E C O N O M I C S   B U L L E T I N

A power comparison among tests for time reversibility

Jorge Belaire−Franch
Department of Economic Analysis, University of Valencia

Dulce Contreras
Department of Economic Analysis, University of Valencia

Abstract

Since time reversibility (TR) is a necessary condition for an independent and identically distributed (iid) sequence, several tests for TR have been suggested to be applied as tests for model misspecification. In this paper, we compare the power of two well known TR tests against two situations: 1) the fitted model is a linear ARMA when the true data generating process is a nonlinear−in−mean model (either threshold autorregresive or bilinear), and 2) the fitted model is a symmetric GARCH model but the true process belongs to the asymmetric GARCH family (either EGARCH or GJR).

The authors acknowledge financial support from the Ministerio de Ciencia y Tecnología, project SEC2003−09205. The authors also acknowledge the many useful comments of the Associate Editor Timothy Conley.
Citation: Belaire−Franch, Jorge and Dulce Contreras, (2004) "A power comparison among tests for time reversibility." *Economics Bulletin,* Vol. 3, No. 23 pp. 1−17
Submitted: April 16, 2004. **Accepted:** July 7, 2004.
URL: http://www.economicsbulletin.com/2004/volume3/EB−04C40003A.pdf

1 Introduction

A stationary time series is said to be time reversible if all its finite-dimensional distributions are invariant to the reversal of time indices. Hence, iid sequences and Gaussian ARMA models are time reversible. Therefore, the concept of time reversibility can be used in nonlinear time series analysis to detect deviations from the Gaussianity/linearity hypothesis.

Rothman (1998) suggests using Ramsey and Rothman (1996) TR test against ARMA residuals for simple bilinear (BL) and threshold autorregresive (TAR) models. The distinct rejection pattern he finds can be used in identification of those time series models. On the other hand, Chen and Kuan (2002) propose using Chen et al. (2000) TR test as model misspecification test against standardised symmetric GARCH residuals from asymmetric GARCH models (EGARCH).

In this paper, we compare the power of Ramsey and Rothman (1996) and Chen et al. (2000) procedures, in order to detect deviations from the linear ARMA and symmetric GARCH processes. Next section reviews both testing procedures; section 3 presents results for the TAR and BL models, whereas section 4 presents results for asymmetric GARCH models. Section 5 explores the possibility of using combinations of both tests in a test sequence. Finally, section 6 concludes.

2 Time reversibility tests

2.1 Ramsey and Rothman TR test

Ramsey and Rothman (1996) propose a procedure (RR test, hereafter) to test for time irreversibility, using the difference between the sample bicovariances for a zero mean stationary time series

$$\hat{\gamma}_{2,1}(k) = \hat{B}_{2,1}(k) - \hat{B}_{1,2}(k),$$

where

$$\hat{B}_{2,1}(k) = (T-k)^{-1} \sum_{t=k+1}^{T} x_t^2 x_{t-k}$$

and

$$\hat{B}_{1,2}(k) = (T-k)^{-1} \sum_{t=k+1}^{T} x_t x_{t-k}^2,$$

restricting the value of the maximum lag up to K (a relatively low value, e.g., 5).

If $\{x_t\}$ is a stationary sequence of zero mean independently and identically distributed random variables, then the exact small sample variance, which they advocate to use if the series is not serially correlated, is

$$Var\left[\hat{\gamma}_{2,1}\left(k\right)\right] = 2\left(\mu_4\mu_3 - \mu_3^2\right)/\left(T-k\right) - 2\mu_2^3\left(T-2k\right)/\left(T-k\right)^2.$$

When $\mu_2 = E[X_t^2]$, $\mu_3 = E[X_t^3]$ and $\mu_4 = E[X_t^4]$ are replaced by their sample counterparts, given a k value, the TR test is

$$TR(k) = \frac{\hat{\gamma}_{2,1}(k)}{\widehat{Var}[\hat{\gamma}_{2,1}(k)]^{1/2}},$$

which is asymptotically distributed as a $N(0,1)$ under the null of time reversibility. The prerequisite of the test is that the data must posses finite sixth moment, but such a condition may be too restrictive for financial data. If the distribution lacks finite sixth moment, the size of RR test can be seriously distorted.

