Diagnostics for Conditional Heteroscedasticity Models: Some Simulation Results

A. K. Tsui

Department of Economics, National University of Singapore, Kent Ridge, Singapore 119260, Singapore (ecsatsui@nus.edu.sg)

Abstract: In this paper, we study the size and power of various diagnostic statistics for univariate conditional heteroscedasticity models. These test statistics include the residual-based tests recently derived by Tse [2001], Li and Mak [1994] and Wooldridge [1991]. Monte Carlo experiments with 1000 replications are conducted to generate conditional variances which follow the ARCH/GARCH processes. We use quasi-maximum likelihood estimation method to obtain estimates of parameters under different ARCH/GARCH models. It is found that the Tse and Li-Mak diagnostics are more powerful.

Keywords: GARCH models; Residual-based diagnostics; Simulation

1. INTRODUCTION

The autoregressive conditional heteroscedasticity (ARCH) model (Engle [1982]) and the generalized ARCH (GARCH) model (Bollerslev [1987]) have been used extensively for analysing time series data. Franses and van Dijk [2000] provide an in-depth treatment of the subject and demonstrated the importance of capturing conditional variance structures in empirical economic and finance research.

While empirical researchers have been enjoying the success of fitting GARCH models in univariate time series, little has been focused on the checking of model adequacy. In fact, the Box-Pierce [1970] type of test statistic is regarded as the most popular diagnostic check. However, Li and Mak [1994] show that it does not follow an asymptotic χ^2 distribution. More recently, Tse [2001] derives a residual-based diagnostic test which accommodates the distortion caused by using OLS estimates. This paper is to compare the size and power of three diagnostic tests proposed by Wooldridge [1990, 1991], Li and Mak [1994] and Tse [2001], respectively. The rest of this paper is as follows. Section 2 gives a brief description of the diagnostic tests for model adequacy. Section 3 presents the Monte Carlo simulation results. Finally Section 4 gives some concluding remarks.

2. TEST STATISTICS

Consider a univariate time series of observations $y_1, \cdots, y_n$ with conditional heteroscedasticity generated by the following equations:

$$y_t - \mu_t = \epsilon_t, \qquad (1)$$

$$\epsilon_t = \sqrt{h_t}\eta_t \qquad (2)$$

where μ_t is the conditional mean and h_t is the conditional variance dependent on the information set Φ_{t-1} at time $t-1$. The conditional mean may be a function of past observations of y_t and values of a vector of weakly exogenous variables. In addition, η_t are independently and identically distributed with mean zero and variance 1. We consider the following GARCH(p,q) models

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{t-i} + \sum_{t=1}^{q} \beta_i \epsilon_{t-i} \quad (3)$$

Let θ be the vector of parameters capturing μ_t and h_t. One can obtain the maximum likelihood estimator (MLE) $\widehat{\theta}$ of θ based on the most commonly used assumption of errors following either normal or Student's t. For notational convenience, denote $\widehat{\epsilon}_t$ and $\widehat{h}_t$ as values of ϵ_t and h_t evaluated at $\widehat{\theta}$, and the estimated standardized residuals as $\widehat{\eta}_t = \widehat{\epsilon}_t/\sqrt{\widehat{h}_t}$. In what follows we briefly summarise the gist of three residual-based statistics.

2.1 Wooldridge test

Wooldridge [1990, 1991] derives a general regression-based diagnostics applicable to a wide class of possible misspecification. Within the framework of GARCH models, we adopt the Wool- dridge's diagnostic test by using the squared standardized residuals. Let's define $\widehat{\omega}_t = (\widehat{\epsilon}^2_{t-1}, \widehat{\epsilon}^2_{t-2}, \cdots, \widehat{\epsilon}^2_{t-m})'$ as the vector of indicator variables. Denote $\nabla_\theta \widehat{h}_t$ as the gradient vector of h_t with respect to θ evaluated at $\widehat{\theta}$, and $\nabla_\theta \widetilde{h}_t = \nabla_\theta \widehat{h}_t / \widehat{h}_t$. Basically, the Wooldridge test statistic is computed by three steps. First, obtain m-element residuals $\widehat{r}_t$ by regressing each elements of $\widehat{\omega}_t$ on $\nabla_\theta \widetilde{h}_t$. Second, regress unity on the vector of m regressors $\widehat{\Phi}_t \widehat{r}_t$, where $\widehat{\Phi}_t = \widehat{\epsilon}^2_t - 1$. Finally, compute $W(m) = N - SSR$, where SSR is the sum of squares of residuals of regression in the second step. Wooldridge shows that in the absence of model misspecification, $W(m)$ is asymptotically distributed as $\chi^2(m)$ with m degrees of freedom.

