eliminated.

KEY WORDS: Serial correlation tests; Box-Pierce Q; blocks of blocks bootstrap, adjusted P-values, double bootstrap.

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, Horowitz and Jing (1995), Hall and Horowitz (1996), Andrews

(2001)). 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. At present, a Monte Carlo study with a studentized version is very time consuming. 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 of Politis and Romano (1992). One reason for choosing this method is based on Monte Carlo evidence reported by Davison and Hinkley (1997, Table 8.2) for time series data. This evidence suggests that the BOB bootstrap is less sensitive to the choice of block length than are alternative blocking methods such as the moving block bootstrap proposed by Künsch (1989). Intuitively speaking, this advantage of the BOB bootstrap is due to the fact that it reduces the influence of edge effects produced by blocking. The BOB bootstrap is, however, a modified version of the moving block bootstrap. Hall, Horowitz and Jing (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). This provides a second motivation for the use of the BOB bootstrap instead of non-overlapping blocks or the stationary bootstrap. Our Monte Carlo experiments results confirm that the BOB bootstrap is rather insensitive to the choice of block length. Further, the difference between the true and nominal probabilities that a test rejects a correct null hypothesis (error in the rejection probability or ERP) is typically much smaller when the P-value of the Q_K test is based on the BOB bootstrap than when the P-value is based on the chi-square P-value.

Beran (1988) gives conditions under which iterating the bootstrap can produce further reductions in the 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).

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 empirical evidence that block bootstrap iteration reduces the finite-sample ERP of a test motivates us to carry out experiments with the iterated blocks-of-blocks bootstrap (double blocks-of-blocks or DBOB bootstrap). We find that the ERPs are usually lower with the DBOB bootstrap than with the non-iterated BOB bootstrap (single blocks-of-blocks or SBOB bootstrap).

The Markov conditional bootstrap (Horowitz, 2001) 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.

This paper investigates the numerical performance of the Q_K test when the P-value is obtained using the SBOB bootstrap and the DBOB bootstrap. We refer to tests that use SBOB and DBOB bootstrap P-values as SBOB and DBOB bootstrap tests, respectively. In the Monte Carlo experiments, the data are generated by stochastic processes that are martingale difference sequences (MDS's) as well as non-MDS processes. The MDS processes considered are a model used 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). A nonlinear moving average model and a bilinear model are used to generate the non-MDS processes.

Finally, the performance of the bootstrapped Q_K test is compared to that of other tests of uncorrelatedness. The other tests are the Q_K* test (Diebold (1986), Lo and MacKinlay (1989) and Lobato, Nankervis and Savin (2001a)), the GP_K test (Guo and Phillips (1998)), and the $\tilde{Q}_K$ test (Lobato, Nankervis and Savin (2001b)). 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 $\tilde{Q}_K$ 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; they can be viewed as extensions of the Q_K test.

For the examples used in the experiments, our results show that 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.

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 $\widetilde{Q}_K$ 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 $\widetilde{Q}_K$ test based on asymptotic P-values. Concluding comments are in Section 5. Some technical computational issues are addressed in the Appendix.

2. Bootstrap Test

The bootstrap provides a first-order asymptotic approximation to the distribution of the Q_K test statistic under the null hypothesis. Thus, 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 bootstraps. The second objective is to describe the prewhitening procedure employed.

Preliminaries. Let $y_1, \ldots, y_n$, denote a real-valued strictly and covariance 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 $n^{½} r = n^{½} [r(1),\ldots, r(K)]'$ is asymptotically normally distributed with asymptotic covariance matrix V, where the ij-th element of V is given by

$$v_{ij} = \gamma(0)^{-2}[c_{i+1,j+1} - \rho(i)c_{1,j+1} - \rho(j)c_{1,i+1} + \rho(i)\rho(j)c_{1,1}] \tag{1}$$

and

$$\begin{aligned} c_{i+1,j+1} = \sum_{d=-\infty}^{d=\infty} \{ & E(y_t - \mu)(y_{t-i} - \mu)(y_{t+d} - \mu)(y_{t+d-j} - \mu) - \\ & E(y_t - \mu)(y_{t-i} - \mu)E(y_{t+d} - \mu)(y_{t+d-j} - \mu)\};\ i, j = 0,1,\ldots,\ K; \end{aligned} \tag{2}$$

see 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-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 $\widetilde{Q}_K$ test replaces V with an estimator that is consistent under the null for both MDS and non-MDS processes; for details, see Lobato, Nankervis and Savin (2001b).