2.2 Chen, Chou and Kuan TR test

In a recent paper, Chen and Kuan (2002) propose using the time reversibility test of Chen, Chou and Kuan (2000), CCK hereafter, to test for misspecification of asymmetric volatility models. A salient feature of this test is that it does not have any moment restrictions, contrary to the RR test which requires finite sixth moment.

Given a stationary time reversible time series, the CCK test is based on the implication that the differences of the series being tested have symmetric marginal distributions, i.e., if x_t is time reversible, the distribution of $z_{t,k} = x_t - x_{t-k}$ is symmetric about the origin.

To test the null hypothesis, we employ the test statistic

$$\frac{\sqrt{T-k}}{\hat{\sigma}_{g,k}}\overline{\Psi}_{g,k}$$

which is asymptotically distributed as a $N(0,1)$ under the null, where

$$\overline{\Psi}_{g,k} = \frac{1}{T-k}\sum \Psi_g(z_{t,k})$$

and $\hat{\sigma}_{g,k}^2$ is a consistent estimator of the variance (for further details, see Chen et al., 2000). The proposed test is a class of tests, depending on the weighting function g. In the present work, we use the exponential density function; then,

$$\Psi_g\left(z\right) = \frac{\beta z}{1+\left(\beta z\right)^2}.$$

However, as Chen (2001) and Chen and Kuan (2002) remark, the CCK test is not directly applicable to model residuals because it is a test of unconditional symmetry. Thus, we follow Chen (2001) and modify the computation of the variance of the statistic, by means of simple bootstrap from the standardised residual series.

3 BL and TAR models

Bilinear and threshold autorregresive processes are popular nonlinear models (for a detailed discussion, see Tong, 1990). As Rothman (1998) notes, bilinear models can capture a wide range of nonlinear behavior, whereas TAR models are specially suited to account for asymmetric limit cycle behavior.

In the bilinear case, we specify the data generating process as:

$$X_t = \gamma \, X_{t-1}\varepsilon_{t-1} + \varepsilon_t$$

for $\gamma = 0.9, 0.8, \ldots, 0.1$. Since the autocovariance function of this model resembles the autocovariance function of an MA(1) process, TR tests power is computed against MA(1) residuals.

In the TAR case, we simulate realizations from the model:

$$X_t = \begin{cases} \alpha \, X_{t-1} + \varepsilon_t & \text{if } X_{t-1} \geq 1, \\ -0.4 \, X_{t-1} + \varepsilon_t & \text{if } X_{t-1} < 1, \end{cases}$$

where $\varepsilon_t \sim N(0,1)$, for $\alpha = -0.9, -0.8, \ldots, -0.1$. As Rothman (1998) points out, the autocorrelation and partial autocorrelation functions of this model is the same as that of an AR(1) process. Therefore, the power of the TR tests are computed against AR(1) residuals. Note that when $\alpha = -0.4$, the null hypothesis is true. In that case, we are also computing the empirical size of the tests.

We have performed 1,000 replicates for each parameter value. RR and CCK tests have been computed on the same data sets. For each replication, the corresponding AR(1) or MA(1) model has been estimated by Maximum Likelihood. When applying the CCK test, we have used 200 simple bootstrap samples from the corresponding estimated residuals, and we have computed the empirical variance of the statistic. Rejection rates at the two-sided 5% significance level are shown in Tables 1–4.

Insert Tables 1–4

In all cases, we have found that the RR test is more powerful than CCK test. Therefore, we could claim that using just the RR test is preferable to detect time irreversibility due to TAR or BL-type processes.

In order to asses the empirical size of the tests against correctly specified MA processes, we have performed simulations of the model:

$$X_t = 0.5 \, \varepsilon_{t-1} + \varepsilon_t.$$

Results in columns labeled 'size' allow us to claim that the RR and CCK tests are not oversized under correctly specified AR(1) or MA(1) models.

4 EGARCH and GJR models

In a large amount of papers, the so-called BDS test of Brock et al. (1996) is applied in order to assess the adequacy of fitted GARCH models. Frequently, the null of independence and identical distribution is accepted whenever a symmetric ARCH or GARCH model is fitted. However, as Brooks and Henry (2000) and Chen and Kuan (2002) show, the BDS test is not useful in detecting deviations from the iid hypothesis in the case of misspecified GARCH models.

