Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 45 (3) 529-533 (2013)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Diagnostics for GARCH-Type Models under Symmetric and Asymmetric Errors

F. IQBAL[++], Y. Z. JAFRI, G. H. TALPUR*

Department of Statistics, University of Balochistan, Quetta – Pakistan

Received 21st December 2012 and Revised 28th May 2013

Abstract: In this paper, the size and power of Ljung-Box and Li-Mak diagnostic tests for univariate autoregressive conditional heteroscedastic models were studied under both symmetric and asymmetric distributions for errors. Monte Carlo simulations with 1000 independent replications are conducted to generate conditional variances with standard normal, Students-*t* and Skewed-*t* distributions. It was found that though the Li-Mak test has higher empirical size than the nominal size of 5% but can be considered a better alternative to the Ljung-Box test in case of asymmetric errors. The empirical power of the Li-Mak test was also found slightly better for asymmetric heavy-tailed data.

Keywords: ARCH-GARCH models, Portmanteau tests, Autocorrelation

1. INTRODUCTION

The autoregressive conditional heteroscedastic (ARCH) model of Engle (1982) and the generalized ARCH (GARCH) model of Bollerslev (1986) have been found to be successful in capturing the volatility or the conditional variance structure of many financial time series. There is a huge literature on modelling these conditional heteroscedastic time series, but not much work has been done on model checking or model selection. Diagnostic is one of the important stages of model building. Residual autocorrelations are used to identify possible departure from the assumption that the white noise disturbances in the specified model are uncorrelated (see Box and Jenkins, 1970).

To check the model adequacy, the asymptotic distribution of the squared and absolute residual autocorrelations derived from such models might be useful. One option is to build a test statistic to test the null hypothesis that the residuals are independent up to a lag *M*. The test statistic can be applied to check for non-linearity in mean and also for nonlinearity in variance. The test statistics usually used are called portmanteau statistics.

One of the widely used portmanteau statistic is the one proposed by Box and Pierce (1970). This statistic is used to test the null hypothesis that the first *M* autocorrelations of a covariance stationary time series are zero. If significant autocorrelation is not found in the residuals from the model, then the model is declared to be adequate. Ljung and Box (1978) discussed the finite sample properties and conservative behaviour of the Box-Pierce statistic.

In financial time series analysis, it is particularly important to check serial correlations of squared series. McLeod and Li (1983) derived a portmanteau test for model adequacy based on the squared residual autocorrelations in ARMA models.

In practice, many researchers apply the Ljung-Box or McLeod-Li tests to the squares of the estimated standardised residuals when testing the adequacy of an ARCH/GARCH model. A χ^2 distribution with M degrees of freedom, as the large sample distribution for these statistics is found misleading and using the squared residual autocorrelations a correct portmanteau test is proposed by Li and Mak (1994).

Tse and Zuo (1997) reported some Monte Carlo results for the finite sample performance of some commonly used diagnostics used in literature and found that the Li-Mak test based on the asymptotic variance under the Gaussian assumption performs favourably among other versions of statistics. Tsui (2004) through Monte Carlo simulation studied the empirical size and power of various tests and found that Li-Mak diagnostics is more powerful.

Jianhong and Lixing (2009) proposed a new approach for checking the adequacy of GARCH-type models. Their tests involved weight functions, which provide them with the flexibility in choosing scores to enhance power performance. Carbon and Francq (2011) derived the asymptotic distribution of a vector of autocorrelation of squared residuals for asymmetric power GARCH models. Portmanteau tests are derived and results are obtained under weak moment assumptions.