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 t. For any fixed number z, let S(z) = $P(Q_K > z \mid H_K)$. The P-value associated with t is p = S(t). The exact symmetric test of H_K rejects at level α if $p \equiv S(t) < \alpha$. 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.

Single Bootstrap. In order to implement the BOB bootstrap, we define a new (K+1)×(n-k) data matrix as $(Y_1, Y_2, \ldots, Y_{n-K})$ where $Y_i = (y_i, y_{i+1}, \ldots, y_{i+K})'$. For the lag K autocorrelation, for example,

$$\begin{pmatrix} Y_1 & Y_2 & \cdots & Y_{n\text{-}K} \end{pmatrix} = \begin{pmatrix} y_1 & y_2 & \cdots & y_{n\text{-}K} \\ & & \cdots & \\ y_{K+1} & y_{K+2} & \cdots & y_n \end{pmatrix} = \begin{pmatrix} y_1^1 & y_2^1 & \cdots & y_{n\text{-}K}^1 \\ & & \cdots & \\ y_1^{K+1} & y_2^{K+1} & \cdots & y_{n\text{-}K}^{K+1} \end{pmatrix}.$$

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, where n = hb. Let B_i be a (K+1)×b matrix given by $B_i = Y_i, \ldots, Y_{i+b-1}$, where i = 1, …, q, and q = n-b-K+1. The SBOB bootstrap test is obtained by 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)×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

$$Q_K{}^S = n\sum_{k=1}^{K}[r^*(k) - r_b(k)]^2 \quad \text{where } r_b(k) \text{ is defined below,}$$

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

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 $r_b(k) = \mathrm{Cov}^*(1, k+1)/[V^*(1)\ V^*(k+1)]^{1/2}$. The terms in the expression for $r_b(k)$ are defined as follows: $\mathrm{Cov}^*(1,k+1) = E^*\overline{(y^1y^{k+1})} - E^*\overline{(y^1)}E^*\overline{(y^{k+1})}$, $V^*(k) = \mathrm{Cov}^*(k,k)$ where E^* denotes the expectation relative to the empirical distribution of the data. The formulae for these expectations are the following:

$$E^*(\overline{y^k}) = \overline{y^k} + \frac{1}{b(n\text{-}K\text{-}b+1)}\left(b(b\text{-}1)\overline{y^k} - \sum_{j=1}^{b-1}(b\text{-}j)(y_j^k + y_{n\text{-}K\text{-}j+1}^k) \right) \text{ with } \overline{y^k} = \frac{1}{n\text{-}K}\sum_{j=1}^{n\text{-}K} y_j^k ,$$

and

$$E^*\left(\overline{y^1y^{k+1}}\right) = \overline{y^1y^{k+1}} + \frac{1}{b(n\text{-}K\text{-}b+1)}\left(b(b\text{-}1)\overline{y^1y^{k+1}} - \sum_{j=1}^{b-1}(b\text{-}j)(y_j^1y_j^{k+1} + y_{n\text{-}K\text{-}j+1}^1 y_{n\text{-}K\text{-}j+1}^{k+1}) \right)$$

with $\overline{y^1y^{k+1}} = \frac{1}{n\text{-}K}\sum_{j=1}^{n\text{-}K} y_j^1 y_j^{k+1}$.

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 $p_K^* = \#(Q_K{}^S > Q_K)/ M_1$. Given a nominal level of α, the SBOB bootstrap test of H_K rejects if $p_K^* < \alpha$.

The bootstrap test based on p_K* has rejection probability α if $P(p_K^* < \alpha | H_K) = \alpha$, 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 $P(p_K^* < \beta | H_K) = F_{p*}(\beta) = \alpha$. 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 β.

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, …,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 $\{B_1^*, \ldots, B_q^*\}$. 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 $Y_i^{**} = (y_i^{1**}, \ldots, y_i^{(K+1)**})'$.

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

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

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

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 $p_{K\alpha}^* = F_{p**}(p_K^*) < \alpha$ where $p_{K\alpha}^*$ is what Davison and Hinkley (1997) call the *adjusted P-value*. The adjusted P-value is estimated by $\#[\, p_K^{**} \le p_K^* \,]/M_1$; this formula is also given by Hinkley (1989).

Davison and Hinkley (1997) strongly recommend the use of adjusted P-values. Politis, Romano and Wolf (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.

