PRELIMINARY AND INCOMPLETE DRAFT

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

Joel L. Horowitz

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

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.

October 18, 1999

Abstract

This paper considers 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 critical 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 and by a theoretical analysis based on a formal Edgeworth expansion.

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 test 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. The Q_K test, however, can produce misleading inferences when the time series is uncorrelated but statistical dependence is present. Romano and Thombs (1996) have proposed using the bootstrap to make robust inferences about the individual autocorrelation correlation coefficients. The rationale for their procedure is that the bootstrap provides a first-order asymptotic approximation to the distribution of the sample autocorrelation coefficients under the null in the presence of statistical dependence.

In this paper, the Q_K test statistic with bootstrap-based critical values is used to test the null that the first K autocorrelations are zero. We use the blocks of blocks (BOB) bootstrap (Politis and Romano (1992)). Monte Carlo evidence reported by Davison and Hinkley (1997, Table 8.2) for time series data suggests that the BOB bootstrap is less sensitive to the choice of block length than are alternative resampling procedures such as the moving block procedure. Intuitively speaking, this advantage of the BOB bootstrap is due to the fact that it reduces the influence of edge effects produced by blocking.

We conducted Monte Carlo experiments to investigate the finite sample performance of the BOB bootstrap procedure for sample of size n = 500; Romano and Thombs (1996) used n =1000. The experimental evidence confirmed that BOB bootstrap is indeed robust to the choice of block length. The performance of the BOB bootstrap, however, turned out to be unsatisfactory despite the fact that the bootstrap produced smaller distortions in the rejection probabilities than the Q_K test based the chi-square distribution. In particular, the differences between the true and nominal rejection probabilities were often large enough to produce substantially misleading inferences in applications.

The less than satisfactory performance of our initial implementation of the bootstrap is not surprising. The bootstrap provides asymptotic refinements to the rejection probabilities of a test statistic if that statistic is asymptotically pivotal. A statistic is asymptotically pivotal if its asymptotic distribution is independent of unknown population parameters. If a test statistic is asymptotically pivotal, then as the sample size increases, the difference between the true and nominal probabilities of rejecting a correct null hypothesis decreases more rapidly with a bootstrap-based critical value than with the asymptotic critical value. However, the Q_K test is asymptotically pivotal only when the data generation process is restricted to the class of independently and identically distributed (iid) processes. The Q_K test is not asymptotically pivotal if the class of data generation processes includes processes that are uncorrelated but depended (e.g., through ARCH or GARCH). When a test statistic is not asymptotically pivotal, the errors in the rejection probabilities with asymptotic and bootstrap critical values decrease to zero at the same rate as the sample size increases. Thus, there is no reason to expect the Q_K test with bootstrap critical values to be more accurate than the same test with asymptotic critical values unless the data generation process is restricted to the class of iid processes.

In settings where the observations are iid, higher order approximations to the distributions of statistics that are not asymptotically pivotal can be obtained using the bootstrap to create an asymptotic pivot. This procedure is called prepivoting (Beran (1987), (1988)). It is also called the iterated or double bootstrap because prepivoting requires two stages of bootstrapping. In this paper, we address the lack of asymptotic pivotality by prepivoting through the use of double

BOB bootstrap procedure. At present, there is no proof that prepivoting produces a higher order approximation in the case of dependent data. We outline such a proof in Section 4.

The numerical performance of the bootstrap tends to be better when the process used to generate the bootstrap samples satisfies the null hypothesis. In general, the first K autocorrelations in the sample are nonzero even if the first K population autocorrelations are zero and consequently the sample does not mimic the population when the null is true. In our application, the null hypothesis can be imposed, at least approximately, by prewhitening the time series and treating the prewhitened series as the sample in bootstrap resampling. The prewhitening is carried out by running a Kth order autoregression where K is the maximum number of autocorrelations to be tested. The residuals from this autoregression are treated as the prewhitened series. Prewhitening reduces the magnitude of the autocorrelation in the sample from $O(n^{-1/2})$ to $O(n^{-1})$ and thereby making the sample autocorrelations asymptotically negligible in our setting.

We investigate the performance of the single and double BOB tests (SBOB and DBOB tests) with prewhitening in Monte Carlo experiments where the time series are generated by martingale difference sequence (MDS) processes as well as non-MDS processes. The DBOB tests are prepivoted. The SBOB tests are not. For the MDS case, we used examples from Romano and Thombs (1996) and Politis, Romano and Wolf (1997) as well as GARCH models. Nonlinear moving average and bilinear models were used to generate time series data in the case of non-MDS processes. The performance of the DBOB test was generally impressive and typically superior to that of the SBOB test. A conjecture suggested by this evidence is that DBOB bootstrap with prewhitening yields an asymptotic refinement.

In this paper, the accuracy of the bootstrap is investigated using a formal Edgeworth expansion in the spirit of Mammen (1993). The expansion provides the basis for a heuristic study of the higher-order distributional properties of the bootstrap assuming that the formal Edgeworth expansion is valid to sufficiently high order. Establishing conditions for the validity expansion is a very difficult problem in probability theory that we do not attempt to solve here. Rather, we use the formal Edgeworth expansion as a heuristic device.

