DETECTING AND DIAGNOSTIC CHECKING MULTIVARIATE
CONDITIONAL HETEROSEDASTIC
TIME SERIES MODELS

H. Wong¹* and W. K. Li²**

¹Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
²Department of Statistics & Actuarial Science, The University of Hong Kong, Hong Kong, China

(Received March 5, 1999; revised October 20, 2000)

Abstract. Two tests for multivariate conditional heteroscedastic models are proposed. One is based on the cross-correlations of standardized squared residuals and the other is a score (Lagrange multiplier) test. The cross-correlations test can be used to detect the presence of multivariate conditional heteroscedasticity whereas the other test can be used for diagnostic checking. Simulation studies on the size and power of the test statistics are reported. The application of the tests is illustrated by an example using the S & P 500 and Sydney All Ordinary Indexes.

Key words and phrases: ARCH models, squared residuals, cross-correlation tests, score test.

1. Introduction

Box and Jenkins (1970) gave a comprehensive summary of linear time series. Then for the past 30 years nonlinear time series found tremendous development and applications. For example, the threshold model (Tong (1980, 1990)), the bilinear model (Granger and Andersen (1978)), the amplitude-dependent exponential autoregressive model (Haggan and Ozaki (1981)), the random coefficient autoregressive (RCA) model (Conlisk (1974); Nicholls and Quinn (1982)), and the autoregressive conditional heteroscedastic (ARCH) model (Engle (1982)). The last model seems to be particularly popular with econometricians and financial researchers and has found great success in those areas. Tsay (1987) showed the second order equivalence of the RCA and ARCH models by observing their common feature—changing conditional variance. Generalizations of the ARCH models include GARCH (Bollerslev (1986); Taylor (1986)), SETAR-ARCH (Tong (1990)), AARCH (Bera et al. (1992)) and many others. Despite of its various generalizations and vast applications, not many researches have been reported on the diagnostic checking of ARCH models. Some early work were Engle and Ng (1993), Chu (1995), and Lundbergh and Teräsvirta (1998).

On the other hand, squared residuals have been used extensively in the identification

*The first author's research was supported by an Internal Competitive Research Grant of the Hong Kong Polytechnic University (Grant Number G-YC15).
**The second author's research was supported by the Hong Kong Research Grants Council.
of ARCH models and the detection of nonlinearities in time series in the last ten to fifteen years. See for example, Tsay (1987), McLeod and Li (1983), Engle (1982), and Granger and Andersen (1978). The McLeod and Li test and Engle’s Lagrange multiplier test were found to be asymptotically equivalent (Luukkonen et al. (1988)). By introducing a generalized squared residual, Li and Mak (1994) developed formal diagnostic tests for univariate ARCH models and its generalizations. In this paper, the Li and Mak techniques will be generalized and used in specifying multivariate ARCH models. In fact, cross-correlation tests for multivariate conditional heteroscedasticity using standardised squared residuals from two univariate generalized ARCH models will be presented. The setup is similar to McLeod (1979). It is assumed that the two squared residuals series are \( M \)-dependent. However, we shall only derive results and statistics for the most important case in applications, i.e., we assume that the squared residual series have instantaneous dependence only. Its generalization to the \( M \)-dependent situation is not difficult. These tests are simple and handy for checking whether a bivariate relationship exists between the variance structures. In comparison, Cheung and Ng (1996) derived cross-correlation tests for squared residuals from two univariate ARMA models under the assumption that the two innovation series are independent. Wong and Li (1996) obtained the more general result when the two innovations are \( M \)-dependent. The present result is a further advancement.

More recently, there has been considerable interest in the application of multivariate ARCH models. See for example, Bollerslev et al. (1992) and Engle and Kroner (1995). Again little work has been reported on the diagnostic checking of these models. Lagrange multiplier type test for model adequacy can be derived from the information matrix of a multivariate model. A Lagrange multiplier test will be derived for checking a general class of multivariate ARCH models against specific alternatives. It is based on a form of the multivariate ARCH model developed in Wong and Li (1997). Organization of this paper is as follows: In Section 2, basic terminologies and the general framework for the diagnostic tests will be introduced. Section 3 contains the main derivations and results. To investigate the empirical performances of the tests, simulation studies and a real example will be reported in Section 4.

2. Basic definitions and assumptions

2.1 The bivariate case

Following Li and Mak (1994), a univariate general conditional heteroscedastic time series model is defined as follows:

Let \( X_t \) be a stationary and ergodic time series. Let \( \psi_t \) be the information set (\( \sigma \)-field) generated by all past observations up to and include time \( t \). Given \( \psi_{t-1} \), the distribution of \( X_t \) is assumed to be Gaussian with conditional mean \( \mu(\beta; \psi_{t-1}) \) and variance \( h(\beta; \psi_{t-1}) \), where \( \beta \) is a \( \ell \times 1 \) vector of parameters. For convenience, let \( \mu_t = \mu(\beta; \psi_{t-1}) \) and \( h_t = h(\beta; \psi_{t-1}) \). Here \( \mu_t \) and \( h_t \) are assumed to be known except for the parameter \( \beta \). Furthermore, it is assumed that both of them have continuous second order derivatives almost surely. Also we have \( \epsilon_t = X_t - \mu_t \). The above formulation will include the ARCH model as a special case, when \( \mu_t = 0 \), and \( h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_r \epsilon_{t-r}^2 \). Through suitable formulation of \( h_t \), many different conditional heteroscedastic models can be classified into several large families of models. The He and Teräsvirta (1999) formulation contained the GARCH(1, 1), absolute value GARCH, nonlinear GARCH(1, 1), volatility switching GARCH(1, 1), and generalized quadratic ARCH models. Hentschel (1995)
encompassed nesting symmetric and asymmetric GARCH models. Ding et al. (1993) discussed long memory models, while Higgins and Bera (1992) considered a family of nonlinear ARCH models.

Clearly the log-likelihood function would have to satisfy certain regularity conditions, and the usual conditions for maximum likelihood estimation, (Hall and Heyde, 1980, p. 156) are assumed to be met. Let \( \hat{\beta} \) be the conditional maximum likelihood estimator of \( \beta \). Let \( \hat{\varepsilon}_t \) be the corresponding residual when \( \beta \) is replaced by \( \hat{\beta} \). Similarly define \( \hat{\mu}_t \) and \( \hat{h}_t \). Now if there are two univariate time series \( X_{1t} \) and \( X_{2t} \) individually satisfying the above conditions, then corresponding definitions for \( \beta_{p_1}, \mu_{p_1}, h_{p_1}, \varepsilon_{p_1}, \hat{\beta}_{p_1}, \hat{\mu}_{p_1}, \hat{h}_{p_1}, \) and \( \hat{\varepsilon}_{p_1}(p = 1, 2) \) are obvious. Now the lag \( k \) squared residual cross-correlation is defined as:

\[
\hat{r}_{12}(k) = \frac{\sum_t (\hat{\varepsilon}_{1t}^2/\hat{h}_{1t} - \bar{\varepsilon}_1)(\hat{\varepsilon}_{2t+k}^2/\hat{h}_{2t+k} - \bar{\varepsilon}_2)}{\left\{\sum_t (\hat{\varepsilon}_{1t}^2/\hat{h}_{1t} - \bar{\varepsilon}_1)^2 \sum_t (\hat{\varepsilon}_{2t+k}^2/\hat{h}_{2t+k} - \bar{\varepsilon}_2)^2\right\}^{1/2}}, \quad k = 0, \pm 1, \pm 2, \ldots
\]