While Brooks and Henry (2000) show that size and bias sign statistics, designed by Engle and Ng (1993), are much more powerful than the BDS test, Chen and Kuan (2002) suggest using the CCK test to detect misspecification in GARCH modelling. In the present work, we analyze the power of RR and CCK tests when testing for time reversibility against the standardised GARCH residuals from asymmetric volatility models.

As in Brooks and Henry (2000), two asymmetric models are considered. The first model is the EGARCH(1,1) or exponential GARCH:

$$\begin{aligned} X_t &= u_t,\, u_t \sim N(0, h_t) \\ h_t &= \alpha_0 + \beta_1 \log(h_{t-1}) + \delta_1 \left(\frac{u_{t-1}}{\sqrt{h_{t-1}}} \right) + a_1 \left[\frac{|u_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right]. \end{aligned}$$

The second model is the GJR or threshold GARCH:

$$\begin{aligned} X_t &= u_t,\, u_t \sim N(0, h_t) \\ h_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \delta \xi_{t-1}^2 \end{aligned}$$

where $\xi_{t-1} = \min\{0, u_{t-1}\}$. In both cases, we have used the parameter values employed in the Monte Carlo study of Brooks and Henry (2000).

We have performed 1,000 replications for each model and parameter setting, and RR and CCK tests have been computed on the same data sets. For each replication, a GARCH(1,1) model has been estimated by Maximum Likelihood. When we apply the CCK test, we perform 200 simple bootstrap samples from the corresponding standardised residuals, and we compute the empirical variance of the statistic. Rejection rates at the two-sided 5% significance level are shown in Tables 5–8.

Insert Tables 5–8

As in the previous section, we have found that the RR test is more powerful than CCK test. Thus, we could claim that the RR test is more suitable to detect time irreversibility due to asymmetric volatility.[1]

[1]In addition, we have also analyzed the behavior of TR tests using the parameter values employed in Chen and Kuan (2002) for the EGARCH model. Results, which are available upon request, are also favourable to the RR test.

However, let us recall that the asymptotic distribution of the RR test relies on the existence up to the sixth population moment, whereas the CCK test computed on estimated residuals from conditional variance models require the error terms to possess just finite fourth moment (see Chen and Kuan, 2002). Since many studies provide evidence that financial data lack higher moments, it is preferable to test for TR using the CCK test despite its lower power.

In addition, we have computed the empirical size of both procedures against a correctly specified GARCH(1,1) process:

$$\begin{aligned} X_t &= u_t,\, u_t \sim N(0, h_t) \\ h_t &= 0.0005 + 0.90\, h_{t-1} + 0.05\, u_{t-1}^2. \end{aligned}$$

Note that the parameters values have been chosen to resemble the high persistence in variance usually found in empirical research. Rejection rates have been computed at two-sided 5% significance level just for $T = 1000$. Results displayed in columns labeled 'size' in Tables 5 and 7, show that the RR and CCK tests are not oversized under correctly symmetric GARCH(1,1) models. Results for other sample sizes are similar.

5 Combination of tests

In this section we investigate whether the power properties of the time reversibility testing may be improved by means of the combination of both the CCK and RR tests. To that end, we have computed the two-tailed 5% rejection rates for the bilinear model with $\gamma = 0.1$, when the rejection rule is "reject the null hypothesis if either the RR test or the CCK test rejects the null". Results are displayed in Table 9.

Insert Table 9

The main conclusion from Table 9 is that power is larger than using each test individually. However, it is well known that if the null hypothesis is true, the final rejection rates from the application of several statistical tests is larger than the nominal size. This well known result is illustrated in Table 10. The rejection rates have been computed for the combined test when the true data generating process is an MA(1) as described in section 3

Insert Table 10

In order to deal with this perverse effect, we suggest to apply the so-called p-value adjustments for multiple tests. This idea has been successfully applied by Psaradakis (2000) in the context of sequential nonlinearity testing.