In this paper, we aim to study the empirical size and power of two important diagnostic tests; the Ljung-Box and Li-Mak tests used for univariate autoregressive conditional heteroscedastic models under

[++]Corresponding author email: F. IQBAL farhatiqb@gmail.com
*Department of Statistics, University of Sindh, Jamshoro –Pakistan.

both symmetric and asymmetric distributions for errors. For this purpose, Monte Carlo simulations are conducted to generate conditional variances from ARCH/GARCH model. The errors are generated form the symmetric (standard normal), heavy-tailed (Students-*t*) and asymmetric heavy-tailed (Skewed-*t* distributions). Our results revealed that under asymmetric errors the performance of Li-Mak test was found slightly better than the Ljung-Box test. It is suggested that these tests may be used with caution when data exhibits asymmetry and heavy-tailedness.

This paper is organized as follows: Section 2 introduces the GARCH model. In Section 3, various diagnostic statistics used for the checking of GARCH-Type models are presented. In Section 4, the results of Monte Carlo simulations are presented and finally Section 5 concludes the paper.

2. The GARCH-Type Models

In this section, we introduce the family of GARCH models. Both symmetric and asymmetric types of models are considered.

In the GARCH*(p,q)* model, the following representation of the series $\{X_t; t \in \mathbb{Z}\}$ is assumed. Observe $\{X_t; 1 \le t \le T\}$ such that

$X_t = h_t^{1/2}\epsilon_t$ with

$$h_t = \omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j}\, h_{t-j},$$

where $p > 0, q \ge 0, \omega_0 > 0, 0 \le \alpha_{0i} < 1$ for $i = 1, \ldots, p, 0 \le \beta_{0j} < 1$ for $j = 1, \ldots, q,$ and $\{\epsilon_t; 1 \le t \le T\}$ are unobservable i.i.d. errors symmetric about zero. When $q = 0$, the GARCH model reduces to the ARCH model.

Let $\boldsymbol{\theta}$ be the vector of parameters capturing h_t. The estimate of $\boldsymbol{\theta}$ denoted by $\widehat{\boldsymbol{\theta}}$ can be obtained using the quasi-maximum likelihood method. We denote $\hat{h}_t$ as value of h_t evaluated at $\widehat{\boldsymbol{\theta}}$, and $\hat{\epsilon}_t = X_t/\sqrt{h_t}$ as standardized residuals.

Usually a GARCH(1,1) model is adequate to obtain a good model fit for financial time series. Standard GARCH models assume that positive and negative values of past observations have a symmetric effect on the volatility. In other words, good and bad news have the same effect on the volatility in the GARCH model and the sign of the shock is irrelevant. In practice this assumption is frequently violated, in particular by stock returns, in that the volatility increases more after bad news than after good news. This is so called the *leverage effect*. However, much applied research is still conducted assuming implicitly the existence of symmetric dynamics, which may lead to model misspecification if dynamic asymmetry is indeed present. The GARCH model is not a suitable choice for modelling the asymmetric effect in the returns.

To capture the asymmetric effect of positive and negative shocks on volatility, a few variants of the basic GARCH model have been proposed. One such model is presented by Glosten, Jaganathan and Runkle (1993). This asymmetric GARCH model commonly known as GJR model is capable of capturing the asymmetric feature.

3. Portmanteau Statistics

In this section we discuss some of the frequently used statistics in time series for diagnostic checking.

3.1 Box-Pierce Statistic

The Box-Pierce statistic is used to test the null hypothesis that the first M autocorrelations of a covariance stationary time series are zero. Under the assumption that the observations are independent and identically distributed, the asymptotic covariance matrix of the vector of sample autocorrelations is the inverse of the sample size times the identity matrix. This test is generally called the classical portmanteau statistic. The lag-*k* residual autocorrelation is defined as:

$$\hat{r}_{1k} = \frac{\sum_{t=k+1}^{T}(\hat{\epsilon}_t - \bar{\epsilon})(\hat{\epsilon}_{t-k} - \bar{\epsilon})}{\sum_{t=1}^{T}(\hat{\epsilon}_t - \bar{\epsilon})^2} \quad \text{for} \quad k = 1,2,\ldots,M,$$

where $\{\hat{\epsilon}_t\}$ are residuals from an autoregressive moving average, ARMA*(p,q)* model, $\bar{\epsilon} = \frac{1}{T}\sum \hat{\epsilon}_t$ and T is the sample size. The Box-Pierce statistic is defined as:

$$Q_{BP}(M) = T\sum_{k=1}^{M} \hat{r}_{1k}^2$$

where $\hat{r}_{1k}^2$ is the sample residual autocorrelation of order *k=1,...,M*. Under the null hypothesis that ARMA*(p,q)*, model is adequate, Q_{BP} is asymptotically distributed as a χ^2 with *(M-p-q)* degrees of freedom.