Here \( \bar{\varepsilon}_p = n^{-1} \sum_t \hat{\varepsilon}_{pt}^2/\hat{h}_{pt}, p = 1, 2 \) and \( n \) is the sample size. Since it may be shown that \( \hat{\varepsilon}_p(p = 1, 2) \) converge to one in probability if the model is correct, \( \hat{r}_{12}(k) \) can be replaced by

\[
\hat{r}_{12}(k) = \frac{\sum_t (\hat{\varepsilon}_{1t}^2/\hat{h}_{1t} - 1)(\hat{\varepsilon}_{2t+k}^2/\hat{h}_{2t+k} - 1)}{\left\{\sum_t (\hat{\varepsilon}_{1t}^2/\hat{h}_{1t} - 1)^2 \sum_t (\hat{\varepsilon}_{2t}^2/\hat{h}_{2t} - 1)^2\right\}^{1/2}}.
\]

Now let \( \hat{r}_k = (\hat{r}_{12}(-k), \hat{r}_{12}(-k + 1), \ldots, \hat{r}_{12}(0), \hat{r}_{12}(1), \ldots, \hat{r}_{12}(k - 1), \hat{r}_{12}(k)) \). Since \( n^{-1} \sum_p (\hat{\varepsilon}_{pt}^2/\hat{h}_{pt} - 1)^2(p = 1, 2) \) converge to two in probability we need only consider the asymptotic distribution of

\[
\hat{C}_{12}(k) = \frac{1}{n} \sum_t (\hat{\varepsilon}_{1t}^2/\hat{h}_{1t} - 1)(\hat{\varepsilon}_{2t+k}^2/\hat{h}_{2t+k} - 1).
\]

Result for \( \hat{r}_k \) follows immediately and will be derived in Section 3.

2.2 The general case

We consider the following random coefficient representation of a general multivariate conditional heteroscedastic model (Wong and Li (1997)):

\[
(2.1) \quad \Phi(B) y_t = \theta(B) a_t,
\]

\[
(2.2) \quad a_t = \sum_{i=1}^r \delta_{it}(B) a_t + \sum_{j=1}^s W_{jt}(B) y_t + e_t.
\]

Here \( \Phi(B) \) and \( \theta(B) \) are \( p \)-th and \( q \)-th order constant coefficient matrix polynomials in the backshift operator \( B \). Also, following Tsay (1987), we shall refer to (2.1) as the observation equation, and (2.2) the innovation equation. Tsay called the above representation a CHARMA model. Accordingly (2.1) and (2.2) will be named as the multivariate CHARMA model. Now \( y_t \) is an \( N \times 1 \) vector. (2.1) is the usual representation of a multivariate ARMA process. \( \delta_{it}(B), i = 1, \ldots, r \), denotes the \( r \)-th order random coefficient matrix in \( B \). For instance, \( \delta_{22t}(B) = \begin{pmatrix} \delta_{2211t} & \delta_{2221t} \\ \delta_{2212t} & \delta_{2222t} \end{pmatrix} B^2 a_t \), where \( \delta_{2211t}, \delta_{2212t}, \delta_{2221t}, \delta_{2222t} \), are zero mean random variables with a constant covariance matrix. The \( \delta_{it} \)'s are independent over \( t \) but may be correlated over \( i = 1, \ldots, r \). A similar definition holds for \( W_{jt}(B) \). The sequences \( \delta_{it} \) and \( W_{it} \) are independent of each other. It can be shown easily that if
\[ D_t = [\delta_{1t} | \delta_{2t} | \cdots | \delta_{rt} | W_{1t} | W_{2t} | \cdots | W_{st}] \]
\[ F_{t-1}^T = (a_{t-1}^T, a_{t-2}^T, \ldots, a_{t-r}^T, y_{t-1}^T, y_{t-2}^T, \ldots, y_{t-s}^T), \]
then
\[ \text{Var}(a_t | \psi_{t-1}) = d_t = (F_{t-1}^T \otimes I_N) \Delta (F_{t-1} \otimes I_N) + G \]

where \( \Delta = E(\text{vec} D_t \text{vec}^T D_t) \), \( G = E(\varepsilon_t e_t^T) \), and \( \psi_t \) is again the information set generated by all past observations up to and include time \( t \). Also the superscript \( T \) denotes matrix transpose. Let the vector of parameters to be estimated in (2.1) be \( \beta_1 \), and the vector of parameters to be estimated in (2.3), i.e. parameters in \( \Delta \) and \( G \), be \( \beta_2 \). We also let \( \beta^T = (\beta_1^T, \beta_2^T) \). Recall as in Subsection 2.1, we have \( \mu_t = \mu(\beta; \psi_{t-1}) \). Let the log-likelihood be \( \ell \), then by the usual conditional log-likelihood decomposition, we have: \( \ell = \sum_{t} \ell_t \), where \( \ell_t = -\frac{1}{2} \log |d_t| - \frac{1}{2} (y_t - \mu_t)^T d_t^{-1} (y_t - \mu_t) \). The crucial step in constructing a Score (Lagrange multiplier) test is to find \( -E(\partial^2 \ell / \partial \beta \partial \beta^T) \), or \( -E(\sum_{t} \partial^2 \ell_t / \partial \beta \partial \beta^T) \). A closed form of the last expression will be given in Subsection 3.2. The test procedure can be extended directly to ARMA-GARCH type models by using numerical estimate of the Information matrix. Some existing algorithms are the BHHH and Fiorentini (Fiorentini et al. (1996)). Alternatively the GARCH models can be put in a CHARMA framework using the approach of Bera et al. (1996).

3. Main results and testing statistics

3.1 The bivariate case and squared residual cross-correlation tests for lagged relationship

Following the development of Subsection 2.1 and McLeod (1979), let there be two stationary and ergodic time series \( X_{1t} \) and \( X_{2t} \). Assume that \( X_{pt} \) is conditionally Gaussian with mean and variance \( \mu_{pt} \) and \( h_{pt} \), \( p = 1, 2 \). Furthermore, \( \varepsilon_{pt} = X_{pt} - \mu_{pt} \), \( \mu_{pt} = \mu(\beta_p, \psi_{t-1}) \), \( h_{pt} = h(\beta_p, \psi_{t-1}) \), \( p = 1, 2 \). Here again \( \psi_t \) is the information set up to time \( t \). In the following derivation we assume that \( \varepsilon_{1t}^2 \) and \( \varepsilon_{2t}^2 \) have no temporal dependence. Namely, \( E(\varepsilon_{1t}^2 \varepsilon_{2t+\ell}^2) - E(\varepsilon_{1t}^2)E(\varepsilon_{2t+\ell}^2) \) can be non-zero only when \( \ell = 0 \). When \( \ell \) is any other integer, \( E(\varepsilon_{1t}^2 \varepsilon_{2t+\ell}^2) - E(\varepsilon_{1t}^2)E(\varepsilon_{2t+\ell}^2) \) must be equal to zero. Now consider the asymptotic distribution and information matrix of \( \beta_p, p = 1, 2 \). It suffices to show the case \( p = 1 \) only. For each \( t \) the contribution to the log likelihood \( \ell_1 \) by \( X_{1t} \) is, apart from a constant, \( \ell_{1t} = -\frac{1}{2} \log h_{1t} - \frac{1}{2} \varepsilon_{1t}^2 / h_{1t} \), and \( \ell_1 = \sum \ell_{1t} \). Thus

\[ \frac{\partial \ell_1}{\partial \beta_1} = \frac{1}{2} \sum \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{1t}^2}{h_{1t}} - 1 \right) + \sum \frac{\varepsilon_{1t}}{h_{1t}} \frac{\partial \mu_{1t}}{\partial \beta_1}. \]