We compute the Sidack-adjusted p-value as:

$$\tilde{p}_i^{(S)} = 1 - (1 - p_i)^m,$$

$i = 1, \ldots, m$, where p_i is the p-value corresponding to the time reversibility test i, and m is the number of statistical tests. The Hochberg (1988) adjusted p-values are obtained as:

$$\tilde{p}_i^{(H)} = \min\{[m - R(p_i) + 1]p_i, 1\},$$

where $R(p_i$ is the rank of the ith p-value. Given a significance level α, the decision rule states that the null is rejected if $\tilde{p}^{(S)} = \min_{1\leq i\leq m} \tilde{p}_i^{(S)} \leq \alpha$ or $\tilde{p}^{(H)} = \min_{1\leq i\leq m} \tilde{p}_i^{(H)} \leq \alpha$.

Next Tables show the rejection rates by using Sidack and Hochberg p-value adjustments, respectively.

Insert Tables 11–14

These new results support the use of p-value adjustments for multiple tests, since the size of the tests sequence is close to the nominal 5%. On the other hand, the power is larger than the power of the CCK test, but it is below the power of the RR test.

6 Conclusions

In this paper, we have compared the power of the TR tests of Ramsey and Rothman (1996) and Chen et al. (2000), in order to detect deviations from the null hypothesis of independence and identical distribution, whenever misspecified models are fitted. To that end, a large Monte Carlo study has been performed.

In the case of misspecification in mean, two common nonlinear processes have been analysed: BL and TAR models. In the case of misspecification in variance, two asymmetric conditionally heteroskedastic processes have been analysed: EGARCH and GJR models.

The results show that, in all considered cases, the RR test is more powerful than the CCK test. However, as it was stressed in section 4, RR test requires the existence of higher moments than the CCK test, to guarantee the asymptotic distribution of the test. Hence, it would be advisable to previously estimate the maximum moment exponent of the (standardised) residuals, as in Chen and Kuan (2002). If the estimate is larger than five, the RR test seems the correct choice. Otherwise, as is usual in financial data, the CCK test is preferable.

Finally, we have investigated the usefulness of a combination of both tests. Results show that if the tests are sequentially computed, the empirical size is larger than the nominal size, as expected. Thus, we suggest to apply p-value adjustments for multiple tests as suggested by Psaradakis (2000) in the context of nonlinearity testing. Applying such adjustments we find that the empirical size is closed to the nominal significance level, and that the power is larger than using just the CCK test, but it is lower than applying just the RR test.

7 References

Brock, W.A., Dechert, W.D., Scheinckman, J.A., LeBaron, B., 1996. A test for independence based on the correlation dimension. Econometric Reviews 15, 197–235.

Brooks, C., Henry, Ó.T., 2000. Can portmanteau nonlinearity tests serve as general misspecification tests? Evidence from symmetric and asymmetric GARCH models. Economics Letters 67, 245–251.

Chen, Y.-T., Kuan , C.-M., 2002. Time irreversibility and EGARCH effects in US stock index returns. Journal of Applied Econometrics 17, 565–578.

Chen, Y.-T., 2001, Testing conditional symmetry with an application to financial returns. Working paper, Institute for Social Sciences and Philosophy, Academia Sinica, Taiwan.

Chen, Y.-T., Chou, R.-Y., Kuan , C.-M., 2000. Testing time reversibility without moment restrictions. Journal of Econometrics 95, 199–218.

Engle, R.F., Ng, V., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

Hochberg, Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800-802.

Psaradakis, Z. 2000. p-Value adjustments for multiple tests for nonlinearity. *Studies in Nonlinear Dynamics and Econometrics*, Vol. 4, Issue 3, Article 1.

Ramsey, J.B., Rothman, P., 1996. Time irreversibility and business cycle asymmetry. Journal of Money, Credit and Banking 28, 3–20.

Rothman, P., 1998. Higher-order residual analysis for simple bilinear and threshold autoregressive models with the TR test. In Rothman, P. (ed.) Nonlinear Time Series Analysis of Economic and Financial Data, Kluwer Academic Press.