3.2 Ljung-Box Statistic

A modified test proposed by Ljung and Box is

$$Q_{LB}(M) = T(T+2)\sum_{k=1}^{M}\frac{\hat{r}_{1k}^2}{T-k}.$$

It has been shown that the finite sample distribution of this statistic is much closer to that of the $\chi^2_{(M-p-q)}$, however its variance could be substantially larger than that of its asymptotic distribution.

3.3 McLeod-Li Statistic

The lag-*k* squared residual autocorrelation is defined as:

$$\hat{r}_{2k} = \frac{\sum_{t=k+1}^{T}(\hat{\epsilon}_t^2 - \bar{\epsilon})(\hat{\epsilon}_{t-k}^2 - \bar{\epsilon})}{\sum_{t=1}^{T}(\hat{\epsilon}_t^2 - \bar{\epsilon})^2} \quad \text{for } k = 1,2,\dots,M,$$

where $\bar{\epsilon} = \frac{1}{T}\sum\hat{\epsilon}_t^2$ and T is the sample size. The McLeod-Li statistic is

$$Q_{ML}(M) = T(T+2)\sum_{k=1}^{M}\frac{\hat{r}_{2k}^2}{T-k}.$$

They showed that, if the eighth order moment of the returns exists, $Q_{ML}(M)$ is distributed asymptotically as χ_M^2. This test is asymptotically equivalent to the Langrange Multiplier (LM) test of Engle (1982). When the $Q_{ML}(M)$ statistic is implemented with absolute values, only the fourth order moment of returns should be finite for the asymptotic distribution to hold.

3.4 Li-Mak Statistic

Li and Mak (1994) derived the asymptotic variance of the correlation coefficients, and suggested some diagnostics for the ARCH/GARCH models. The lag-*k* correlation coefficient $\hat{r}_{3k}$ is defined as:

$$\hat{r}_{3k} = \frac{\sum_{t=k+1}^{T}(\hat{\epsilon}_t^2 - 1)(\hat{\epsilon}_{t-k}^2 - 1)}{\sum_{t=1}^{T}(\hat{\epsilon}_t^2 - 1)^2},$$

where $\hat{\epsilon}_t$ are the standardised residuals from GARCH model estimated by QMLE. Li and Mak (1994) showed that $\sqrt{T}\hat{\mathbf{r}}_3$ is asymptotically normally distributed with mean **0** and covariance matrix **V**, where $\hat{\mathbf{r}}_3$ denotes the vector of sample correlation coefficients defined by $\hat{\mathbf{r}}_3 = (\hat{r}_{31}, \hat{r}_{32}, \dots, \hat{r}_{3M})'$ and **V** can be consistently estimated by $\hat{\mathbf{V}} = \mathbf{I}_M - \left(\frac{1}{4}\right)\mathbf{X}\hat{\mathbf{G}}^{-1}\mathbf{X}'$ where $\mathbf{I}_M$ is the *M* x *M* identity matrix, $\hat{\mathbf{G}}^{-1}$ is a consistent estimate of the asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta_0)$ and $\mathbf{X} = (X_1, \dots, X_M)'$ with

$$\hat{X}_k = -\frac{1}{T}\sum_{t=k+1}^{T}(\hat{\epsilon}_{t-k}^2 - 1)\frac{\nabla_\theta \hat{h}_t}{\hat{h}_t},$$

$\hat{h}_t$ is the estimate of the conditional variance of the GARCH model and $\nabla_\theta f(\theta)$ represents the derivative of $f(\theta)$ w.r.t θ.