Differentiating again and taking iterative expectations with respect to \( \psi_{t-1} \) (Higgins and Bera (1992)), we have

\[ E \left( \frac{\partial^2 \ell_1}{\partial \beta_1 \partial \beta_1^T} \right) = -\frac{1}{2} \sum E \left[ \frac{1}{h_{1t}^2} \left( \frac{\partial h_{1t}}{\partial \beta_1} \right) \left( \frac{\partial h_{1t}}{\partial \beta_1} \right)^T \right] - \sum E \left[ \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\partial \mu_{1t}}{\partial \beta_1} \right)^T \right]. \]

Under the usual regularity conditions (Hall and Heyde (1980), p. 156) for maximum likelihood estimators, \( \sqrt{n}(\hat{\beta}_1 - \beta_1) \) can be shown to be asymptotically normally distributed.
with mean zero and variance $G_1^{-1} = -E(n^{-1} \partial^2 \ell_1 / \partial \beta_1 \partial \beta_1^T)^{-1}$. Thus asymptotically we have

$$(\hat{\beta}_p - \beta_p) \overset{d}{=} (nG_p)^{-1} \partial \ell_p / \partial \beta_p, \quad G_p \overset{d}{=} -E(n^{-1} \partial^2 \ell_p / \partial \beta^2_p), \quad \text{for } p = 1, 2.$$ 

Now let $C_M = (C_{12}(-M), C_{12}(-M+1), \ldots, C_{12}(1), C_{12}(0))^T$, and $\hat{C}_M = (\hat{C}_{12}(-M), \hat{C}_{12}(-M+1), \ldots, \hat{C}_{12}(1), \hat{C}_{12}(0))^T$, for some $M > 0$. Here $C_{12}(k) = \frac{1}{n} \sum (\varepsilon_{1t}^2/h_{2t-k} - 1)/(\varepsilon_{2t-k}^2/h_{2t-k})$ for $k = -M, \ldots, 0, \ldots, M$. Similarly define $r_M$ and $\hat{r}_M$. Now let $\beta^T = (\beta_1^T, \beta_2^T)$ and $\hat{\beta}^T = (\hat{\beta}_1^T, \hat{\beta}_2^T)$. By a Taylor series expansion of $\hat{C}_M$ about $\beta$ and evaluated at $\beta$ we have

$$\hat{C}_M \overset{d}{=} C_M + \frac{\partial C_M}{\partial \beta^T}(\hat{\beta} - \beta)$$

where $\partial C_M / \partial \beta^T$ is a matrix with rows $(\partial C_{12}(k)/\partial \beta)^T$ for $k = -M, -M+1, \ldots, -1, 0, 1, \ldots, M-1, M$. For instance, if $\beta_1$ and $\beta_2$ consist of $q_1$ and $q_2$ parameters respectively, then $\partial C_M / \partial \beta^T$ is a $2M+1$ by $q_1 + q_2$ matrix. Now we consider the structure of this matrix in detail.

Let $C_{M-} = (C_{12}(-M), C_{12}(-M+1), \ldots, C_{12}(-1))^T$, $\hat{C}_{M-} = (\hat{C}_{12}(-M), \hat{C}_{12}(-M+1), \ldots, \hat{C}_{12}(-1))^T$, $C_{M+} = (C_{12}(1), C_{12}(2), \ldots, C_{12}(M))^T$, and $\hat{C}_{M+} = (\hat{C}_{12}(1), \hat{C}_{12}(2), \ldots, \hat{C}_{12}(M))^T$. Thus $C_T = (C_{M-}, C_{12}(0), C_{M+})$ and

$$\frac{\partial C_M}{\partial \beta^T} = \begin{pmatrix}
\frac{\partial C_{M-}}{\partial \beta_1^T} & \frac{\partial C_{M-}}{\partial \beta_2^T} \\
\frac{\partial C_{12}(0)}{\partial \beta_1^T} & \frac{\partial C_{12}(0)}{\partial \beta_2^T} \\
\frac{\partial C_{M+}}{\partial \beta_1^T} & \frac{\partial C_{M+}}{\partial \beta_2^T}
\end{pmatrix}_{(2M+1) \times (q_1 + q_2)}.$$ 

Now by taking iterative expectations with respect to $\psi_{t-1}$ and using similar techniques as in Mcleod and Li (1983), it can be shown that if $X_{1t}$ and $X_{2t}$ have instantaneous dependence in variance only, then $\sqrt{n}C_{M+}$ and $\sqrt{n}C_{M-}$ are both asymptotically normally distributed with mean zero and variance $4I$, where $I$ is the $M$ by $M$ identity matrix. Next we consider entries of $\partial C_{M-} / \partial \beta_1^T$.

For $\partial C_{12}(k)/\partial \beta_1$, $k = 1, 2, \ldots, M$, observe

$$\frac{\partial C_{12}(k)}{\partial \beta_1} = n^{-1} \frac{\partial}{\partial \beta_1} \sum \left( \frac{\varepsilon_{1t}^2}{h_{1t}} - 1 \right) \left( \frac{\varepsilon_{2t-k}^2}{h_{2t-k}} - 1 \right)$$

$$= n^{-1} \left\{ \sum \left( \frac{2\varepsilon_{1t}}{h_{1t}} \left( - \frac{\partial \mu_{1t}}{\partial \beta_1} \right) - \frac{\varepsilon_{1t}^2}{h_{1t}} \right) \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{2t-k}^2}{h_{2t-k}} - 1 \right) \right\}. \tag{3.2}$$

Therefore

$$E \left( \frac{\partial}{\partial \beta_1} C_{12}(k) \mid \psi_{t-1} \right) = \frac{1}{n} \left\{ \sum - \frac{1}{h_{1t}} \cdot \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{2t-k}^2}{h_{2t-k}} - 1 \right) \right\} = X_{1k}.$$ 