Table 1: Estimated rejection rates of RR test at lag k against MA(1) residuals from Bilinear models at the 5% significance level

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	size
$T = 100$										
$k = 1$	0.160	0.421	0.636	0.657	0.554	0.448	0.393	0.388	0.432	0.059
$k = 2$	0.063	0.062	0.070	0.110	0.192	0.271	0.333	0.382	0.409	0.048
$k = 3$	0.051	0.057	0.057	0.056	0.059	0.077	0.099	0.134	0.176	0.059
$k = 4$	0.050	0.046	0.049	0.057	0.045	0.053	0.060	0.069	0.110	0.052
$k = 5$	0.061	0.062	0.053	0.047	0.041	0.031	0.041	0.051	0.072	0.040
$T = 250$										
$k = 1$	0.321	0.812	0.936	0.898	0.744	0.530	0.437	0.500	0.592	0.052
$k = 2$	0.045	0.057	0.085	0.176	0.307	0.450	0.569	0.608	0.620	0.050
$k = 3$	0.059	0.061	0.071	0.068	0.071	0.118	0.176	0.255	0.323	0.062
$k = 4$	0.056	0.051	0.047	0.052	0.051	0.051	0.079	0.125	0.178	0.054
$k = 5$	0.052	0.046	0.040	0.039	0.038	0.029	0.044	0.067	0.101	0.054
$T = 500$										
$k = 1$	0.582	0.981	0.994	0.973	0.836	0.567	0.462	0.599	0.670	0.050
$k = 2$	0.045	0.057	0.107	0.247	0.470	0.651	0.749	0.771	0.748	0.054
$k = 3$	0.062	0.062	0.060	0.048	0.062	0.141	0.244	0.379	0.455	0.046
$k = 4$	0.061	0.060	0.058	0.056	0.054	0.066	0.104	0.190	0.263	0.065
$k = 5$	0.055	0.054	0.044	0.038	0.031	0.030	0.048	0.088	0.148	0.053

The first arrow displays the values of the γ parameter. Column labeled 'size' displays the empirical rejection rates against a correctly specified MA(1) process.

Table 2: Estimated rejection rates of RR test at lag k, against AR(1) residuals from TAR models at the 5% significance level

	-0.1	-0.2	-0.3	-0.5	-0.6	-0.7	-0.8	-0.9	size
$T = 100$									
$k = 1$	0.099	0.066	0.058	0.068	0.093	0.125	0.181	0.242	0.065
$k = 2$	0.058	0.053	0.056	0.057	0.060	0.060	0.059	0.063	0.057
$k = 3$	0.057	0.064	0.061	0.065	0.066	0.063	0.062	0.067	0.062
$k = 4$	0.053	0.053	0.052	0.047	0.050	0.049	0.051	0.051	0.050
$k = 5$	0.042	0.043	0.043	0.048	0.051	0.046	0.045	0.049	0.046
$T = 250$									
$k = 1$	0.182	0.124	0.084	0.081	0.130	0.217	0.310	0.424	0.057
$k = 2$	0.054	0.050	0.048	0.049	0.059	0.063	0.074	0.079	0.052
$k = 3$	0.048	0.050	0.048	0.046	0.045	0.049	0.045	0.064	0.047
$k = 4$	0.052	0.051	0.048	0.055	0.058	0.058	0.054	0.051	0.050
$k = 5$	0.057	0.056	0.056	0.054	0.055	0.054	0.048	0.047	0.054
$T = 500$									
$k = 1$	0.319	0.175	0.091	0.078	0.177	0.330	0.500	0.685	0.054
$k = 2$	0.062	0.060	0.062	0.056	0.063	0.068	0.083	0.094	0.059
$k = 3$	0.049	0.046	0.045	0.043	0.050	0.050	0.055	0.062	0.043
$k = 4$	0.049	0.050	0.047	0.052	0.052	0.056	0.061	0.063	0.048
$k = 5$	0.060	0.057	0.057	0.058	0.060	0.058	0.059	0.065	0.058

The first arrow displays the values of the ρ_1 parameter. Column labeled 'size' displays the empirical size of the test against a correctly specified AR(1) process (TAR model with $\rho_1 = -0.4$).