2.2 Li-Mak test

Li and Mak [1994] question the adequacy of using the popular Box-Pierce statistic as applied to the squared standardized residuals. They show that the Box-Pierce statistic does not converge to a χ^2 distribution asymptotically. Working on correlation coefficients of the squared standardized residuals, Li and Mak [1994] derive an appropriate asymptotic distribution and provide some diagnostic tests for GARCH models.

Let $\widehat{C}_k$ be defined as

$$\widehat{C}_k = \frac{1}{n} \sum_{t=k+1}^{n} (\widehat{\eta}^2_t - 1)(\widehat{\eta}^2_{t-k} - 1) \quad (4)$$

Then, the lag-k correlation coefficient $\widehat{r}_k$ of the squared standardized residual $\widehat{\eta}_t$ can be defined as

$$\widehat{r}_k = \frac{\sum_{t=k+1}^{n} (\widehat{\eta}^2_t - 1)(\widehat{\eta}^2_{t-k} - 1)}{\sum_{t=1}^{n} (\widehat{\eta}^2_t - 1)^2} \quad (5)$$

$\widehat{r}_k$ can then be rewritten as $\widehat{C}_k/\widehat{C}_0$. Let $\widehat{r} = (\widehat{r}_1, \widehat{r}_2, \cdots, \widehat{r}_m)'$ denote the vector of correlation coefficients containing elements from lags 1 to m. Li and Mak [1994] show that $\sqrt{n}\widehat{r}$ has an asymptotic normal distribution. The consistent variance of $\sqrt{n}\widehat{r}$ can be estimated by $\widehat{V} = I - (XG^{-1}X')/4$, where I is the $U \times U$ identity matrix, G^{-1} is a consistent estimate of the asymptotic variance of $\sqrt{n}(\widehat{\theta} - \theta)$ and $X = (X_1, X_2, \cdots, X_m)'$ with

$$X_k = -\frac{1}{n} \sum_{t=k+1}^{n} \frac{\nabla_\theta \widehat{h}_t}{\widehat{h}_t} (\widehat{\eta}^2_t - 1), \quad k = 1, 2, \cdots, m \quad (6)$$

Hence, if the model is specified correctly, $Q(m) = n\widehat{r}'\widehat{V}^{-1}\widehat{r}$ will be asymptotically distributed as a χ^2 with m degrees of freedom. For practical purposes, the factor 1/4 can be replaced by $1/\widehat{C}^2_0$ because $\widehat{C}_0$ converges to 2 when ϵ_t is Gaussian. We mention in passing that the Box-Pierce statistic is not considered as Li and Mak [1994] show that $\widehat{V}$ is generally not idempotent even asymptotically.

2.3 Tse test

Tse [2001] proposes a residual-based diagnostic for the adequacy of the conditional-variance structure by modifying the ordinary least squares (OLS) procedure which includes lagged squared standardized residuals in the regression. As warned by Wooldridge [1990], the inference procedure based on the conventional OLS estimates are questionable. Tse [2001] makes correction for the asymptotic variance of the OLS estimates and derives a residual-based statistic which follows an asymptotic χ^2 distribution. Sketch of the Tse test is summarized as follows. First, regress $\widehat{\eta}^2_t - 1$

on a vector of lagged squared standardized residuals $\widehat{d}_t = (\widehat{\eta}^2_{t-1}, \widehat{\eta}^2_{t-2}, \cdots, \widehat{\eta}^2_{t-m})$, consisting of the following structure.

$$\widehat{\eta}^2_t - 1 = \widehat{d}_t\delta + v_t \tag{7}$$

where δ is a $m \times 1$ vector of regression parameters. Second, applying the established results of Pierce [1982] which deals with substituting estimators for parameters, Tse obtains the following test statistic

$$T(m) = n\widehat{\delta}'\widehat{L}\widehat{G}^{-1}\widehat{L}\widehat{\delta} \tag{8}$$

where

$$\widehat{G} = \widehat{c}\widehat{L} - \widehat{S}\widehat{R}\widehat{S}' \tag{9}$$

$$\widehat{L} = \sum \widehat{d}_t\widehat{d}'_t/n \tag{10}$$

$$\widehat{S} = \sum \widehat{d}_t(\partial\widehat{\eta}^2_t/\partial\theta')/n \tag{11}$$

$$\widehat{c} = \sum(\widehat{\eta}^2_t - 1)^2/n \tag{12}$$

$\widehat{\delta}$ is the OLS estimator of δ, and $\widehat{R}$ is a consistent estimate of the asymptotic variance of $\sqrt{n}(\widehat{\theta} - \theta)$.