The Q_K statistic is only one of several test statistics that can be employed to test for uncorrelatedness. A number of 'robust' tests have been developed to address the shortcomings of the Q_K test. These include the Q_K* test, which is a modified Box-Pierce test, (Diebold (1986), Lo and MacKinlay (1989) and Lobato, Nankervis and Savin (1999a)), the GP_K test (Guo and Phillips (1998)), the $\tilde{Q}_K$ test (Lobato, Nankervis and Savin (1999b)) and the T_K test (Lobato (1999). The Q_K* test is designed for the case where the time series is generated by a MDS process and the asymptotic covariance matrix of the sample autocorrelations is diagonal. The GP_K test is also designed for the MDS case but relaxes the assumption that the asymptotic covariance matrix is diagonal. Finally, the $\tilde{Q}_K$ and T_K tests are asymptotically valid for both MDS and non-MDS processes, and hence they are natural competitors to the bootstrap test. We provide Monte Carlo evidence on the performance of these tests.

We also report the results of Monte Carlo experiments where the single and double BOB bootstrap is carried out using the Q_K* and the GP_K test statistics. The performance of the SBOB and DBOB tests which exploit the more robust test statistics is usually unsatisfactory compared to that SBOB and DBOB tests based on the Q_K statistic. The explanation of this result is a topic on our current research agenda.

The organization of the paper is the following. Section 2 describes the Q_K test with bootstrap-based P-values. Section 3 reports the results of Monte Carlo experiments for MDS examples and Section 3 for non-MDS examples. Section 4 presents the results of a heuristic examination of the accuracy of the single and double bootstrap with prewhitening. Section 5 reports the results of Monte Carlo experiments using robust test statistics and Section 6 presents the results of a small Monte Carlo power study for the SBOB and DBOB tests based on the Q_K statistics. The concluding comments are in Section 7. Some technical computational issues in are addressed in the Appendix.

2. BOOTSTRAP TEST

The bootstrap provides at least a first-order asymptotic approximation to the distribution of the Q_K test statistic under the null. Thus, the null can be tested by comparing the Q_K test 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 double BOB bootstrap with prewhitening to calculate the P-values. In the Monte Carlo experiments we compare the performance of the single and double BOB tests. The first objective of this section is to describe the calculation of bootstrap P-values for the Q_K test using the single BOB bootstrap. 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 $\gamma(j) = E(y_t - m)(y_{t+j} - m)$ 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 T where the ij-th element of T is given by

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

where

$$c_{i+1,j+1} = \sum_{d=-\infty}^{d=\infty}\{E(y_t - m)(y_{t-i} - m)(y_{t+d} - m)(y_{t+d-j} - m) - E(y_t - m)(y_{t-i} - m)E(y_{t+d} - m)(y_{t+d-j} - m)\};\ i, j = 0,1,\ldots, K; \qquad (2)$$

see Hannan and Heyde (1972) and Romano and Thombs (1996). Assuming T is known, H_K: $\rho = [\rho(1), \ldots, \rho(K)]' = 0$ can be tested using a test statistic of the form $nr'T^{-1}r$, which asymptotically follows a chi-square with K degrees of freedom when H_K is true. In practice, T is unknown. A feasible test can be obtained either by replacing T by a known matrix or by estimating T.

The Box-Pierce Q_K statistic (Box-Pierce (1970) replaces T with the identity matrix. The Q_K* test replaces T 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 T with an estimator that is consistent under the null for MDS processes. The $\tilde{Q}_K$ test replaces T with an estimator that is consistent under the null for both MDS and non-MDS processes; for details, see Lobato, Nankervis and Savin (1999b).

In this paper, the null H_K is tested using the P-value of the Q_K test statistic. Each sample of n observations $y_1,\ldots, y_n$ produces a specific value of the test statistic, 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 if $p < \alpha$ where α is the probability of a Type I error. The p-value can be calculated from some predetermined distribution or estimated by the bootstrap. We now show to obtain an estimate of the p value using the single BOB bootstrap and double BOB bootstrap.

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 $Y_i = (y_i, y_{i+1}, \ldots, y_{i+K})'$. For the lag one autocorrelation, for example,

$$\left(Y_1, Y_2, \ldots, Y_{n-1}\right) = \begin{pmatrix} y_1 \; y_2 \ldots y_{n-1} \\ y_2 \; y_3 \ldots \; y_n \end{pmatrix}.$$

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 $(K+1)\times b$ matrix given by $B_i = Y_i, \ldots, Y_{i+b-1}$, where i = 1, …, q, and q = n-b-K+1. The single BOB 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 $(K+1)\times n$, which is the bootstrap sample and is denoted by $Y_1^*,\ldots, Y_n^*$, 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

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

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

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

3. Repeat steps 1 and 2 M_1 times.

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 single BOB 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 single BOB 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 double BOB 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 $Y_1^{**}, \ldots, Y_n^{**}$.

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

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

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

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

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 double BOB 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 $\#[\, p_K^{**} \leq p_K^* \,]/M_1$; this formula is also given by Hinkley (1989).

Davison and Hinkley (1997) strongly recommend the use of adjusted P-values. Politis, 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 has not been established in the case of dependent data.

Pre-whitening. The key idea of the bootstrap is to replace the unknown population by the sample for the purpose of approximating the sampling distribution, $P(Q_K > t \mid H_K)$. In general, the sample autocorrelations are nonzero even when the null H_K is true. Thus, in order to approximate the sampling distribution of Q_K when H_K is true, the condition that the first K autocorrelations are zero has to be imposed when resampling, that is, when generating the bootstrap samples. We employ pre-whitening to make the autocorrelations in the sample asymptotically negligible, that is, to impose H_K, at least approximately. The pre-whitening is carried out 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 SBOB with prewhitening we first fit an AR(K)to the original data to obtain the residuals, $e_t = y_{t+K} - \hat{y}_{t+K}$. From the residuals we calculate the restricted sample autocorrelations R(k). 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 $E^* = (E_1^*,\ldots,E_n^*)$ from which r*(k) is calculated and $Q_K^S = n\,\Sigma[r^*(k)-R(k)]^2$.