Pre-whitening. We investigated the performance of the bootstrap with and without pre-whitening of the data series. The details of the pre-whitening procedure are described below. The rationale for pre-whitening 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 pre-whitening 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, pre-whitening increases the ERP. In Section 4 we present only the results of the experiments with pre-whitening. The results of experiments without pre-whitening are available upon request.

The pre-whitening 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)×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)×n bootstrap sample $E^* = (E_1^*,\ldots,E_n^*)$. This bootstrap sample is used in place of Y* in steps 1 to 3 in calculating the SBOB bootstrap test.

For DBOB we first prewhiten the SBOB sample by fitting an AR(K) to the (K+1) × n elements of E* (or equivalently using appropriate weighting of repeated elements). In order to prewhiten DBOB we consider an augmented (2K+1) × 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)×b blocks are resampled from the matrix U* to form the (K+1)×n matrix U**. This bootstrap sample is used in place of Y** in steps 1′ to 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.

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 the Appendix. For further details see Nankervis (2001).

The tables in this section report the empirical rejection probabilities of bootstrap tests of H_K: $\rho(1) = \ldots = \rho(K) = 0$, K = 1, 5, 10, for samples of n = 500. The empirical rejection probabilities for the bootstrap tests are calculated using 5,000 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 $\widetilde{Q}_K$ test is implemented using the VARHAC procedure described in Lobato, Nankervis and Savin (2001b).

The random number generator used in the experiments was the very long period generator RANLUX with luxury level p = 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. Diagonal Case. Let $y_t = z_t \bullet z_{t-1}$ where $\{z_t\}$ is a sequence of iid N(0,1) random variables. 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 is about 0.12 when the nominal rejection probability is 0.01 and 0.23 when it is 0.10.

(ii) The DBOB bootstrap eliminates the distortions in the rejection probabilities for the first two hypotheses: the MAD is about 0.003 at 0.01 and 0.011 at 0.10. For the third hypothesis the distortions are almost eliminated: MAD is about 0.003 at 0.01 and 0.017 at 0.10

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

(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 $\widetilde{Q}_K$ test under-rejects for all three hypotheses: the MAD is 0.005 at 0.01 and 0.014 at 0.10. 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. 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 $\beta^2 + 2\alpha_1\beta\nu_2 + \alpha_1\nu_4 < 1$ where $\nu_i = E|z_t|^i$. 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 $\beta = 0.90$. 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, $\gamma_0 = E(y_t - \mu)^2 = 0.1$, $E(y_t - \mu)^3/\gamma_0^{3/2} = 0$, $E(y_t - \mu)^4/\gamma_0^2 = 4.5$, 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 $v_{11} = 6.303$. Lobato, Nankervis and Savin (2001a) have also used this example.

Example 3. Chi-square (3) GARCH. 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 Tersäsvirta (1999) condition is also satisfied when z_t is a chi-square (3) random variable. In this case (the skewness is an estimate), $\gamma_0 = E(y_t - \mu)^2 = 0.1$, $E(y_t - \mu)^3 / \gamma_0^{3/2} = 1.85$, $E(y_t - \mu)^4 / \gamma_0^{2} = 10.9$ 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 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 for 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 $\widetilde{Q}_K$ test works satisfactorily for H_1 and H_5 for Gaussian GARCH and for all three hypotheses for GARCH with chi-square (3) errors.

Non-MDS Examples

The first uncorrelated non-MDS process is generated by an 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 r(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 $N(0, \sigma^2)$ 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 of the asymptotic Q_K test: the MAD is about 0.25 at 0.01 and 0.36 at 0.10. 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 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. We studied the power properties of these tests by conducting Monte Carlo experiments where the data generation process was given by the moving average process $w_t = y_t + \theta y_{t-1}$ and where y_t was generated by the uncorrelated processes used in Examples 1 and 4 above. The value of θ was chosen so that the lag-1 autocorrelations took the values 0.05, 0.1, 0.2 and 0.3. 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 experiment 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) Under the alternative hypothesis, the powers of the DBOB Q_K test are generally similar to those of asymptotic $\tilde{Q}_K$ test for H_1 but are often substantially larger for H_5 and H_{10}.

(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 much less sensitive to K than the powers of the asymptotic $\tilde{Q}_K$ test.

(iv) 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. 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 double blocks-of-blocks procedure with prewhitening. Monte Carlo experiments were conducted to investigate the true rejection probability of the Q_K test with block-of-blocks 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 blocks-of-blocks 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 blocks-of-blocks 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. On the basis of this evidence, we recommend using the double blocks-of-blocks bootstrap procedure with prewhitening.