Note that the conditional expectation of the expression

$$n^{-1} \sum (2\varepsilon_{1t}/h_{1t})(-\partial \mu_{1t}/\partial \beta_1)(\varepsilon_{2t-k}^2/h_{2t-k} - 1)$$

from (3.2) will vanish only in the no temporal dependence case. In the $M$-dependent case we have to keep this term. By the ergodic theorem $E(\partial/\partial \beta_1 C_{12}(-k))$ can be consistently
estimated by \( X_{1k} \). Let the resultant \( M \times q_1 \) matrix be \( X_1 \) when \( \partial/\partial \beta_1 C_{12}(-k), k = 1, \ldots, M \), are estimated by \( X_{1k} \), i.e. \( X_1 = (X_{11}, X_{12}, \ldots, X_{1M})^T \). Next we consider entries of \( \partial C_{M-}/\partial \beta_2^T \). Now

\[
\frac{\partial C_{12}(-k)}{\partial \beta_2} = \frac{1}{n} \sum \left( \frac{\epsilon_{1t}^2}{h_{1t}} - 1 \right) \left[ \frac{2\epsilon_{2t-k}}{h_{2t-k}} \left( -\frac{\partial \mu_{2t-k}}{\partial \beta_2} \right) - \frac{\epsilon_{2t-k}^2}{h_{2t-k}^2} \frac{\partial h_{2t-k}}{\partial \beta_2} \right].
\]

Therefore, \( E(\partial C_{12}(-k)/\partial \beta_2 | \psi_{t-1}) = 0 \).

Note again the above expression vanishes only in the case of no temporal dependence. Similarly it can be shown that

\[
\frac{\partial C_{12}(0)}{\partial \beta_1} = \frac{1}{n} \sum \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} - \frac{\epsilon_{2t}^2}{h_{2t}} \frac{\epsilon_{1t}}{h_{2t}} \frac{\partial \mu_{1t}}{\partial \beta_1} - \frac{\epsilon_{2t}^2}{h_{2t}} \frac{\epsilon_{1t}^2}{h_{2t}^2} \frac{\partial h_{1t}}{\partial \beta_1},
\]

and

\[
\frac{\partial C_{12}(0)}{\partial \beta_2} = \frac{1}{n} \sum \frac{1}{h_{2t}} \frac{\partial h_{2t}}{\partial \beta_2} - \frac{\epsilon_{1t}^2}{h_{1t}} \frac{\epsilon_{2t}}{h_{2t}} \frac{\partial \mu_{2t}}{\partial \beta_2} - \frac{\epsilon_{1t}^2}{h_{1t}} \frac{\epsilon_{2t}^2}{h_{2t}^2} \frac{\partial h_{2t}}{\partial \beta_2}.
\]

Note that the expectation of the second term in (3.4) and (3.5) vanishes under the normality assumption. The remaining terms in equations (3.4) and (3.5) provide consistent estimates for \( E(\partial C_{12}(0)/\partial \beta_1) \) and \( E(\partial C_{12}(0)/\partial \beta_2) \). Let the estimates be denoted as \( \hat{C}_{12}(0)/\partial \beta_1 \) and \( \hat{C}_{12}(0)/\partial \beta_2 \) respectively. Finally consider entries of \( \partial C_{M+}/\partial \beta_2^T \) and \( \partial C_{M-}/\partial \beta_2^T \). Now it is well known \( C_{12}(+k) = C_{21}(-k) \) for \( k = 1, \ldots, M \). Thus

\[
\frac{\partial C_{12}(+k)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum \left( \frac{\epsilon_{2t}^2}{h_{2t}} - 1 \right) \left( \frac{\epsilon_{1t-k}^2}{h_{1t-k}} - 1 \right) = \sum \left( \frac{\epsilon_{2t}^2}{h_{2t}} - 1 \right) \left( \frac{2\epsilon_{1t-k}}{h_{1t-k}} \left( -\frac{\partial \mu_{1t-k}}{\partial \beta_1} \right) - \frac{\epsilon_{1t-k}^2}{h_{1t-k}^2} \frac{\partial h_{1t-k}}{\partial \beta_1} \right).
\]

Therefore, \( E(\partial C_{12}(+k)/\partial \beta_1 | \psi_{t-1}) = 0 \). And

\[
\frac{\partial C_{12}(+k)}{\partial \beta_2} = \frac{1}{n} \frac{\partial}{\partial \beta_2} \sum \left( \frac{\epsilon_{2t}^2}{h_{2t}} - 1 \right) \left( \frac{\epsilon_{1t-k}^2}{h_{1t-k}} - 1 \right) = \frac{1}{n} \sum \left( \frac{\epsilon_{2t-k}^2}{h_{2t-k}} - 1 \right) \left[ 2\frac{\epsilon_{2t}}{h_{2t}} \left( -\frac{\partial \mu_{2t}}{\partial \beta_2} \right) - \frac{\epsilon_{2t}^2}{h_{2t}^2} \frac{\partial h_{2t}}{\partial \beta_2} \right].
\]

Therefore,

\[
E \left( \frac{\partial C_{12}(+k)}{\partial \beta_2} | \psi_{t-1} \right) = \frac{1}{n} \left[ \sum \left( \frac{\epsilon_{2t-k}^2}{h_{2t-k}} - 1 \right) \left( -\frac{1}{h_{2t}} \frac{\partial h_{2t}}{\partial \beta_2} \right) \right].
\]

Hence by the ergodic theorem again, \( E(\partial C_{12}(+k)/\partial \beta_2) \) can be consistently estimated by

\[
X_{2k} = -\frac{1}{n} \sum \left( \frac{\epsilon_{2t-k}^2}{h_{2t-k}} - 1 \right) \left( \frac{1}{h_{2t}} \frac{\partial h_{2t}}{\partial \beta_2} \right).
\]

Let \( X_2 \) denote the matrix \( (X_{21}, X_{22}, \ldots, X_{2M})^T \). Hence if we let

\[
X = \left( \begin{array}{c} X_1 \\ \partial \hat{C}_{12}(0)/\partial \beta_1^T \\ \partial \hat{C}_{12}(0)/\partial \beta_2^T \\ X_2 \end{array} \right)_{(2M+1) \times (q_1+q_2)}
\]
then \( \hat{C}_M \equiv C_M + X(\hat{\beta} - \beta) \).

\( \hat{C}_M \) and hence \( \hat{r}_M \) can be shown to be asymptotically jointly normal by using the Mann-Wald device and the martingale central limit theorem (Billingsley (1961)). Now we arrive at the following theorem:

**Theorem 1.** Assume \( \varepsilon^2_{1t} \) and \( \varepsilon^2_{2t} \) have no temporal dependence. The asymptotic distributions of \( \sqrt{n}\hat{r}_M \) and \( \sqrt{n}\hat{r}_M+ \) are both normal with means \( \sqrt{n}\hat{r}_M \) and \( \sqrt{n}\hat{r}_M+ \) respectively. And the covariance matrices are \( \mathbf{I} - \frac{1}{4}X_1G_1^{-1}X_1^T \) and \( \mathbf{I} - \frac{1}{4}X_2G_2^{-1}X_2^T \) respectively. Here \( G_p = -E(\partial^2\ell_p/\partial\beta_p\partial\beta_p^T) \) for \( p = 1, 2 \).