The Li-Mak statistic is

$$Q_{LM}(M) = T\,\hat{\mathbf{r}}_3'\hat{\mathbf{V}}^{-1}\hat{\mathbf{r}}_3.$$

If the model is correct, $Q_{LM}(M)$ asymptotically follows χ^2 distribution with *M* degrees of freedom.

4. RESULTS AND DISCUSSION

In this section, we check the performance of two widely used diagnostic tests; the Ljung-Box (Q_{LB}) and Li-Mak (Q_{LM}) tests under both symmetric and asymmetric errors. Monte Carlo simulations are used to examine the empirical size and power of the portmanteau statistics (Q_{LB}) and (Q_{LM}). We use 1000 replications and the sample size of *T*=100, 500 and 1000 for all experiments. All simulations are performed using MATLAB software.

The data are generated from the following two data generating processes (DGPs), denoted by M1 for GARCH(1,1) and M2 for ARCH(2) model:

M1: $X_t = h_t^{1/2}\epsilon_t, \quad h_t = 0.5 + 0.3\,X_{t-1}^2 + 0.2\,h_{t-1}.$

M2: $X_t = h_t^{1/2}\epsilon_t, \quad h_t = 0.1 + 0.8\,X_{t-1}^2 + 0.1\,X_{t-2}^2.$

These data generating processes (DGPs) are used to exhibit low and high volatile series. Other DGPs with different parameters values may also be used. For each DGP, the errors ($\hat{\epsilon}_t$) are generated from the standard normal distribution (*N(0,1)*), student-*t* distribution with 3 degrees of freedom (*t(3)*) and from Hansen's skewed-*t* distribution (Hansen, 1994) with 4 degrees of freedom and skewness parameter of 0.25 (*ST(4,0.25)*). The skewed-*t* distribution is chosen to observe the effect of asymmetry on the size and power of these tests.

The GARCH(1,1) and ARCH(2) models are fitted using maximum likelihood method assuming normal errors for each DGPs and the portmanteau statistics (Q_{LB}) and (Q_{LM}) are computed.. The resulting estimators are called quasi-maximum likelihood estimators (QMLEs). For this study, we choose *M* = 3, 6, and 10 as the magnitude of *M* is independent of the validity of the diagnostics. As a benchmark for comparison, the nominal size of 5% is used.

The rejection frequency represents the estimated size of the tests when the underlying estimated models (EMs) are GARCH(1,1) and ARCH(2) for DGPs M1 and M2, respectively. For empirical power, the ARCH(2) and GARCH(1,1) models are fitted to DGPs M1 and M2, respectively. Although, like residual-based diagnostics, these portmanteau tests have no specific alternatives, it should not be construed that the portmanteau tests are consistent against all model misspecifications.

(Table 1) shows the empirical frequency of rejection in percentage (empirical sizes) of diagnostic tests for ARCH/GARCH models. It can be seen that both tests generally have reasonable reliable size under the normal errors with Li-Mak test under rejecting the null hypothesis under standard normal errors for *M*=3. Tse and Zuo (1997) also found in their simulation study that the Li-Mak and Box-Pierce tests under reject the null hypothesis significantly for small values of *M*. Our study is different from Tse and Zuo (1997) in the sense that we also examined the performance of these tests under asymmetric heavy-tailed distribution.

For students-*t (3)*, both tests slightly over-rejected the null hypothesis with Li-Mak test showing better results as compared to Ljung-Box. Tse (2000) also noted that Li-Mak test generally has reliable size although there is slight tendency to over-reject the null. Under the skewed heavy-tailed distribution, it is found that the rejection frequencies of both tests are on the higher side. The Ljung-Box test over-rejected the null hypothesis nearly more than twice the desired level. The Li-Mak test though has higher empirical sizes than the nominal size of 5% but can be considered a better alternative Ljung-Box test in case of asymmetric errors.