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). Each element e_t in E* is replaced by $v_t = e_t - \hat{e}_t$ to form the prewhitened SBOB sample $V^* = (V_1^*,\ldots, V_n^*)$ from which the restricted autocorrelations are calculated, say R*(k). For DBOB (K+1)×b blocks are resampled from the matrix V* to form the (K+1) n matrix V** from which r**(k) is calculated and $Q_K^D = n\,\Sigma[r^{**}(k)-R^*(k)]^2$.

3. MONTE CARLO EVIDENCE

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

MDS Examples

The first two MDS examples are motivated by experiments conducted by Romano and Thombs (1996) and Politis, Romano and Wolf (1997). The first example illustrates how the tests perform for a simple one-dependent MDS process where the asymptotic covariance matrix of the sample autocorrelations is diagonal under the null. In the second example, seasonal heteroskedasticity is added to the first example; the purpose of second example is to examine the robustness of the tests to heteroskedasticity. The third and fourth example 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 with 3 degrees of freedom. Under the null, 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 single BOB bootstrap-based P-values, SBOB, are calculated using $M_1 = 599$ replications. The double BOB- based tests, DBOB, are calculated using $M_1 = 599$ and $M_2 = 199$ 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.

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, $\tilde{Q}_K$ and T_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, Nankervis and Savin (1999b).

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 Sun Enterprise 3000 server 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, $\gamma_0 = E(y_t - m)^2 = 1$, $E(y_t - m)^3 / \gamma_0^{3/2} = 0$, $E(y_t - m)^4 / \gamma_0^{2} = 9$, and T is the identity matrix except that $\tau_{11} = 3$. 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.

Example 2. Heteroskedastic Diagonal Case. Let $y_t = a_t \bullet z_t \bullet z_{t-1}$ where $\{z_t\}$ is a sequence of iid N(0,1) random variables and where $\{a_t\}$ is the infinite repetition of the sequence {1,1,1,2,3,1,1,1,1,2,4,6}. This sequence was used in the heteroskedastic autoregressive AR(1) example with n = 256 considered by Politis, Romano and Wolf (1997). The y_t process is uncorrelated for all k and t, but not independent. The mean, variance, skewness and kurtosis for this heteroskedastic process are 0.00, 6.3, 0.0 and 31.26. This example is used to examine the effect of heteroskedasticity on the performance of the tests. For some nonstationary processes, Q_K can be used to test the null hypothesis that the lag-j autocorrelations are zero for all t and j = 1,…, K. Politis, Romano and Wolf (1997) prove that the single moving block procedure is robust to mild forms of heteroskedasticity under certain conditions, which are satisfied by Example 2. This robustness presumably holds for the BOB bootstrap as well.

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

(i) For 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 essentially eliminates the distortions in the rejection probabilities for all three hypotheses: the MAD is about 0.004 at 0.01 and 0.015 at 0.10.

(iii) The SBOB bootstrap substantially reduces the distortions in the empirical rejection probabilities and in several instances essentially eliminates the distortions: the MAD is about 0.01 when the nominal rejection probability is 0.01 and about 0.04 when it is 0.10.

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

The first panel also shows that the asymptotic $\tilde{Q}_K$ and T_K tests tend to under-reject for the each hypothesis, but not by a large margin. The MAD is 0.005 at 0.01 and 0.014 at 0.10 for the $\tilde{Q}_K$ test and about 0.01 at 0.01 and 0.07 at 0.10 for the T_K test. Note that the asymptotic confidence intervals for the rejection probabilities are tighter for the asymptotic $\tilde{Q}_K$ and T_K tests than for the bootstrap tests because the performance of the asymptotic tests is investigated using 25, 000 replications.

The second panel of Table 1 reports the result for the heteroskedastic case. The results are similar to those in the first panel. The main difference is that the performance of the DBOB test is less satisfactory in the presence of heteroskedasticity than in the homoskedastic case. The $\tilde{Q}_K$ and T_K tests again under-reject for each hypothesis, but by a larger margin then in the first panel.

Example 3. 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 in order that y_t is covariance stationary. He and Teräsvirta (1999) show that the unconditional four 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, $\gamma_0 = E(y_t - m)^2 = 0.1$, $E(y_t - m)^3/\gamma_0^{3/2} = 0$, $E(y_t - m)^4/\gamma_0^2 = 4.5$, and T is diagonal where the diagonal elements follow the recursion $\tau_{jj} = (1-\alpha_0-\beta) + (\alpha_0+\beta)\tau_{j-1,j-1}$ where $\tau_{11} = 6.303$. See Lobato, Nankervis and Savin (1999a).

Example 4. Chi-square (3) GARCH. This GARCH (1,1) model is the same as in Example 3 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), $\gamma_0 = E(y_t - m)^2 = 0.1$, $E(y_t - m)^3/\gamma_0^{3/2} = 1.85$, $E(y_t - m)^4/\gamma_0^2 = 10.9$ where T 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}. These distortions are much reduced when SBOB and DBOB bootstrap-based P-values are employed. In particular, the distortions are essentially eliminated by the DBOB test for both GARCH models. SBOB is a close competitor, especially in the case of Gaussian GARCH. Nonetheless, the Monte Carlo evidence favors DBOB over SBOB.