**Proof.** It suffices to show the \( \hat{r}_M^- \) case. Similar results for \( \hat{r}_M^+ \) follow by direct computation. Normality of \( \hat{r}_M^- \) follows from that of \( \hat{r}_M^- \). For the correlation matrix of \( \hat{r}_M^- \), consider first the asymptotic cross-covariance of \( \sqrt{n}(\hat{\beta}_1 - \beta_1) \) and \( \sqrt{n}C_{M-} \). As 
\[
\hat{\beta}_1 - \beta_1 \approx (nG_1)^{-1}\partial\ell_1/\partial\beta_1,
\]
then
\[
E(\hat{\beta}_1 - \beta_1) = \hat{C}_M = C_M + X(\hat{\beta} - \beta).
\]

Recall \( C_{M-} = (C_{12}(-M), C_{12}(-M+1), \ldots, C_{12}(-k), \ldots, C_{12}(-1))^T \). From (3.1) \( E(\partial\ell_1/\partial\beta_1 \cdot C_{M-}) \) is equal to
\[
G_{-1}^{-1}E\left\{ \sum_{t} \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{1t}^2}{h_{1t}} - 1 \right) + \sum_{t} \frac{1}{h_{1t}} \frac{\partial \mu_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{1t}^2}{h_{1t}} - 1 \right) \right\}.
\]

By taking iterative expectation it can be shown that the crosscovariance of \( \varepsilon_{1t}/h_{1t} \cdot \varepsilon_{1t}/h_{1t} \) is zero. Similarly it can be shown that
\[
E\left\{ \sum_{t} \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{1t}^2}{h_{1t}} - 1 \right) \cdot \sum_{t'} \frac{1}{h_{1t'}} \frac{\partial h_{1t'}}{\partial \beta_1} \left( \frac{\varepsilon_{1t'}^2}{h_{1t'}} - 1 \right) \cdot \frac{\varepsilon_{2t'-k}^2}{h_{2t'-k}} - 1 \right\}
\]
is non-zero only when \( t = t' \). In which case, observe
\[
E\left\{ \sum_{t} \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{1t}^2}{h_{1t}} - 1 \right) \cdot \sum_{t'} \frac{1}{h_{1t'}} \frac{\partial h_{1t'}}{\partial \beta_1} \left( \frac{\varepsilon_{1t'}^2}{h_{1t'}} - 1 \right) \cdot \frac{\varepsilon_{2t'-k}^2}{h_{2t'-k}} - 1 \right\}
\]
\[
= E\left\{ \frac{1}{2n} \sum_{t} \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{1t}^2}{h_{1t}} - 1 \right) \cdot \frac{\varepsilon_{2t-k}^2}{h_{2t-k}} - 1 \right\}
\]
\[
\approx \frac{1}{n} \sum_{t} \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{1t-k}^2}{h_{1t-k}} - 1 \right), \quad \text{as} \quad E\left( \frac{\varepsilon_{1t}^2}{h_{1t}} - 1 \right)^2 = 2.
\]

Therefore, \( E(\partial\ell_1/\partial\beta_1 \cdot C_{12}(-k)) \) is equal to
\[
\frac{1}{n} \sum_{t} E\left( \frac{1}{h_{1t}} \frac{\partial h_{1t}}{\partial \beta_1} \left( \frac{\varepsilon_{2t-k}^2}{h_{2t-k}} - 1 \right) \right).
\]

Comparing with the expectation of (3.2), we see that asymptotically \( E(\partial\ell_1/\partial\beta_1 \cdot C_{M-}) = -G_1^{-1}X_1^T \). Now as \( \dot{r}_2 \equiv \dot{C}_{12}(-k)/\sqrt{\dot{C}_{11}(0)\dot{C}_{22}(0)} \equiv \dot{C}_{12}(-k)/2 \), it is clear that asymptotically \( \text{Cov}(\sqrt{n}\hat{r}_{M-}) = \mathbf{I} - \frac{1}{4}X_1G_1^{-1}X_1^T \). The theorem now follows.
Now if we let \( V_1 = I - \frac{1}{4} X_1 G_1^{-1} X_1^T \) and \( V_2 = I - \frac{1}{4} X_2 G_2^{-1} X_2^T \), then the portmanteau statistics \( \hat{Q}_{M-} = n r_{M-}^T \hat{V}_1^{-1} \hat{r}_{M-} \) and \( \hat{Q}_{M+} = n r_{M+}^T \hat{V}_2^{-1} \hat{r}_{M+} \) are clearly tests for temporal dependence in variance. Note here for \( \hat{V}_1 \), the estimate of \( V_1 \), we can obtain it by replacing entries of \( G_1 \) by its sample averages. \( \hat{V}_2 \) can naturally be obtained by the same approach. Under the null of instantaneous dependence in variance only, then both \( \hat{Q}_{M-} \) and \( \hat{Q}_{M+} \) will follow the \( \chi^2 \) distribution with \( M \) degrees of freedom. Rejecting any of the two statistics will imply that there is temporal dependence in variance of the two series under investigation. Equivalently, this is an indication of the presence of multivariate conditional heteroscedasticity and thus multivariate ARCH models should be considered.

Another important point is that if \( \varepsilon_{1t}^2 \) and \( \varepsilon_{2t}^2 \) are uncorrelated, then the conditional expectation of \( \partial C_{12}(k)/\partial \beta_2 \) in (3.6) would be zero. Then both the \( X_1 \) and \( X_2 \) matrices are zero, whence \( V_1 \) and \( V_2 \) reduce to the identity matrix. It follows then the two portmanteau statistics in the last paragraph reduce to \( nr_{M-}^T r_{M-} \) and \( nr_{M+}^T r_{M+} \), when \( \varepsilon_{1t}^2 \) and \( \varepsilon_{2t}^2 \) are uncorrelated. These statistics are now in the form of the Box-Pierce statistic (Box and Pierce (1970)).

Let \( Q_{B-} = nr_{M-}^T r_{M-} \) and \( Q_{B+} = nr_{M+}^T r_{M+} \).

Now since \( V_1^{-1} \) and \( G_1^{-1} \) are both positive semi-definite, it follows immediately that \( Q_{B-} \leq Q_{M-} \). Equality holds when \( \varepsilon_{1t}^2 \) and \( \varepsilon_{2t}^2 \) are uncorrelated. If temporal dependence in variance exists, then in general \( Q_{B-} < Q_{M-} \), and \( Q_{B-} \) will be a conservative statistic asymptotically. This phenomenon will be demonstrated in the simulations in Section 4. Similar argument applies to the \( Q_{B+} \) and \( Q_{M+} \) statistics.