Tse [2001] shows that $T(m)$ is asymptotically χ^2 distributed with m degrees of freedom if the model is correct. In practice, derivatives in matrix $\widehat{S}$ can be computed by numerical differentiation, or alternatively by using recursive formulae given by Fiorentini *et al.* [1996], or by Tse [2000].

3. SIMULATION RESULTS

We conduct Monte Carlo experiments to examine the empirical size and power of the tests proposed by Wooldridge, Li and Mak, and Tse. We confine to low-order ARCH and GARCH processes for the conditional variance h_t. They include the following data generating processes (DGP) denoted by M1 for ARCH(1), M2 for ARCH(2) and M3 for GARCH(1,1), respectively.

$$M1: \; h_t = 0.2 + 0.6\epsilon^2_{t-1}, \tag{13}$$

$$M2: \; h_t = 0.2 + 0.6\epsilon^2_{t-1} + 0.2\epsilon^2_{t-2}, \tag{14}$$

$$M3: \; h_t = 0.2 + 0.7h_{t-1} + 0.1\epsilon^2_{t-1}, \tag{15}$$

Without loss of generality, the conditional mean of each DGP is assumed to be zero. For each DPG, we generate observations of sample size N, for $N = 200$, 500 and 1000, respectively and fit an ARCH/GARCH model to the data. The estimation models (EM) include ARCH(1), ARCH(2) and GARCH(1, 1). The relevant diagnostic statistics are computed from the sample estimates.

To gauge the performance of various tests for Gaussian and non-Gaussian errors, we consider processes that follow [i] a standard normal distribution, and [ii] a Student's t distribution with 8 degrees of freedom. In addition, all MLE are obtained under the assumption of normality, and all estimations are computed by using GAUSS programmes based on 1000 replications. As noted by Wooldridge (1991), the loss of efficiency in the quasi-MLE under Student's t may be insignificant.

The empirical size of various diagnostic tests is obtained by matching the DGP with the appropriate EM. We choose m = 1, 2, 3 and 4 because the magnitude of m is independent of the validity of the diagnostics. The empirical size is obtained when each of the appropriate model is estimated from the corresponding DGP. We use the 5 percent nominal size as benchmark for comparison. As can be seen from Table 1, W(m), Q(m) and T(m) tests generally have reasonably reliable size under the normal errors and t_8 errors. In fact, they yield good empirical size even for such a sample size of 200. However, all tests tend to over-reject the null hypothesis, except the Li-Mak and Tse tests which under-reject the null at m = 1 for the GARCH(1, 1) model.

Table 2 tabulates the empirical power of the three diagnostic tests. To avoid nesting between DGP and EM, we consider four combinations, including {M2, ARCH(1)}, {M2, GARCH(1, 1)}, {M3, ARCH(1)} and {M3, ARCH(2)}, respectively. Other combinations such as {M1, ARCH(1)} and {M1, GARCH(1, 1)} are excluded.

It can be observed that all three diagnostic tests have the lowest empirical power when $m = 1$, and no clear-cut options among $m =$

Table 1. Empirical Size of Diagnostics for Univariate Conditional Heteroscedasticity.