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. By comparison, the asymptotic T_K test under-rejects by a large margin; the under-rejection is very striking in the case of H_{10}.

Non-MDS Examples

The first uncorrelated non-MDS processes is generated by a nonlinear moving average model, and the second by a bilinear model. These nonlinear models are described in Tong (1990, pp.114-115) and also Granger and Teräsvirta (1993). For these two examples, the asymptotic matrix of the sample autocorrelations is nondiagonal under the null and the elements of T under the null are given in the Appendix of Lobato, Nankervis and Savin (1999b).

Example 5. Nonlinear Moving Average Case. Let $y_t = z_t \bullet z_{t-2} \bullet (z_{t-2} + z_t + c)$ 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 6. Bilinear Case. Let $y_t = z_t + b\, z_{t-1} y_{t-2}$ 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 $m = 0$, $\gamma_0 = E(y_t - m)^2 = \sigma^2/(1 - b^2 \sigma^2) = 1.333$, $E(y_t - m)^3/ \gamma_0^{3/2} = 0$, $E(y_t - m)^4/ \gamma_0^{2} = 3(1 - b^4\sigma^4)/(1 - 3b^4\sigma^4) = 3.462$.

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 test work satisfactorily for both of the non-MDS examples. 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 0.25 at 0.01 and 0.45 at 0.10. The distortions are considerably less for the bilinear model, but they are large nonetheless. Note that the SBOB test tends over-reject, more so for Example 5 than Example 6. For example, in the case of Example 5, the MAD between the empirical and nominal rejection probability is about 0.035 at 0.01 and 0.08 at 0.10.

The asymptotic $\tilde{Q}_K$ test tends to perform better than the asymptotic T_K test for the nonlinear moving average model while the reverse is true for the bilinear model. The asymptotic $\tilde{Q}_K$ test works satisfactorily in several instances for Example 5 and similarly for asymptotic T_K test for Example 6.

4. OUTLINE OF BOOTSTRAP THEORY

The aim is to use the block bootstrap with prepivoting to obtain an asymptotic refinement of the rejection probability of Q. "Asymptotic refinement" means that the difference between the true and nominal rejection probabilities converges to zero more rapidly than it would if the critical value were obtained from the first-order asymptotic distribution (limiting distribution) of Q under the null hypothesis H_0. This section outlines the as yet incomplete calculations. The calculations use certain non-standard Edgeworth expansions. The literature on Edgeworth expansions does not contain results giving regularity conditions for the validity of the expansions that are needed. Derivation of these conditions is beyond the scope of this paper. Thus, we assume that the required expansions exist without providing regularity conditions. Accordingly, our derivation is heuristic.

Let Q_n be the Q statistic based on a sample of size n. Denote the sample by c. Assume that H_0 is true. Let G_n denote the CDF of Q_n. That is, $G_n(q) = P(Q_n \leq t)$ for any fixed t. Then $G_n(Q_n) \sim U[0,1]$. If G_n were known, it would be a pivotal statistic and could be used to test H_0. Specifically, H_0 is rejected at the a level if $G_n(Q_n) > 1 - a$. This test statistic is not feasible, however, because G_n is unknown in applications.

G_n can be estimated by the (block) bootstrap. Let Q_n^* be the version of Q_n that is obtained from the block bootstrap sample c^*. By forming repeated block bootstrap samples, one obtains repeated values of Q_n^*. The empirical distribution of these estimates G_n. Let G_n^* denote the block bootstrap estimator of G_n. That is, $G_n^*(t) = P^*(Q_n^* \leq t)$, where P^* is the probability measure induced by block bootstrap sampling. Note that G_n^* can be estimated with arbitrary accuracy in an application. Consider, therefore, the test statistic $t_n \equiv G_n^*(Q_n)$. This is a feasible test statistic. Its asymptotic distribution under H_0 is U[0,1], regardless of the details of the data generation process. Thus, t_n is asymptotically pivotal, and its asymptotic a-level critical value is $1 - a$.

Bootstrap iteration can be used to obtain an asymptotic refinement for t_n. To do this, let c_j^* denote the j'th block bootstrap sample. Draw second-stage block bootstrap samples from each first-stage sample c_j^*. Let Q_{nj}^{**} denote the version of Q_n that is obtained by sampling c_j^*. By repeatedly sampling c_j^*, one can obtain the empirical distribution $G_{nj}^{**}(t) = P_j^{**}(Q_{nj}^{**} \leq t)$, where P_j^{**} is the probability measure induced by block bootstrap sampling from c_j^*. Define $t_{nj}^* \equiv G_{nj}^{**}(Q_{nj}^*)$. Let t_n^* denote the generic bootstrap random variable of which t_{nj}^* is the j'th realization. Let v_a^* satisfy

$$P^*(t_n^* \leq v_a^*) = 1 - a \ . \qquad (1)$$

Then v_a^* is the double block bootstrap critical value of t_n. That is, we reject H_0 at the nominal a level if $t_n > v_a^*$.