3.2 A score test for the multivariate CHARMA model

3.2.1 The information matrix

The score (Lagrange multiplier) test is a well known technique for testing model adequacy. For a recent survey on the historical developments and new results, see Bera and Bilias (2001). The usual difficulty in constructing a score test is to find the information matrix. However, in this case if we use the estimation technique developed in Mak (1993), then the estimation and the information matrix problem can be solved simultaneously. Details of the derivation of the information matrix and the estimation procedure of the multivariate CHARMA model can be found in Wong and Li (1997). The results are described below. Recall our model equation as defined in (2.1), (2.2) and (2.3). Let the vector of parameters in the observation equation (2.1) be \( \beta_1 \). The dimension of \( \beta_1 = K \). Similarly let the vector of parameters in the innovation equation (2.3) be \( \beta_2 \). The dimension of \( \beta_2 = L \). Following Wong and Li (1997), we arrive at the following theorem:

**Theorem 2.** Let \( G(\beta) \) be the information matrix of the model defined by (2.1), (2.2) and (2.3), then

\[
G(\beta) = \begin{pmatrix} G_{11} & G_{12} \\ G_{12}^T & G_{22} \end{pmatrix}_{(K+L) \times (K+L)}
\]

where

\[
G_{11} = -\frac{1}{2} \sum_t \frac{\partial \text{vec}^T d_t^{-1}}{\partial \beta_1} \left( \frac{\partial}{\partial \beta_1^T} \text{vec} d_t \right) + \frac{1}{2} \sum_t (\text{vec}^T d_t^{-1} \otimes I_K)
\cdot \left\{ I_N \otimes \left[ \frac{\partial \mu_t}{\partial \beta_1} \left( \frac{\partial \mu_t^T}{\partial \beta_1^T} \right) + I_{(K,N)} \left( I_K \otimes \frac{\partial \mu_t}{\partial \beta_1} \right) \left( \frac{\partial \mu_t^T}{\partial \beta_1} \otimes I_K \right) \right] \right\} (\text{vec} I_N \otimes I_K),
\]
here $I_{(K,N)}$ is the commutation matrix as defined in Magnus ((1988), p. 35), $I_{(K,N)}$ is also known as the permuted identity matrix (Rogers (1980)).

$$G_{12} = -\frac{1}{2} \sum \frac{\partial \text{vec}^T d_{t-1}}{\partial \beta_1} \cdot \frac{\partial \text{vec} d_t}{\partial \beta_2^T}$$

and

$$G_{22} = \frac{1}{2} \sum \frac{\partial d_t}{\partial \beta_2} \ast (I_N \otimes \text{vec} I_L) \left( d_t^{-1} \frac{\partial d_t}{\partial \beta_2^T} (d_t^{-1} \otimes I_L) \right).$$

3.2.2 The score (Lagrange multiplier) test

Now our problem is to determine whether the orders in (2.2), i.e. $r$ and $s$, are adequate or not. For ease of exposition, suppose we want to test the null hypothesis $H_0 : r = m$ against the alternative $r = m + p, p \geq 1$. Following Harvey ((1990), §5.5), the above problem can be formulated as follows:

Let the vector of parameters of the full model be $\xi$. If $\xi$ is partitioned as $\xi^T = (\xi_{(1)}, \xi_{(2)})$, where $\xi_{(1)}$ is the vector of parameters corresponds to the model under $H_0$, then our testing problem is equivalent to a test of $H'_0 : \xi_{(2)} = 0$.

Partitioning the information matrix conformably with $\xi_{(1)}$ and $\xi_{(2)}$ yields

$$I(\xi) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}.$$ 

Following the standard technique, $\xi$ will be estimated under the restricted model using maximum likelihood. $I(\xi)$ can be found using the result of Subsection 3.2.1. The LM test is valid under the conditions for asymptotic normality of the maximum likelihood estimator. Let the maximum likelihood estimate under $H'_0$ be $\hat{\xi}_0$. Let $D_2(\hat{\xi}_0)$ be the vector of partial derivatives of $\ell(\xi)$ (the log-likelihood) with respect to $\xi_{(2)}$ and evaluated at $\xi_0$. The Lagrange multiplier statistic is: $LM = \{D_2(\xi_0)\}' [I_{22} - I_{21}I_{11}^{-1}I_{12}]^{-1} \{D_2(\xi_0)\}$, where $I_{ij}, i, j = 1, 2$ are evaluated at $\xi_0$. If $\xi_{(2)}$ is of dimension $M$, then under the null $LM$ will follow a $\chi^2_M$ distribution. Under the null of a constant and diagonal $d_t$, the score test for no temporal causality of volatility can be shown to be asymptotically equivalent to the sum of the squared $Q_{M-}$ and $Q_{M+}$ statistics. Lundbergh and Teräsvirta (1998) obtained a similar result for the univariate portmanteau test in Li and Mak (1994). It should be noted, however, that the cross-correlation test and the score test have different alternative hypotheses.

4. Monte carlo results and empirical example

4.1 Monte Carlo results

Example 1. Cross-correlation tests for data generated from univariate series. We consider series from two ARCH(1) models. The two statistics $Q_{M-}$ and $Q_{M+}$ are portmanteau tests for dependence in variance. Because of symmetry, we only considered the $Q_{M-}$ in the simulations. It was compared with the Box-Pierce statistic, $Q_{B-} = n \sum_{j=-M}^{-1} \bar{r}_j^2$. We looked at the empirical sizes and power of the two statistics for 4 different cases and the sample sizes ($n$) considered are 200, 400, and 600. In the power study, the perturbation coefficient is just a moderate value of 0.3. In summary, the four cases considered were:

$$X_{1t} = \sqrt{\alpha_0 + \alpha_1 X_{1t-1}^2 + p X_{2t-1}^2} \cdot (\rho \varepsilon_{2t} + \varepsilon_{1t}), \quad X_{2t} = \sqrt{\alpha_0 + \alpha_1 X_{2t-1}^2} \cdot \varepsilon_{2t}.$$
Table 1a. Empirical size of the $Q_{M-}$ and $Q_{B-}$ statistics.

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$\rho$</td>
<td>$Q_{M-}$</td>
<td>$Q_{B-}$</td>
<td>$Q_{M-}$</td>
<td>$Q_{B-}$</td>
<td>$Q_{M-}$</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>0.043</td>
<td>0.041</td>
<td>0.059</td>
<td>0.056</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.063</td>
<td>0.066</td>
<td>0.060</td>
<td>0.056</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.062</td>
<td>0.052</td>
<td>0.054</td>
<td>0.048</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.051</td>
<td>0.041</td>
<td>0.059</td>
<td>0.050</td>
<td>0.049</td>
</tr>
<tr>
<td>400</td>
<td>0</td>
<td>0.066</td>
<td>0.065</td>
<td>0.046</td>
<td>0.044</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.044</td>
<td>0.042</td>
<td>0.065</td>
<td>0.064</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.051</td>
<td>0.048</td>
<td>0.046</td>
<td>0.040</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.063</td>
<td>0.054</td>
<td>0.049</td>
<td>0.041</td>
<td>0.046</td>
</tr>
<tr>
<td>600</td>
<td>0</td>
<td>0.047</td>
<td>0.046</td>
<td>0.048</td>
<td>0.048</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.054</td>
<td>0.053</td>
<td>0.052</td>
<td>0.052</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.042</td>
<td>0.038</td>
<td>0.057</td>
<td>0.050</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.050</td>
<td>0.041</td>
<td>0.056</td>
<td>0.064</td>
<td>0.063</td>
</tr>
</tbody>
</table>

Table 1b. Empirical power of the $Q_{M-}$ and $Q_{B-}$ statistics.