We conducted a Monte Carlo investigation of the asymptotic 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 asymptotic Q_K*, GP_K and $\tilde{Q}_K$ tests performed similarly when the null hypothesis is true. This said, the GP_K test tended to under-reject compared to the Q_K* and $\tilde{Q}_K$ tests.

Finally, we investigated the power of the Q_K test with bootstrap-based P-values against the power of the $\tilde{Q}_K$ test. Because the $\tilde{Q}_K$ test is asymptotically valid for both MDS and non-MDS process, it is a natural competitor to the Q_K test with bootstrap-based P-values. The empirical powers of the Q_K test with bootstrap P-values were similar to or better than those of the $\tilde{Q}_K$ test. In particular, the empirical powers of the Q_K test with bootstrap P-values were substantially higher than those of the $\tilde{Q}_K$ test when both the number of autocorrelations being tested was large ($K \geq 5$) and the autocorrelations under the alternative were substantially different from zero. The poor power of the $\tilde{Q}_K$ test may be explained by the imprecision of the HAC estimator of the covariance matrix of the sample autocorrelations.

Our recommendation is subject to qualification that the performance of the bootstrap is sensitive to the kurtosis of the time series process. We have chosen examples for which the kurtosis is moderate, but relevant for economics and financial time series. It is easy to construct examples where the kurtosis is several orders of magnitude larger than in our examples. As is well known, high kurtosis can cause the bootstrap test to perform poorly.

A further qualification is that we have not attempted a detailed investigation of Andrews' and Ploberger's (1996, hereinafter AP's) tests of the hypothesis that a series is iid against the alternative that it is ARMA(1,1). AP present Monte Carlo evidence indicating that their tests are more powerful than the Q_K test when the alternative model is, in fact, ARMA(1,1). However, AP obtained critical values for their tests only under the null hypothesis of an iid series. We carried out Monte Carlo experiments using the series $y_t = z_t \bullet z_{t-1}$, $z_t \sim N(0,1)$, which is uncorrelated but not iid. We found that when AP's 0.05-level critical values are used and the sample size is either 500 or 20,000, AP's tests reject the hypothesis $\rho(1) = 0$ with probability exceeding 0.20. Thus, AP's tests with AP's critical values are not reliable procedures for testing the hypothesis that a series is uncorrelated though not necessarily serially independent. It may be possible to overcome this problem by using the bootstrap to obtain critical values for AP's tests. Doing this, however, presents theoretical and computational challenges whose solution is beyond the scope of this paper. The investigation of the properties of AP's tests with bootstrap critical values is left for future research.

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. John Nankervis gratefully acknowledges financial support from the ESRC through Research Grant no. R000222581.

Appendix: Bootstrap Stopping Rules

To reduce the computation time for the double bootstrap tests we use a number of stopping rules. These stopping rules are implemented by first doing the M_1 single bootstrap calculations, saving all single bootstrap samples, estimated coefficients and test statistics. The single bootstrap P-values, p_K^* are 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 $p_K^* = 1$ for any K then $p_{Ka}^* = \#(p_K^{**} \leq p_K^*)/M_1 = 1$ and there is no need for double bootstrap calculations. This occurs about N/M_1 times in every N Monte Carlo experiments where the null hypothesis is true.

Stopping Rule 2: The adjusted P-value is calculated as $\#(p_K^{**} \leq p_K^*)/M_1 = \#(\ \#(Q_K^D$

$>Q_K^S)/M_2 \leq p_K*)/M_1$. We can express $\#(Q_K^D > Q_K^S) \leq M_2 p_K*$ as $\sum_{i=1}^{M_2} I(Q_{K_i}^{\ D} > Q_K^{\ S}) \leq M_2 p_K*$.

We avoid unnecessary replications by stopping after m_2 replications if $\sum_{i=1}^{m_2} I(Q_{K_i}^{\ D} > Q_K^{\ S})$ 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. Under the null, this has the effect of reducing the number of double bootstrap replications by approximately one half in our experiments.

Stopping Rule 3: Since we report rejection probabilities for a maximum nominal level of 0.1, we stop doing double bootstrap replications if the adjusted P-value must exceed 0.1; i.e. stop after m_1 sets of double bootstrap replications if $\sum_{i=1}^{m_1} I(p_{K_i}** \leq p_K*)$ exceeds $0.1M_1$. Under the null, this has the effect of requiring only about $M_1/3$ sets of double bootstrap replications.