Table 3: Estimated rejection rates of CCK test at lag k against MA(1) residuals from Bilinear models at the 5% significance level

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	size
$T = 100$										
$k = 1$	0.133	0.309	0.462	0.484	0.396	0.242	0.137	0.085	0.092	0.059
$k = 2$	0.054	0.055	0.064	0.069	0.096	0.129	0.169	0.204	0.232	0.048
$k = 3$	0.061	0.058	0.057	0.047	0.039	0.041	0.048	0.076	0.089	0.059
$k = 4$	0.049	0.048	0.052	0.052	0.054	0.049	0.057	0.062	0.067	0.052
$k = 5$	0.046	0.047	0.051	0.049	0.049	0.048	0.046	0.059	0.067	0.040
$T = 250$										
$k = 1$	0.228	0.648	0.847	0.888	0.791	0.531	0.232	0.098	0.099	0.052
$k = 2$	0.046	0.046	0.064	0.107	0.177	0.291	0.403	0.486	0.534	0.050
$k = 3$	0.053	0.055	0.056	0.054	0.060	0.057	0.057	0.076	0.114	0.062
$k = 4$	0.057	0.052	0.051	0.050	0.051	0.047	0.058	0.066	0.072	0.054
$k = 5$	0.044	0.048	0.039	0.045	0.044	0.041	0.048	0.048	0.061	0.054
$T = 500$										
$k = 1$	0.445	0.934	0.990	0.996	0.973	0.823	0.412	0.122	0.113	0.050
$k = 2$	0.047	0.058	0.103	0.175	0.317	0.502	0.666	0.762	0.805	0.054
$k = 3$	0.051	0.048	0.054	0.068	0.066	0.071	0.074	0.115	0.178	0.046
$k = 4$	0.068	0.067	0.055	0.058	0.053	0.046	0.063	0.068	0.103	0.065
$k = 5$	0.047	0.052	0.046	0.042	0.054	0.044	0.052	0.050	0.067	0.053

The first arrow displays the values of the γ parameter. Column labeled 'size' displays the empirical size of the test against a correctly specified MA(1) process.

Table 4: Estimated rejection rates of CCK test at lag k, against AR(1) residuals from TAR models at the 5% significance level

	-0.1	-0.2	-0.3	-0.5	-0.6	-0.7	-0.8	-0.9	size
$T = 100$									
$k = 1$	0.090	0.067	0.061	0.069	0.076	0.102	0.137	0.169	0.060
$k = 2$	0.068	0.075	0.072	0.062	0.062	0.059	0.063	0.059	0.070
$k = 3$	0.039	0.044	0.045	0.051	0.054	0.056	0.061	0.061	0.047
$k = 4$	0.049	0.052	0.057	0.056	0.060	0.059	0.056	0.066	0.057
$k = 5$	0.044	0.041	0.040	0.046	0.052	0.045	0.047	0.047	0.043
$T = 250$									
$k = 1$	0.182	0.119	0.085	0.069	0.108	0.161	0.249	0.342	0.057
$k = 2$	0.055	0.051	0.053	0.059	0.063	0.068	0.071	0.082	0.059
$k = 3$	0.046	0.044	0.041	0.039	0.041	0.045	0.054	0.051	0.042
$k = 4$	0.054	0.053	0.055	0.051	0.052	0.054	0.055	0.056	0.053
$k = 5$	0.050	0.047	0.046	0.051	0.059	0.062	0.062	0.053	0.047
$T = 500$									
$k = 1$	0.296	0.166	0.082	0.087	0.156	0.263	0.403	0.565	0.058
$k = 2$	0.046	0.049	0.049	0.057	0.063	0.063	0.075	0.086	0.056
$k = 3$	0.055	0.047	0.054	0.050	0.049	0.042	0.040	0.048	0.054
$k = 4$	0.044	0.044	0.049	0.048	0.048	0.052	0.053	0.053	0.048
$k = 5$	0.052	0.051	0.049	0.050	0.045	0.048	0.047	0.049	0.050

The first arrow displays the values of the ρ_1 parameter. Column labeled 'size' displays the empirical size of the test against a correctly specified AR(1) process (TAR model with $\rho_1 = -0.4$).

Table 5: Estimated rejection rates of RR test at lag k against GARCH(1,1) residuals from EGARCH(1,1) models at the 5% significance level

	$T = 250$	$T = 500$	$T = 1000$	size ($T = 1000$)
$k = 1$	0.075	0.123	0.202	0.057
$k = 2$	0.066	0.088	0.134	0.060
$k = 3$	0.067	0.087	0.118	0.047
$k = 4$	0.059	0.059	0.082	0.053
$k = 5$	0.061	0.063	0.062	0.049

Column labeled 'size' displays rejection rates against a correctly specified GARCH(1,1) process.