Next, we examine the power of both tests. The results of empirical powers of two diagnostic tests used are displayed in **(Table 2).** It can be observed from the results that both tests have very low empirical powers especially for small lag (M = 3). The empirical powers of tests increased with the lag length and sample size. The {M2, ARCH(2)} combination has the lowest power for Students-*t(3)* errors. The power of the tests drops when the true residuals are from heavy tailed and skewed heavy tailed distributions. Tse (2002) also noted this and argued that this may be due to the loss of efficiency in the QMLE which assumes normal errors. Tsui (2004) concluded that the Li-Mak test and Tse test (Tse, 2000) should be regarded as more appropriate tools for checking the model adequacy of univariate conditional heteroscedasticity specifications. The Li-Mak test is found to be more powerful than the Ljung-Box test for all errors distributions considered in this research.

Table 1: Empirical sizes of diagnostic tests for conditional heteroscedastic models

Error Distribution	DGP	EM	*T*	*M=3*		*M=6*		*M=10*	
				Q_{LB}	Q_{LM}	Q_{LB}	Q_{LM}	Q_{LB}	Q_{LM}
N(0, 1)	M1	GARCH(1, 1)	100	4.7	3.8	4.1	4.3	6.1	6.2
			500	4.6	4.2	4.7	4.4	5.4	6.1
			1000	5.1	4.1	4.7	4.4	5.3	5.5
	M2	ARCH(2)	100	6.2	6.0	6.0	6.6	6.1	6.5
			500	5.8	6.3	5.9	6.9	5.6	7.2
			1000	6.4	5.9	5.4	6.1	6.8	6.8
t(3)	M1	GARCH(1, 1)	100	6.1	5.8	6.9	6.0	7.2	6.5
			500	7.0	5.4	7.6	6.1	6.2	6.9
			1000	6.9	5.5	6.5	5.7	6.0	5.9
	M2	ARCH(2)	100	6.4	5.9	7.1	6.2	7.7	6.9
			500	5.8	5.5	6.5	5.7	6.8	6.8
			1000	5.4	5.5	6.2	5.1	6.7	6.4
ST(4, 0.25)	M1	GARCH(1, 1)	100	8.8	6.8	8.9	7.1	9.9	8.1
			500	7.5	7.1	7.2	6.8	8.1	7.4
			1000	6.9	6.5	6.8	6.9	7.0	6.8
	M2	ARCH(2)	100	9.2	6.9	10.5	9.2	11.9	10.8
			500	7.9	6.6	9.9	7.8	11.8	8.8
			1000	7.1	5.9	8.0	7.6	10.0	7.9

DGP: data generating process, EM: estimated model, Q_{LB} and Q_{LM} are Ljung-Box and Li-Mak test, respectively. *M* is the lag length. The nominal size of the test is 5%.

Table 2: Empirical powers of diagnostic tests for conditional heteroscedastic models

Error Distribution	DGP	EM	*T*	*M=3*		*M=6*		*M=10*	
				Q_{LB}	Q_{LM}	Q_{LB}	Q_{LM}	Q_{LB}	Q_{LM}
N(0, 1)	M1	ARCH(2)	100	6.9	7.7	6.5	8.9	7.0	10.6
			500	7.6	8.8	6.6	10.3	7.8	9.9
			1000	9.1	8.9	6.2	10.9	8.2	10.2
	M2	GARCH(1, 1)	100	6.6	2.5	6.5	8.0	7.2	15.6
			500	6.7	3.4	6.8	9.4	7.5	19.7
			1000	6.1	4.6	7.1	10.1	7.8	22.8
t(3)	M1	ARCH(2)	100	6.4	7.9	6.9	8.2	6.9	9.1
			500	6.8	8.8	7.0	9.3	7.8	10.5
			1000	7.7	9.4	8.1	11.4	9.1	10.8
	M2	GARCH(1, 1)	100	6.5	3.9	6.9	7.8	7.0	13.1
			500	6.5	4.3	7.2	8.1	7.1	14.1
			1000	6.0	4.5	6.9	8.9	7.2	18.2
ST(4, 0.25)	M1	ARCH(2)	100	4.1	6.6	7.1	9.7	7.1	10.0
			500	5.7	8.3	6.9	11.9	7.7	12.6
			1000	5.9	11.2	7.4	12.1	11.3	13.2
	M2	GARCH(1, 1)	100	5.2	6.1	5.6	8.8	11.9	11.1
			500	5.4	8.2	6.7	10.2	11.8	13.0
			1000	6.1	10.8	6.9	10.9	10.0	15.5