MODE	N	Test Statistics											
		W(1)	W(2)	W(3)	W(4)	Q(1)	Q(2)	Q(3)	Q(4)	T(1)	T(2)	T(3)	T(4)
Panel A: N(0, 1) errors													
MA1	200	5.7	5.7	5.2	6.3	6.2	5.6	4.9	4.6	6.2	6.3	5.8	6.0
	500	7.1	5.9	5.8	7.7	5.7	6.1	5.3	5.4	5.7	6.0	6.2	5.5
	1000	6.1	5.0	5.2	6.5	5.1	4.3	4.3	4.3	5.1	4.8	4.5	3.9
MA2	200	6.4	6.3	6.0	6.6	6.2	7.8	7.4	6.6	6.3	8.4	7.2	7.4
	500	5.5	5.5	5.6	6.5	7.5	8.6	9.1	10	7.6	8.4	9.4	11
	1000	7.7	5.8	7.1	8.0	5.4	7.4	7.9	7.8	5.4	7.4	7.5	7.3
MG	200	5.5	4.8	5.1	7.3	2.3	5.3	7.9	6.3	2.3	4.8	8.6	6.8
	500	5.9	6.4	5.8	6.5	2.8	4.8	5.4	6.0	2.8	4.9	5.8	6.5
	1000	6.3	5.1	6.4	6.0	2.5	4.0	5.7	6.0	2.5	3.9	5.8	6.1
Panel B: t_8 errors													
MA1	200	6.1	5.4	6.1	6.1	5.8	5.3	5.2	5.8	5.9	5.8	5.8	6.0
	500	4.8	5.0	5.3	5.8	4.2	4.4	5.1	5.2	4.3	4.4	5.2	4.7
	1000	4.9	5.7	5.7	6.1	4.7	4.3	4.7	5.5	4.7	4.4	5.1	5.5
MA2	200	6.4	5.9	6.0	6.9	3.9	4.9	4.7	6.2	4.0	5.3	4.7	6.2
	500	5.0	4.7	5.3	5.4	5.5	5.7	6.6	6.6	5.5	6.0	6.9	7.4
	1000	5.9	4.1	4.9	4.5	3.9	4.9	6.3	6.6	3.9	5.2	6.2	6.7
MG	200	5.9	5.5	6.9	7.6	1.4	3.2	5.4	6.4	1.3	3.1	6.1	7.2
	500	5.5	5.8	7.0	7.2	1.2	4.2	5.9	5.8	1.2	4.2	6.2	6.6
	1000	5.9	5.3	5.4	5.1	3.0	4.3	5.3	5.6	3.1	4.2	5.5	6.0

Notes: MA1 stands for (M1, ARCH(1)), MA2 for (M2, ARCH(2)) and MG for M3, GARCH(1,1)). W(m) is the Wooldridge test. Q(m) is the Li-Mak test. T(m) is the Tse test. We consider m = 1, 2, 3 and 4. The figures in the table are the empirical frequency of rejection in percentage. The nominal size of the tests is 5 percent. The estimation is based on 1000 replications.

2, 3, or 4. Power of all tests decreases when the true residuals are t_8 distributed. The {M3, ARCH(2)} combination give the lowest power among all three tests. One reason is due to the loss of efficiency in the quasi-MLE under the assumption of normal-distributed errors.

We mention in passing that the pattern of simulated empirical size and power of the three diagnostics under study are consistent with [i] various values of ARCH(1), ARCH(2) and GARCH(1, 1) models used in the data generating processes; [ii] number of replications (i.e. 1000 or 10000); and [iii] choice of degrees of freedom in the t-distribution.

4. CONCLUDING REMARKS

We have studied the empirical size and power of the diagnostic tests derived by Wooldridge, Li-Mak and Tse using Monte Carlo simulations. It is found that all three tests have reasonably reliable empirical size. However, the Li-Mak and Tse tests are more powerful than the Wooldridge test. As such, the Li-Mak and Tse tests should be regarded as more appropriate diagnostic tools for checking the model adequacy of univariate conditional heteroscedasticity specifications.

Table 2. Empirical Power of Diagnostics for Univariate Conditional Heteroscedasticity.