We expect to be able to use arguments similar to those of Beran (1988) to show that this procedure gives an asymptotic refinement to the rejection probability of a test based on t_n. That is, we expect to be able to show that the difference between the nominal and true rejection probabilities decreases to zero more rapidly when v_a^* is used as the critical value than when 1 - a is used. As in Beran (1988), the argument will be based on Edgeworth expansions, but the expansions in our case are non-standard due to the effects of blocking. Consequently, we do not have regularity conditions under which the expansions are valid. In addition, since block-bootstrap cumulants converge to population cumulants more slowly than do iid bootstrap sample cumulants, our block bootstrap prepivoting procedure will not necessarily produce the same degree of asymptotic refinement as can be obtained through prepivoting under iid sampling.

5. ROBUST TESTS

As noted in the introduction, there are a number of robust test statistics that can be used to test for uncorrelatedness. The single and double BOB bootstrap can be based on these robust statistics. The robust tests tend to perform better than the Q_K test when each the test is based on first-order asymptotic theory: the robust tests tend to produce smaller distortions in the rejection probabilities under the null than the Q_K test. This raises an interesting question: does the bootstrap perform better in the case of the robust test statistics than with the Q_K statistic?

This section examines the performance of the BOB bootstrap tests with prewhitening when they are based on robust test statistics. The robust test statistics employed are Q_K^*, GP_K and T_K. The empirical rejection probabilities under the null are calculated in a set of Monte Carlo experiments that use a subset of the examples described in Section 3, namely the one-dependent example and the nonlinear moving average example.

The tests with single BOB bootstrap-based P-values, SBOB, are calculated using $M_1 = 999$ replications. The double BOB- based tests, DBOB, are calculated using $M_1 = 999$ and $M_2 = 249$ replications. Tables 4 and 5 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 2,000 replications. The results for the bootstrap tests are reported for three block lengths, b = 4, 10 and 20.

Tables 4 and 5 report the empirical rejection probabilities of the robust tests when they are based on first order asymptotic theory. Among the three tests, the Q_K*, GP_K and T_K tests, Q_K* has displays the best performance for both the one-dependent and nonlinear moving average examples. This is despite the fact that the Q_K* test is not valid for the nonlinear moving average model. For this model the MAD for the Q_K* test is about 0.002 at 0.01 and 0.05 and 0.10.

Turning to the question of interest, the results in Tables 4 and 5 show that the bootstrap using the robust test statistics does not generally outperform the bootstrap using the Q_K statistic. The bootstrap with robust statistics performs best when testing H_1. For this null, the bootstrap with robust statistics performs about the same as the bootstrap with Q_K. However, the bootstrap with the robust statistics is generally inferior to that with Q_K when testing H_5 and $H_{10.}$ The bootstrap test using Q_K* tends to under-reject for H_{10}. The bootstrap tests using GP_K and T_K substantially under-reject for H_5 and H_{10}: for DBOB the MAD is about 0.01 at 0.01 and 0.09 at 0.10. We note that the Q_K* and GP_K statistics are identical when K =1: this explains why the empirical rejection probabilities of the two tests are identical for H_1.

6. POWER EXPERIMENTS

7. 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, we test the null hypothesis of uncorrelatedness using the Q_K test statistic with bootstrap-based P-values. In our Monte Carlo experiments, we investigated whether the true rejection probability of the bootstrap test is close to the nominal rejection probability when the null hypothesis is true. The bootstrap was implemented using both a single and a double blocks of blocks procedure with prewhitening. The experiments were carried using MDS and non-MDS processes for samples of size 500. For the processes considered in this study, there were large distortions in the empirical rejection probabilities when the Q_K test was based on asymptotic P-values.

The main Monte Carlo findings of this study are twofold. First, our bootstrap procedure works well: the bootstrap essentially eliminates the distortions in the empirical rejection probabilities that are present when the Q_K test is based on first order asymptotic theory. Second, the results are generally robust to the choice of the block length. On the basis of this evidence, we recommend using the double blocks of blocks procedure with prewhitening. We also conducted a Monte Carlo investigation of the this bootstrap procedure using several robust test statistics for testing uncorrelatedness, namely Q_K*, GP_K and T_K . The performance of the bootstrap was unsatisfactory when using the robust test statistics.

The simulation results suggest that the bootstrap procedure yields an asymptotic refinement in the case of the Q_K statistic. In the case of an asymptotic refinement, the difference between the true and nominal rejection probabilities converges to zero more rapidly than it would if the critical value or P-value were obtained from the first-order asymptotic distribution of Q_K under the null hypothesis. The suggestion is investigated using a formal Edgeworth expansion. Our theoretical calculations are as yet incomplete. We noted that the literature does not contain the regularity conditions for the validity of the expansions that are needed. Thus, our theoretical results are heuristic.

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.

ACKNOWLEDGEMENTS: We thank Paul Weller and seminar participants at Monash University for useful comments. 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 $\#(p_K^{**} \leq p_K^*)/M_1 = \#(\ \#(Q_K^{DB}$

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

We avoid unnecessary replications by stopping after m_2 replications if $\sum_{i=1}^{m_2} I(Q_{K_i}{}^{DB} > Q_K{}^B)$ 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. 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 $\sum_{i=1}^{m_1} I(p_{K_i} ** \le p_K *)$ exceeds $0.1M_1$. 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^B. The purpose of this ordering is to exploit the negative correlation between p_K** and Q_K^B so that $\sum_{i=1}^{m_1} I(p_{K_i} ** \le p_K *)$ 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.