DGP: data generating process, EM: estimated model, Q_{LB} and Q_{LM} are Ljung-Box and Li-Mak test, respectively. *M* is the lag order. The nominal size of the test is 5%.

5. CONCLUSION

In this paper we studied the empirical size and power of two important diagnostic tests; the Ljung-Box test and the Li-Mak test for ARCH/GARCH models under asymmetric and heavy tailed errors using Monte Carlo simulations. We found that the both tests over-predicted the null hypothesis when errors were generated from skewed-*t* distribution. The Li-Mak test comparatively produced better empirical sizes than the Ljung-Box test. Though the empirical sizes of both tests were found very low for all errors distributions, again the Li-Mak test were found reliable than the Ljung-Box. Based on these findings we concluded that these tests may be used with caution and more powerful tests need to be developed for ARCH/GARCH models exhibiting asymmetry and heavy-tailedness.

REFERENCES:

Bollerslev, T. (1986) Generalised autoregressive conditional heteroscedasticity. J. Econometrics, 31 (**3**): 307–327.

Box, G. E. P., and Jenkins, G. M. (1970) *Time Series Analysis-Forecasting and Control.* Holden-Day, San Francisco.

Box, G. E. P., and Pierce, D. (1970) Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. J. Am. Stat. Assoc. **(65)**: 1509-1526.

Carbon, M., and Francq, C. (2011) Portmanteau goodness-of-fit test for asymmetric power GARCH models. Aus. J. Stat. **(40)**: 55 – 64.

Engle, R. F. (1982) Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation. Econometrica, **(50)**: 987–1008.

Glosten, L., Jagannathan, R., and Runkle, D. (1993) On the relation between the expected value and the volatility on the nominal excess return on stocks. J. Financ. 48 (**5**): 1779–1801.

Hansen, B. E. (1994) Autoregressive Conditional Density Estimation. International Econ. Rev. 35 (**3**): 705–730.

Jianhong, W., and Lixing, Z. (2009) Testing the adequacy of GARCH-type models in time series. Acta. Math. Sci. 29B (**2**): 327 – 340.

Li, W. K., and Mak, T. K. (1994) On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. J. Time. Ser. Anal. 15 (**6**): 627–636.

Ljung, G. M., and Box, G. E. P. (1978) On a measure of lack of fit in time series models. Biometrika, **(65)**: 297–303.

McLeod, A. I., and Li, W. K. (1983) Diagnostic checking ARMA time series models using squared residual autocorrelations. J. Time. Ser. Anal. 4 **(4):** 269–273.

Tse, Y. K., and Zuo, X. L. (1997) Testing for conditional heteroscedasticity: some monte carlo results. J. Stat. Comput. Sim. 58 (**3**): 237–253.

Tse, Y. K. (2000) A test for constant correlation in multivariate GARCH model. J. Econ. **(98)**: 107–127.

Tse, Y. K. (2002) Residual-based diagnostics for conditional heteroscedasticity models. Econ. J. **(5)**: 358–373.

Tsui, A. K. C. (2004) Diagnostic for conditional heteroscedastic models: some simulation results. Math. Comput. Sim. **(64)**: 113–119.