DGP	EM	N	Test	N(0, 1) errors m = 1	m = 2	m = 3	m = 4	t_8 errors m = 1	m = 2	m = 3	m = 4
M_2	ARCH(1)	200	W(m)	4.9	5.6	6.7	8.3	5.6	6.2	6.8	7.5
			Q(m)	11.1	37.4	39.6	38.9	6.0	25.0	28.6	27.3
			T(m)	11.2	35.6	41.5	38.9	6.0	24.0	28.0	27.2
M_2	ARCH(1)	500	W(m)	5.5	6.8	7.1	8.3	4.8	5.2	5.3	6.9
			Q(m)	21.5	73.5	77.2	77.4	9.8	53.5	57.3	57.2
			T(m)	21.8	71.9	77.6	76.9	9.8	51.6	58.2	56.9
M_2	ARCH(1)	1000	W(m)	4.2	5.6	6.6	7.1	4.6	6.8	7.0	7.7
			Q(m)	36.4	93.9	95.7	96.4	15.7	79.5	83.4	82.8
			T(m)	36.4	93.4	95.9	96.3	15.7	78.7	83.7	82.7
M_2	GARCH(1,1)	200	W(m)	5.9	5.6	6.9	7.2	5.3	5.9	6.6	6.6
			Q(m)	4.2	12.1	12.3	10.4	2.8	9.3	9.4	9.5
			T(m)	4.2	12.3	12.1	11.2	2.8	9.0	8.9	9.8
M_2	GARCH(1,1)	500	W(m)	5.0	4.7	5.4	6.2	4.0	4.8	5.1	5.6
			Q(m)	3.3	17.0	16.1	14.5	3.3	11.4	11.0	12.0
			T(m)	3.3	16.1	15.4	15.1	3.3	11.3	10.9	11.8
M_2	GARCH(1,1)	1000	W(m)	5.1	6.2	5.8	7.5	5.0	4.0	4.9	5.6
			Q(m)	3.8	20.6	19.6	18.6	3.3	14.5	13.4	13.9
			T(m)	3.9	20.4	19.0	18.2	3.3	14.0	12.9	13.4
M_3	ARCH(1)	200	W(m)	4.2	5.4	6.3	7.0	6.2	5.5	6.8	7.5
			Q(m)	7.2	12.6	14.2	13.9	5.6	8.4	10.2	11.2
			T(m)	7.3	12.3	14.9	12.8	5.7	8.3	11.0	11.8
M_3	ARCH(1)	500	W(m)	5.7	5.6	6.0	5.9	5.5	5.0	5.9	7.2
			Q(m)	9.0	23.4	28.4	26.1	4.9	18.5	22.0	22.1
			T(m)	9.1	22.0	28.1	26.6	5.0	17.8	22.4	22.3
M_3	ARCH(1)	1000	W(m)	6.5	7.5	6.6	6.6	5.7	4.2	4.8	5.1
			Q(m)	12.6	38.3	44.8	45.6	5.2	25.4	31.3	32.0
			T(m)	12.7	37.4	44.2	46.1	5.1	24.7	32.0	32.4
M_3	ARCH(2)	200	W(m)	6.5	4.8	4.7	6.4	6.3	6.2	6.1	5.8
			Q(m)	7.2	7.2	6.0	6.6	5.1	5.2	5.9	5.0
			T(m)	7.3	8.0	7.0	8.0	5.0	5.6	6.3	5.6
M_3	ARCH(2)	500	W(m)	5.3	4.3	5.7	7.0	5.2	4.8	5.3	6.4
			Q(m)	5.1	6.6	6.9	7.4	5.3	7.3	7.0	6.8
			T(m)	5.3	6.3	7.4	7.5	5.2	7.3	7.6	7.2
M_3	ARCH(2)	1000	W(m)	5.5	4.5	5.2	5.6	6.3	4.6	5.5	6.1
			Q(m)	6.5	7.1	8.0	7.0	4.9	5.6	7.7	6.8
			T(m)	6.4	7.2	7.9	7.5	4.8	5.7	7.2	6.9

Notes: DGP is the data generating process. EM is the estimated model. W(m) is the Wooldridge test. Q(m) is the Li-Mak test. T(m) is the Tse test. We consider m = 1, 2, 3 and 4. The figures in the table are the empirical frequency of rejection in percentage. The nominal size of the tests is 5 percent. The estimation is based on 1000 replications. In addition, those combinations for which the DGP is nested within the EM are excluded.

5. REFERENCES

Bollerslev, T., A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics*, 69, 542-549, 1987.

Box, G.E.P. and Pierce, D.A., Distribution of the residual autocorrelations in autoregressive integrated moving average time series models, *Journal of American Statistical Association*, 65, 1509-26, 1970.

Engle, R.F., Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom Inflation, *Econometrica*, 50, 987-1007, 1982.

Fiorentini, G, G. Calzolari and Panattoni, L., Analytic derivatives and the computation of GARCH estimates, *Journal of Applied Econometrics*, 11, 399-417, 1996.

Franses, P.H. and van Dijk, D., *Non-linear time series models in empirical finance*, Cambridge, 2000.

Li, W.K. and Mak, T.K., On the squared residual autocorrelations in non-linear time series with conditional heteroscedasticity, *Journal of Time Series Analysis*, 15, 627-636, 1994.

Pierce, D.A., The asymptotic effect of substituting estimators for parameters in certain types of statistics, *Annals of Statistics*, 10, 475-478, 1982.

Tse, Y.K., A test for constant correlations in a multivariate GARCH model, *Journal of Econometrics*, 98, 107-127, 2000.

Tse, Y.K., Residual-based diagnostics for conditional heteroscedasticity models, *Econometric Journal*, forthcoming.

Wooldridge, J.M., A unified approach to robust, regression-based specification tests, *Econometric Theory*, 6, 17-43, 1990.

Wooldridge, J.M., On the application of robust, regression-based diagnostics to models of conditional means and conditional variances, *Journal of Econometrics*, 47, 5-46, 1991.