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$\rho$</td>
<td>$Q_{M-}$</td>
<td>$Q_{B-}$</td>
<td>$Q_{M-}$</td>
<td>$Q_{B-}$</td>
<td>$Q_{M-}$</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>0.604</td>
<td>0.600</td>
<td>0.638</td>
<td>0.634</td>
<td>0.641</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.574</td>
<td>0.566</td>
<td>0.584</td>
<td>0.581</td>
<td>0.610</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.461</td>
<td>0.447</td>
<td>0.441</td>
<td>0.428</td>
<td>0.440</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.307</td>
<td>0.290</td>
<td>0.257</td>
<td>0.246</td>
<td>0.264</td>
</tr>
<tr>
<td>400</td>
<td>0</td>
<td>0.851</td>
<td>0.850</td>
<td>0.874</td>
<td>0.873</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.840</td>
<td>0.838</td>
<td>0.840</td>
<td>0.835</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.730</td>
<td>0.723</td>
<td>0.665</td>
<td>0.658</td>
<td>0.702</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.504</td>
<td>0.486</td>
<td>0.476</td>
<td>0.452</td>
<td>0.426</td>
</tr>
<tr>
<td>600</td>
<td>0</td>
<td>0.953</td>
<td>0.955</td>
<td>0.955</td>
<td>0.955</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.950</td>
<td>0.948</td>
<td>0.949</td>
<td>0.948</td>
<td>0.959</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.876</td>
<td>0.873</td>
<td>0.856</td>
<td>0.853</td>
<td>0.851</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.687</td>
<td>0.677</td>
<td>0.614</td>
<td>0.596</td>
<td>0.573</td>
</tr>
</tbody>
</table>

Here $\alpha_0 = 0.00001; \alpha_1 = 0.2, 0.3, 0.4$ or $0.5; \rho = 0$ or $0.3$; and $\rho = 0, 0.3, 0.6$ or $0.9$. $\varepsilon_{1t}$ and $\varepsilon_{2t}$ are independent $N(0,1)$ sequences. The choice of $\alpha_0$ is inspired by an example of Engle (1983). In his case, $\alpha_0 = 0.00006$, and our value is a comparable one.

Clearly at $\rho = 0$, the empirical power is the empirical size. When $\rho = 0$, both the innovation and squared innovation series of $X_{1t}$ and $X_{2t}$ are uncorrelated. In this case, $Q_{M-}$ and $Q_{B-}$ will be asymptotically equivalent. When $\rho \neq 0$, as argued at the end of Subsection 3.1, we expect the size of $Q_{B-}$ to be conservative asymptotically. 1000 simulations were performed for every possible combination of the sample sizes, the perturbations, the correlations and the $\alpha_1$ values. Simulation results for the size and power of $Q_{M-}$ and $Q_{B-}$ were summarised in Table 1a and Table 1b. The degrees of freedom for
the two statistics are 4 and hence the critical value at 5% level is 9.4877. For Table 1a, when \( n = 600 \) and \( \rho = 0 \), the average size of the \( Q_{M-} \) and \( Q_{B-} \) statistics are 0.0510 and 0.0508 respectively in the 4000 simulations. 0.05 is of course the theoretical size. This confirms our earlier assertion that the two statistics are asymptotically equivalent. On the other hand, when \( n = 600 \) and \( \rho \neq 0 \), the average size of \( Q_{M-} \) and \( Q_{B-} \) are respectively 0.0502 and 0.0463 in the 12,000 simulations. Thus as expected, \( Q_{B-} \) is conservative and \( Q_{M-} \) is almost equal to the theoretical value. Now in financial data analysis, the need of large data sets is self-explanatory. Moreover, examples of the presence of instantaneous dependence are abundant. Thus the \( Q_{M-} \) statistic should be of practical importance. From Table 1b, recalling that our perturbation term is not large, we see that the power of the portmanteau statistics is good for a moderate sample size, say \( m \geq 400 \).

**Example 2.** Cross-correlations and score tests for data generated by multivariate ARCH models. It will be interesting to see the performance of the test statistics for data generated by the same mechanism. We consider the following two multivariate ARCH models:

\[
\text{Model 1: } \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}, \quad d_t = \begin{pmatrix} 0.01 + 0.3a_{1t-1}^2 & 0 \\ 0 & 0.01 + 0.5a_{2t-1}^2 \end{pmatrix};
\]

and

\[
\text{Model 2: } \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}, \quad d_t = \begin{pmatrix} 0.005 + 0.5a_{1t-1}^2 & 0 \\ 0 & 0.005 + 0.4a_{2t-1}^2 \end{pmatrix}.
\]

Here \((a_{1t}a_{2t})^T\) are assumed to be first order white noise sequences, and \(d_t\) is the conditional covariance matrix at time \(t\). Note that, in Model 1, the bivariate series \((X_{1t}X_{2t})^T\) can be regarded as two univariate series. Similar logic holds for Model 2. In considering the power of the test statistics, the innovation equations would be

\[
\text{Model 1': } d_t = \begin{pmatrix} E_t & 0 \\ 0 & H_t \end{pmatrix}
\]

where \(E_t = 0.01 + 0.3a_{1t-1}^2 + 0.5a_{2t-2}^2\), \(H_t = 0.01 + 0.5a_{2t-1}^2 + 0.25a_{1t-2}^2\);

\[
\text{Model 2': } d_t = \begin{pmatrix} E_t & 0 \\ 0 & H_t \end{pmatrix}
\]

where \(E_t = 0.005 + 0.5a_{1t-1}^2 + 0.3a_{2t-2}^2\), \(H_t = 0.005 + 0.4a_{2t-1}^2 + 0.2a_{1t-2}^2\).

Three sample sizes, \(n = 100, 200, \) and 400 were considered in the simulations. 1000 simulations were performed from each model. For the cross-correlations tests, we still consider the first four cross-correlation coefficients. For the score test, we consider a

<table>
<thead>
<tr>
<th>(n)</th>
<th>Model 1</th>
<th>(Q_{M-})</th>
<th>(Q_{B-})</th>
<th>Score</th>
<th>(Q_{M-})</th>
<th>(Q_{B-})</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.044</td>
<td>0.041</td>
<td>0.047</td>
<td>0.034</td>
<td>0.032</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.057</td>
<td>0.054</td>
<td>0.043</td>
<td>0.055</td>
<td>0.051</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.050</td>
<td>0.046</td>
<td>0.053</td>
<td>0.048</td>
<td>0.048</td>
<td>0.047</td>
<td></td>
</tr>
</tbody>
</table>
Table 2b. Empirical power of the cross-correlation and score test statistics.

<table>
<thead>
<tr>
<th>n</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q_M$-</td>
<td>$Q_B$-</td>
<td>Score</td>
<td>$Q_M$-</td>
</tr>
<tr>
<td>100</td>
<td>0.885</td>
<td>0.868</td>
<td>0.746</td>
<td>0.271</td>
</tr>
<tr>
<td>200</td>
<td>0.992</td>
<td>0.988</td>
<td>0.928</td>
<td>0.443</td>
</tr>
<tr>
<td>400</td>
<td>0.998</td>
<td>0.998</td>
<td>0.987</td>
<td>0.712</td>
</tr>
</tbody>
</table>

second order model versus a first order model. Results are summarized in Table 2a and Table 2b.