The effectiveness of Stopping Rule 3 is enhanced by doing the calculations for the sets of double bootstrap replications in an order corresponding to decreasing size of Q_K^S. The purpose of this ordering is to exploit the negative correlation between p_K** and Q_K^S so that $\sum_{i=1}^{m_1} I(p_{K_i}** \leq p_K*)$ more quickly exceeds the limit $0.1M_1$ if this limit is to be exceeded. In our experiments this re-ordering and Stopping Rule 3 had the combined effect of requiring only about $M_1/6$ sets of double bootstrap replications when the null hypothesis was true.

Nankervis (2001) finds that using the above stopping rules give rise to similar computational savings in the case where the null hypothesis is not true. The combined effect of all these rules is that we require only from $M_1M_2/15$ to $M_1M_2/11$ double bootstrap replications in our experiments.

References

Andrews, D. W. K., 2001. Higher-order improvements of a computationally attractive k-step boostsrap for extremum estimators. *Econometrica*, forthcoming.

Andrews, D.W.K. and W. Ploberger, 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., and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge, U.K.

Diebold, F. X., 1986. Testing for serial correlation in the presence of heteroskedasticity. *Proceedings of the American Statistical Association, Business and Economics Statistics Section,* 323-328.

El Babsiri, M. and Zakoian, J.-M., (2001). Contemporaneous asymmetry in GARCH processes. *Journal of Econometrics*, 101, 257-294.

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, VIII, 65-74.

Horowitz, J. L., 2001. Bootstrap methods for markov processes. Unpublished manuscript, University of Iowa, Department of Economics.

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.

Lobato, I., Nankervis J.C., Savin N. E., 2001a. Testing for autocorrelation using a modified Box-Pierce Q test," *International Economic Review* 42, 187-205.

Lobato, I., Nankervis J.C., Savin N. E., 2001b. Testing for zero autocorrelation in the presence of statistical dependence." Forthcoming, *Econometric Theory*.

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.

Nankervis, J.C., 2001. Stopping rules for double bootstrap tests. Unpublished manuscript, University of Surrey, Department of Economics.

Politis, D. N., Romano, J.P., 1992. General resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Annals of Statistics*, 20, 1985-2007.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1303-131.

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, Stanford University, Department of Statistics.

Tong, H., 1990. *Nonlinear Time Series*. Oxford University Press, Oxford.

Table 1
Rejection Probabilities (Percent) of Tests: Diagonal MDS, n = 500[a]

	H_1			H_5			H_{10}		
Tests	1	5	10	1	5	10	1	5	10
	Diagonal One-Dependent Homoskedastic Case								
Q_K	12.3	24.8	33.3	6.4	15.0	22.4	4.5	11.6	18.4
SBOB Q_K									
b = 4	2.0	7.0	12.4	1.3*	5.7**	10.8*	1.0*	4.8*	10.3*
b = 10	2.4	8.1	13.3	1.5	6.3	12.3	1.2*	5.8**	11.7
b = 20	2.9	8.9	14.3	1.7	7.3	13.5	1.4**	6.7	13.6
DBOB Q_K									
b = 4	1.3*	5.6*	10.8*	0.9*	4.5*	9.2*	0.7*	3.8	8.3
b = 10	1.0*	6.2	11.1**	0.9*	4.9*	9.9*	0.9*	4.6*	9.7*
b = 20	1.1*	6.2	11.0**	0.9*	5.2*	10.5*	0.9*	5.3*	10.8*

Q_K*	0.7	4.8*	9.7*	0.9*	4.5	9.5**	0.9*	4.6	9.3
GP_K	0.7	4.8*	9.7*	0.5	3.9	8.8	0.4	3.4	7.8
$\tilde{Q}_K$	0.5	4.2	9.3	0.7	4.3	9.3	0.7	3.8	8.6

[a] Notes: 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 denotes acceptance of the nominal rejection probability by a 0.05 symmetric asymptotic test, and two asterisks denote acceptance by a 0.01 symmetric asymptotic test.

Table 2
Rejection Probabilities (Percent) of Tests: GARCH(1,1) Models, n = 500 [a]

	H_1			H_5			H_{10}		
Tests	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
$\widetilde{Q}_K$	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
$\widetilde{Q}_K$	1.0*	5.0*	10.1*	1.1*	5.3*	10.8	0.9*	5.0*	10.3*

[a] Notes: See Table 1.