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

Bera, A.K., and Higgins M.L. (1997), "ARCH and Bilinearity as Competing Models for Nonlinear Dependence," *Journal of Business and Economic Statistics*, 15, 43-51.

Beran, R (1987), " Prepivoting to Reduce Level Error of Confidence Sets, " *Biometrika* 74, 457-468.

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., and 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.

Davison, A.C., and Hinkley, D.V. (1997), *Bootstrap Methods and Their Application*, Cambridge, U.K.: Cambridge University Press.

Den Haan, W. J., and Levin, A. (1997), "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics*: *Robust Inference* (Vol 15), eds. G. S. Maddala and C.R. Rao, Amsterdam: North-Holland, 291-341.

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.

Granger, C. W. J. and Andersen, A. P (1978), *An Introduction to Bilinear Time Series Models*. Gottingen: Vanenhoek and Ruprecht.

Granger, C. W. J. and Teräsvirta,T. (1993), *Modelling Nonlinear Economic Relationships*. Oxford: University of Oxford Press.

Guo, B. B. and 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. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Hall, P., and 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., and Jing, B-Y. (1995), "On Blocking Rule for the Bootstrap with Dependent Data," *Biometrika*, 82, 3, 561-574.

Hamilton, K.G. and James, F. (1997), "Acceleration of RANLUX", *Computer Physics Communications*, 101, 241-248.

Hannan, E. J., and Heyde, C. C. (1972), "On Limit Theorems for Quadratic Functions of Discrete Time Series," *Annals of Mathematical Statistics*, 43, 2058-2066.

He, C. and 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., (1998), "The Bootstrap." In: Heckman, J.J., and Leamer,E.E. (Eds.), *Handbook of Econometrics*, vol 5, Elsevier Science, Amsterdam. Forthcoming.

Künsch, H.R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217-1241.

Lo, A.W., and 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, (1999), "Testing that a Dependent Process is Uncorrelated," unpublished manuscript, Centro de Investigacion Economica, ITAM, Mexico City.

Lobato, I., Nankervis J.C., and Savin N. E. (1999a), "Testing for Autocorrelation Using a Modified Box-Pierce Q Test," unpublished manuscript, University of Iowa, Department of Economics. Forthcoming in the *International Economic Review.*

Lobato, I., Nankervis J.C., and Savin N. E. (1999b), "A Robust Test for Autocorrelation in the Presence of Statistical Dependence," unpublished manuscript, University of Iowa, Department of Economics.

Newey, W.K., and West, K.D. (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

Numerical Algorithms Group (1997), NAG Fortran 77 Library Manual, Mark 17, Numerical Algorithms Group, Oxford.

Politis, D. N., and Romano, J.P. (1992), "A 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., and Romano, J.P. (1994), "Large Sample Confidence Regions based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 22, 2031-2050.

Politis, D. N., and Romano, J.P., Wolf, M. (1997), "Subsampling for Heteroskedastic Time Series," *Journal of Econometrics*, 81, 281-318.

Robinson, P.M., and Velasco, C. (1997),"Autocorrelation Robust Inference," in *Handbook of Statistics*: *Robust Inference* (Vol 15), eds. G. S. Maddala and C.R. Rao, Amsterdam: North-Holland, 267-298.

Romano, J.L., and Thombs L. A. (1996), "Inference for Autocorrelations under Weak Assumptions," *Journal of American Statistical Association*, 91, 590-600.

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

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

	H_1			H_5			H_{10}		
Tests	1	5	10	1	5	10	1	5	10
	Diagonal One-Dependent Homoskedastic Case								
Q_K	12.7	24.6	33.0	6.4	14.9	22.4	4.6	12.0	18.6
$\tilde{Q}_K$	0.5	4.2	9.3	0.7	4.3	9.3	0.7	3.8	8.6
T_K	0.7	4.6**	9.5**	0.4	3.4	7.9	0.1	1.5	4.2
SBOB Q_K									
b = 4	1.5	6.7	12.1	1.0*	6.0	11.1*	0.7*	5.2*	9.8*
b = 10	1.8	7.1	12.9	1.2*	6.6	12.1	1.0*	5.5*	10.6*
b = 20	2.2	7.6	13.8	1.4**	7.4	13.1	1.1*	5.5*	11.8
DBOB Q_K									
b = 4	0.6**	4.8*	9.8*	0.7*	4.7*	9.4*	0.6**	4.3**	8.5
b = 10	0.9*	4.9*	10.2*	0.8*	4.9*	9.7*	0.8*	4.2**	8.9**
b = 20	0.8*	5.4*	10.1*	1.0*	5.0*	9.9*	0.8*	4.3**	9.1**
	Diagonal One-Dependent Heteroskedastic *Case								
Q_K	19.4	32.4	40.8	7.7	14.8	20.5	3.9	8.7	12.6
$\tilde{Q}_K$	0.3	3.1	8.0	0.5	3.1	7.6	0.4	3.3	7.9
T_K	0.5	3.9	8.8	0.3	2.6	6.5	0.1	0.9	2.7
SBOB Q_K									
b = 4	1.3*	6.4	12.3	1.0*	5.3*	10.9**	0.5	4.4**	9.6*
b = 10	1.9	7.7	14.0	1.5	6.4	12.5	0.8*	5.0*	10.6*
b = 20	2.4	8.2	14.7	1.7	7.0	13.2	0.9*	5.7**	11.6
DBOB Q_K									
b = 4	0.5	3.8	8.6	0.5	3.4	8.3	0.3	3.1	7.5
b = 10	0.8*	4.4*	9.4*	0.8*	4.1	9.0**	0.5	3.8	8.6
b = 20	1.0*	4.9*	9.5*	0.8*	4.5*	9.4*	0.7*	4.1	8.8