It is interesting to observe that the empirical power of the cross-correlation and score tests differ in the two models. In Model 1 cross-correlation is better, but in Model 2, it is the other way round. This illustrates that the magnitude of the coefficients in the models affect the power of the tests.

4.2 An empirical example—dependence in variance in the New York and Sydney stock indices

As an illustrative example for the cross-correlation and the Lagrange multiplier (LM) test, we consider the first difference of logs i.e. the return of the Standard and Poor’s 500 (SP500) and the Sydney All Ordinaries (SAO) stock indices during the period 1993 to 1997. Data were obtained from Datastream. The data are the daily closing price of the two indices and the number of observations for each index is 1242. The return series were scaled up by a factor of 100. Let $X_{1t}$ and $X_{2t}$ be the SP500 and the SAO series respectively. Based on previous reports (Bera and Higgins (1993); Bollerslev et al. (1994)), it is reasonable to fit GARCH(1,1) models to the individual series. The GARCH(1,1) models were found to be:

\[
X_{1t} = \varepsilon_{1t},
\]

\[
h_{1t} = 0.0173 + 0.0847 \varepsilon_{1t-1}^2 + 0.888 h_{1t-1},
\]

and

\[
X_{2t} = \varepsilon_{2t},
\]

\[
h_{2t} = 0.202 + 0.153 \varepsilon_{2t-1}^2 + 0.532 h_{2t-1}.
\]

Here values in brackets were standard errors. Now Table 3 shows the sample autocorrelation coefficients (SAC) and sample cross-correlation coefficients (SCC) of the standardized squared residuals from the univariate GARCH models.

The conventional error band is $\pm \frac{2}{\sqrt{1241}} = 0.0568$. Clearly the whole picture was dominated by $\hat{r}_{12}(1), 0.41$, meaning that the volatility of the SAO index was driven overwhelmingly by the SP500 index. The $Q_{M+}$ and $Q_{M-}$ statistics are 215.99 and 6.08 respectively. This phenomenon was natural from conventional market wisdom. Table 3 also indicates that a first order multivariate model for the innovation equation is reasonable. For the observation equation, we look at the sample autocorrelations and cross-correlations of the two return series. They are summarized in Table 4. Note that each of the two series is a white noise sequence.

Again the whole picture was dominated by the cross-correlation at lag 1. The table suggested that the conditional mean might be fitted as an MA model, with order at
Table 3. SAC and SCC for standardized squared residuals from GARCH(1,1) models.

<table>
<thead>
<tr>
<th>Lag</th>
<th>SAC (SP500)</th>
<th>SAC (SAO)</th>
<th>SCC (SP500 &amp; SAO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.064</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$4.4 \times 10^{-4}$</td>
<td>$-3.5 \times 10^{-3}$</td>
<td>0.41</td>
</tr>
<tr>
<td>2</td>
<td>$-0.41 \times 10^{-3}$</td>
<td>0.023</td>
<td>0.015</td>
</tr>
<tr>
<td>3</td>
<td>-0.022</td>
<td>0.019</td>
<td>$8.4 \times 10^{-3}$</td>
</tr>
<tr>
<td>4</td>
<td>0.037</td>
<td>$9.4 \times 10^{-4}$</td>
<td>$7.1 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 4. SAC and SCC for SP500 and SAO return series.

<table>
<thead>
<tr>
<th>Lag</th>
<th>SAC (SP500)</th>
<th>SAC (SAO)</th>
<th>SCC (SP500 &amp; SAO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.079</td>
<td>0.079</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.022</td>
<td>0.040</td>
<td>0.522</td>
</tr>
<tr>
<td>2</td>
<td>-0.044</td>
<td>-0.012</td>
<td>-0.071</td>
</tr>
<tr>
<td>3</td>
<td>-0.049</td>
<td>0.014</td>
<td>-0.005</td>
</tr>
<tr>
<td>4</td>
<td>0.041</td>
<td>-0.040</td>
<td>0.013</td>
</tr>
</tbody>
</table>

most 2. Thus both MA(1)–GARCH(1, 1) and MA(2)–GARCH(1, 1) model were fitted to the data. However, in fitting the MA(2)–GARCH(1, 1) model, it was found that there is a negative coefficient in the innovation equation and the lag 2 coefficients for the observation equations were insignificant. The lag 2 coefficients and their standard errors (in brackets) are shown in the following matrix:

$$
\begin{pmatrix}
0.0216(0.0275) & -0.0077(0.0240) \\
-0.0128(0.0325) & -0.0234(0.0331)
\end{pmatrix}
$$

Thus the MA(2)–GARCH(1, 1) model was abandoned.

The MA(1)–GARCH(1, 1) can be represented as:

$$
\begin{pmatrix}
X_{1t} \\
X_{2t}
\end{pmatrix} = \begin{pmatrix}
0.0841(0.0265) & 0.0161(0.0259) \\
0.5176(0.0293) & 0.0719(0.0303)
\end{pmatrix} \begin{pmatrix}
a_{1t-1} \\
a_{2t-1}
\end{pmatrix} + \begin{pmatrix}
a_{1t} \\
a_{2t}
\end{pmatrix} + d_t = \text{diag}(h_{1t}, h_{2t}),
$$

where

$$
\begin{pmatrix}
h_{1t} \\
h_{2t}
\end{pmatrix} = \begin{pmatrix}
0.0020(0.0019) \\
0.0289(0.0136)
\end{pmatrix} + \begin{pmatrix}
0.0821(0.0277) & 0.0115(0.0137) \\
0.0389(0.0179) & 0.0849(0.0311)
\end{pmatrix} \begin{pmatrix}
a_{1t-1} \\
a_{2t-1}
\end{pmatrix}$$
\[ 
\begin{pmatrix}
0.8930(0.033) & 0 \\
0 & 0.5475(0.1601)
\end{pmatrix}
\begin{pmatrix}
{h_1}_{t-1} \\
{h_2}_{t-1}
\end{pmatrix}. 
\]

Note here in the coefficient matrix of \((a^2_{1t-1} a^2_{2t-1})^T\), the \((2, 1)\)-th entry is 0.0389, which has a t-ratio greater than 2, indicating that the Australian index is driven by the American index. Results in Table 3 actually suggest that a higher order model in the innovation equation is unnecessary. To confirm this, we perform the score test for an additional second order ARCH term in the above model. The extra number of parameters introduced were four and the LM statistic would thus be based on a \(\chi^2_4\) distribution. The information matrix was obtained via the BHHH algorithm and the LM statistic was 7.33, which was not significant at the 5% level.

In all, the simulations and the last example show that both tests have good empirical performance. As mentioned in the beginning, there is as yet not too much work on the detecting and diagnostic checking of multivariate conditional heteroscedastic models. We hope these tests can serve as tools to handle the relevant problems.

Acknowledgements

The authors would like to thank the referees for comments that led to the improvement of the paper. The authors would like to thank Dr. W. S. Chan for computational assistance.

REFERENCES