Table 6: Estimated rejection rates of RR test at lag k against GARCH(1,1) residuals from GJR models at the 5% significance level

	$T = 250$	$T = 500$	$T = 1000$
$k = 1$	0.125	0.247	0.404
$k = 2$	0.083	0.150	0.247
$k = 3$	0.087	0.125	0.170
$k = 4$	0.077	0.085	0.121
$k = 5$	0.077	0.074	0.084

Table 7: Estimated rejection rates of CCK test at lag k against GARCH(1,1) residuals from EGARCH(1,1) models at the 5% significance level

	$T = 250$	$T = 500$	$T = 1000$	size ($T = 1000$)
$k = 1$	0.084	0.114	0.172	0.059
$k = 2$	0.072	0.083	0.110	0.069
$k = 3$	0.056	0.062	0.098	0.049
$k = 4$	0.051	0.067	0.074	0.057
$k = 5$	0.053	0.058	0.053	0.047

Column labeled 'size' displays rejection rates against a correctly specified GARCH(1,1) process.

Table 8: Estimated rejection rates of CCK test at lag k against GARCH(1,1) residuals from GJR models at the 5% significance level

	$T = 250$	$T = 500$	$T = 1000$
$k = 1$	0.100	0.180	0.314
$k = 2$	0.078	0.115	0.180
$k = 3$	0.067	0.091	0.128
$k = 4$	0.061	0.081	0.097
$k = 5$	0.061	0.069	0.078

Table 9: Estimated rejection rates of combined test at lag k against MA(1) residuals from bilinear model at the 5% significance level

	$T = 100$	$T = 250$	$T = 500$
$k = 1$	0.205	0.365	0.621
$k = 2$	0.103	0.077	0.090
$k = 3$	0.094	0.098	0.091
$k = 4$	0.091	0.100	0.115
$k = 5$	0.092	0.093	0.089

Table 10: Estimated rejection rates of combined test at lag k against MA(1) residuals from a correctly specified MA(1) model at the 5% significance level

	$T = 100$	$T = 250$	$T = 500$
$k = 1$	0.085	0.077	0.080
$k = 2$	0.088	0.076	0.073
$k = 3$	0.097	0.100	0.087
$k = 4$	0.107	0.093	0.101
$k = 5$	0.083	0.090	0.091

Table 11: Estimated rejection rates using Sidack-corrected p-values at lag k against MA(1) residuals from bilinear model at the 5% significance level

	$T = 100$	$T = 250$	$T = 500$
$k = 1$	0.137	0.269	0.515
$k = 2$	0.051	0.044	0.046
$k = 3$	0.050	0.055	0.050
$k = 4$	0.051	0.054	0.049
$k = 5$	0.053	0.056	0.061

Table 12: Estimated rejection rates using Sidack-corrected p-values at lag k against MA(1) residuals from a correctly specified MA(1) model at the 5% significance level

	$T = 100$	$T = 250$	$T = 500$
$k = 1$	0.049	0.041	0.046
$k = 2$	0.052	0.038	0.041
$k = 3$	0.059	0.049	0.041
$k = 4$	0.053	0.053	0.048
$k = 5$	0.045	0.051	0.055

Table 13: Estimated rejection rates using Hochberg-corrected p-values at lag k against MA(1) residuals from bilinear model at the 5% significance level

	$T = 100$	$T = 250$	$T = 500$
$k = 1$	0.141	0.282	0.526
$k = 2$	0.058	0.043	0.051
$k = 3$	0.051	0.056	0.053
$k = 4$	0.053	0.056	0.052
$k = 5$	0.054	0.056	0.060

Table 14: Estimated rejection rates using Hochberg-corrected p-values at lag k against MA(1) residuals from a correctly specified MA(1) model at the 5% significance level

	$T = 100$	$T = 250$	$T = 500$
$k = 1$	0.052	0.044	0.049
$k = 2$	0.056	0.040	0.045
$k = 3$	0.061	0.052	0.046
$k = 4$	0.053	0.054	0.049
$k = 5$	0.048	0.053	0.056