Table 3
Rejection Probabilities (Percent) of Tests: Non-MDS, n = 500 [a]

	H_1			H_5			H_{10}		
Tests	1	5	10	1	5	10	1	5	10
	Nonlinear Moving Average Case								
Q_K	26.1	38.9	46.9	19.1	31.2	39.0	14.7	24.6	32.3
SBOB Q_K									
b = 4	3.6	11.6	19.0	2.3	8.3	15.6	1.5	6.3	12.6
b = 10	4.3	12.2	19.6	2.6	9.6	17.0	1.4	7.1	13.7
b = 20	4.6	12.9	20.6	2.7	10.5	18.4	1.7	8.0	15.3
DBOB Q_K									
b = 4	1.3*	7.0	13.1	0.8*	5.2*	10.2*	0.7*	3.9	8.1

b = 10	1.3*	6.3	12.5	0.9*	5.8**	10.8*	0.9*	4.5*	8.9**
b = 20	1.6	7.3	12.9	1.1*	5.9	11.6	1.0*	4.7*	9.7*
Q_K*	1.3	7.2	14.6	1.0*	5.2*	10.9	1.1*	5.1*	10.0*
GP_K	1.3	7.2	14.6	0.5	4.0	9.3	0.7	3.3	7.9
$\tilde{Q}_K$	0.7	4.4	9.9*	1.3	4.9*	9.9*	1.9	5.5	10.2*
	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

[a] Notes: See Table 1.

Table 4
Powers (Percent) for the 0.05 Nominal $\tilde{Q}_K$, Q^*_K, GP_K and Double Bootstrap Q_K Tests, n = 500 [a]

ρ	Tests	One-Dependent Homoskedastic Case				Nonlinear Moving Average Case		
		H_1	H_5	H_{10}		H_1	H_5	H_{10}
ρ = 0.0	$\tilde{Q}_K$	4.2	4.3	3.8		4.4	4.9*	5.5
	Q^*_K	4.7*	4.6**	4.4		7.3	5.0*	4.9*
	GP_K	4.7*	3.9	3.5		7.3	3.7	3.2
	DBOB Q_K							
	b = 4	5.8**	4.6*	3.5		6.8	5.9	3.5
	b = 10	5.8**	5.2*	4.1		6.5	6.0	3.7
	b = 20	6.0	5.7**	4.4*		6.4	6.4	4.0

ρ = 0.05	$\tilde{Q}_K$	8.3	6.1	5.0		4.7	4.8	5.6
	Q^*_K	9.6	6.2	5.9		7.6	4.9	4.7
	GP_K	9.6	5.4	4.2		7.6	4.1	3.8
	DBOB Q_K							
	b = 4	10.6	8.5	6.2		6.3	6.0	4.0
	b = 10	10.4	9.2	6.8		5.9	6.3	4.3
	b = 20	10.3	9.2	7.2		6.2	6.5	4.8
ρ = 0.1	$\tilde{Q}_K$	23.4	13.4	9.3		11.3	7.2	7.1
	Q^*_K	25.1	13.1	10.7		16.0	8.1	7.1
	GP_K	25.1	11.6	7.8		16.0	8.1	6.1
	DBOB Q_K							
	b = 4	26.4	20.9	16.0		14.0	12.5	9.4
	b = 10	26.2	21.6	17.1		13.2	12.5	10.1
	b = 20	25.8	21.2	18.5		13.4	12.5	10.3
ρ = 0.2	$\tilde{Q}_K$	74.5	52.6	37.2		45.2	24.8	17.6
	Q^*_K	76.6	49.1	37.3		51.6	26.2	20.1
	GP_K	76.6	48.3	40.0		51.6	28.4	18.7
	DBOB Q_K							
	b = 4	75.1	68.2	61.1		51.2	45.6	38.9
	b = 10	73.5	67.9	61.1		47.6	45.3	39.7
	b = 20	72.3	67.2	61.0		45.9	44.4	40.1
ρ = 0.3	$\tilde{Q}_K$	98.2	92.3	81.5		79.0	60.0	42.7
	Q^*_K	98.6	91.3	80.4		84.0	56.3	44.1
	GP_K	98.6	91.4	78.4		84.0	59.8	42.4
	DBOB Q_K							
	b = 4	97.5	96.0	93.5		80.3	78.1	76.9
	b = 10	96.8	95.4	94.0		77.0	76.8	76.7
	b = 20	95.6	95.3	93.5		74.2	75.6	75.8

[a] Notes: The number of replications for the Q_K test with BOB bootstrap-based P-values is 5,000. The number of replications for the $\tilde{Q}_K$, Q^*_K, GP_K tests is 25,000.