NOTE: The number of replications for the Q_K test with BOB bootstrap-based P-values is 5,000. The number of replications for the Q_K , $\tilde{Q}_K$ and T_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

Tests	H_1			H_5			H_{10}		
	1	5	10	1	5	10	1	5	10
	GARCH(1,1) with Normal Errors								
Q_K	2.3	7.9	14.2	4.0	11.9	19.2	6.1	15.1	22.9
$\tilde{Q}_K$	0.8**	4.9*	9.9*	0.8**	4.5	9.5**	0.6	4.2	8.8
T_K	0.4	2.8	6.7	0.2	1.7	4.1	0.1	0.9	2.2
SBOB Q_K									
b = 4	1.2*	5.5*	10.8*	1.1*	4.9*	10.0*	1.3*	5.3*	10.2*
b = 10	1.4*	6.5	11.4	1.2*	5.6**	11.3	1.3*	5.4*	10.9**
b = 20	2.2	7.3	12.7	1.3*	6.3	12.2	0.9*	5.7**	12.1
DBOB Q_K									
b = 4	1.0*	4.9*	9.6*	0.9*	4.2	8.8	1.2*	4.8*	8.9**
b = 10	1.0*	5.5*	10.3*	0.9*	5.0*	9.7*	1.0*	4.7*	9.6*
b = 20	1.1*	6.1	10.9**	0.9*	4.8*	9.3*	0.6**	4.5	9.6*
	GARCH(1,1) with Chi-Square(3) Errors								
Q_K	3.5	10.5	17.4	7.6	18.1	26.7	11.2	23.6	32.8
$\tilde{Q}_K$	1.0*	5.0*	10.1*	1.1*	5.3*	10.8	0.9*	5.0*	10.3*
T_K	0.3	2.5	6.2	0.1	0.7	1.9	0.0	0.1	0.3
SBOB Q_K									
b = 4	1.4**	6.4	11.7	1.3*	5.6*	11.5	1.2*	5.1*	10.4*
b = 10	1.8	6.8	12.7	1.4**	6.4	12.8	1.1*	5.5*	11.7
b = 20	2.3	8.0	13.9	1.6	7.2	14.3	1.2*	6.1	13.3
DBOB Q_K									
b = 4	0.8*	5.1*	9.9*	0.8*	4.4*	9.0**	0.9*	4.4*	8.8
b = 10	1.2*	5.1*	10.6*	1.0*	4.6*	10.5*	0.9*	4.4*	9.2*
b = 20	1.2*	5.8**	11.4	1.0*	5.2*	10.6*	0.8*	4.4*	9.9*

NOTE: See Table 1.

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

	H_1			H_5			H_{10}		
Tests	1	5	10	1	5	10	1	5	10
	Nonlinear Moving Average Case								
Q_K	25.7	38.3	46.3	18.9	30.4	38.5	14.4	24.4	31.6
$\tilde{Q}_K$	0.7	4.4	9.9*	1.3	4.9*	9.9*	1.9	5.5	10.2*
T_K	0.6	3.9	9.1	0.2	2.2	5.8	0.0	0.4	1.5
SBOB Q_K									
b = 4	3.7	10.4	16.9	2.3	8.4	15.3	1.4	6.6	12.6
b = 10	4.1	10.9	17.4	2.6	9.2	16.1	1.5	7.0	13.3
b = 20	4.5	11.1	17.9	2.8	9.6	16.9	1.5	7.3	14.4
DBOB Q_K									
b = 4	1.2*	6.0	11.0**	1.0*	5.3*	10.0*	0.7*	4.1	8.9**
b = 10	1.1*	6.0	10.8*	1.0*	5.3*	10.5*	0.7*	4.4*	9.1**
b = 20	0.8*	6.0	11.2	1.3*	5.4*	10.2*	0.7*	4.1	9.2*
	Bilinear Case								
Q_K	5.4	14.2	21.5	5.8	15.3	23.8	4.1	12.1	19.5
$\tilde{Q}_K$	1.4	6.3	12.3	1.6	6.8	13.0	1.1*	5.4**	10.9
T_K	0.9*	5.0*	10.1*	1.2**	5.4**	10.6	0.6	3.7	8.7
SBOB Q_K									
b = 4	2.4	7.8	13.9	1.6	6.6	12.9	1.4**	6.0	11.5
b = 10	2.3	7.2	13.4	1.7	6.8	12.9	1.3*	5.8**	11.4
b = 20	2.7	7.4	13.3	1.7	7.1	13.2	1.1*	5.8**	12.1
DBOB Q_K									
b = 4	1.4*	6.1	11.5	1.1*	4.8*	10.6*	1.2*	4.8*	9.7*
b = 10	1.1*	5.3*	10.1*	1.2*	4.8*	9.9*	1.1*	4.6*	9.6*
b = 20	1.1*	5.4*	10.2*	1.0*	5.3*	10.0*	0.8*	4.5**	9.4*

NOTE: See Table 1.

Table 4. Rejection Probabilities (Percent) of Robust Tests: Diagonal MDS, n = 500

	H_1			H_5			H_{10}		
Tests	1	5	10	1	5	10	1	5	10
	One-dependent Case								
Q_K	12.3	24.9	33.0	6.8	15.2	23.6	5.0	12.4	19.2
Q_K*	0.8*	4.5*	9.7*	0.9*	5.2*	10.1*	0.9*	4.7*	9.9*
GP_K	0.8*	4.5*	9.7*	0.8*	4.7*	9.9*	0.5**	3.9**	8.6**
T_K	0.9*	4.2*	10.4*	0.6*	4.0*	8.1	0.1	2.1	4.8
SBOB Q_K*									
b = 4	0.9*	4.0*	9.0*	0.5**	4.2*	8.5**	0.4**	3.0	7.2
b = 10	0.9*	4.7*	9.2*	0.8*	4.6*	9.2*	0.4**	3.0	7.4
b = 20	1.4*	5.5*	10.4*	0.9*	5.0*	9.5*	0.2	3.2	7.9
DBOB Q_K*									
b = 4	0.7*	3.4	7.5	0.5**	3.8**	8.1	0.4**	3.0	6.7
b = 10	0.7*	3.6	7.8	0.7*	4.4*	8.6**	0.4**	3.2	6.8
b = 20	0.7*	4.7*	8.2	0.7*	4.2*	8.6**	0.2	3.1	7.1
SBOB GP_K									
b = 4	0.9*	4.0*	9.0*	0.3	3.0	7.7	0.1	1.4	7.1
b = 10	0.9*	4.7*	9.2*	0.4**	3.5	8.2	0.2	1.7	4.9
b = 20	1.4*	5.5*	10.4*	0.5**	3.5	8.6**	0.1	1.7	4.7
DBOB GP_K									
b = 4	0.7*	3.4	7.5	0.3	2.9	7.5	0.4**	1.8	4.7
b = 10	0.7*	3.6	7.8	0.5**	3.4	7.7	0.4**	2.0	4.8
b = 20	0.7*	4.7*	8.2	0.6*	3.5	7.3	0.3	1.9	5.1
SBOB T_K									
b = 4	0.8*	3.9	9.1*	0.2	3.0	6.3	0.1	1.1	3.3
b = 10	0.9*	4.0*	9.3*	0.3	2.9	5.8	0.1	0.8	2.8
b = 20	0.8*	3.9	9.2*	0.2	2.5	5.2	0.1	0.6	2.5
DBOB T_K									
b = 4	0.9*	4.0*	8.7*	0.4**	3.1	6.5	0.1	1.4	3.4
b = 10	0.8*	3.9**	8.9*	0.4**	3.2	6.1	0.1	0.9	3.1
b = 20	0.7*	3.7**	8.6**	0.2	2.5	5.1	0.1	1.2	2.6

NOTE: The number of replications for the Q_K*, GP_K and T_K tests is 2,000. 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 5. Rejection Probabilities (Percent) of Robust 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	24.2	37.0	45.0	17.8	29.3	37.1	13.1	22.7	30.8
Q_K*	1.1*	7.2	15.0	0.8*	4.6*	10.1*	0.8*	4.0*	8.5**
GP_K	1.1*	7.2	15.0	0.5	3.2	8.1	0.8*	2.9	6.9
T_K	0.7*	3.6	8.6	0.2	1.8	5.8	0.0	0.2	1.2
SBOB Q_K*									
b = 4	1.5	6.2**	11.5**	0.2	1.9	4.4	0.1	0.8	2.6
b = 10	1.6**	6.5	11.9	0.2	2.0	4.9	0.1	0.7	2.5
b = 20	1.9	6.7	11.8	0.2	1.8	5.7	0.1	0.8	2.8
DBOB Q_K*									
b = 4	1.4*	5.0*	9.0*	0.1	1.9	3.6	0.2	0.7	1.9
b = 10	1.4*	4.9*	9.1*	0.4	1.7	3.9	0.1	0.7	2.1
b = 20	1.4*	5.3*	8.8**	0.2	1.9	4.4	0.1	0.7	2.5
SBOB GP_K									
b = 4	1.5**	6.2**	11.5**	0.1	0.9	2.7	0.1	0.4	1.1
b = 10	1.6**	6.5	11.9	0.1	0.9	3.1	0.1	0.5	1.2
b = 20	1.9	6.7	11.8	0.1	1.1	3.2	0.1	0.3	1.2
DBOB GP_K									
b = 4	1.4*	5.0*	9.0*	0.1	0.9	2.6	0.1	0.4	1.1
b = 10	1.4*	4.9*	9.1*	0.1	1.1	2.9	0.1	0.5	1.2
b = 20	1.4*	5.3*	8.8**	0.0	1.1	2.8	0.1	0.4	1.1
SBOB T_K									
b = 4	0.7*	3.4	7.4	0.2	1.0	3.2	0.0	0.1	0.2
b = 10	0.7*	3.2	7.4	0.2	0.9	2.9	0.0	0.1	0.5
b = 20	0.8*	3.0	6.9	0.2	0.7	2.5	0.0	0.1	0.5
DBOB T_K									
b = 4	0.9*	3.6	7.4	0.4	1.4	3.5	0.1	0.1	0.6
b = 10	0.7*	3.5	7.6	0.3	1.6	3.6	0.1	0.4	1.1
b = 20	0.7*	3.3	6.8	0.2	0.9	2.9	0.1	0.4	1.1

NOTE: The number of replications for the Q_K*, GP_K and T_K tests is 2,000. 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.