Large Scale Conditional Covariance Matrix

Modeling, Estimation and Testing

Zhuanxin Ding[1]

Frank Russell Company, Tacoma, WA 98401, USA

and

Robert F. Engle

University of California, San Diego, La Jolla, CA 92093, USA

June 1996

This Revision May 2001

Abstract

A new representation of the diagonal *Vech* model is given using the Hadamard product. Sufficient conditions on parameter matrices are provided to ensure the positive definiteness of covariance matrices from the new representation. Based on this, some new and simple models are discussed. A set of diagnostic tests for multivariate ARCH models is proposed. The tests are able to detect various model misspecifications by examing the orthogonality of the squared normalized residuals. A small Monte-Carlo study is carried out to check the small sample performance of the test. An empirical example is also given as guidance for model estimation and selection in the multivariate framework. For the specific data set considered, it is found that the simple one and two parameter models and the constant conditional correlation model perform fairly well.

Keywords: conditional covariance, Multivariate ARCH, Hadamard product, M-test.

---

[1] We would like to thank Clive Granger, Hong-Ye Gao, Doug Martin, Jeffrey Wang and Hal White for many helpful discussions.

1. Introduction

It has now become common knowledge that the risk of the financial market, which is mainly represented by the variance of individual stock returns and the covariance between different assets or with the market, is highly forecastable. The research in this area however is far from complete and the application of different econometric models is just at its beginning. Most research in this area focuses on the modeling and forecasting of stock volatilities instead of the covariation of different assets even though the covariance plays a key role in the standard CAPM theory and the mean-variance asset allocation model. One main reason for this is probably the complexity in modeling the conditional covariance of multiple time series faced by everybody who tries to tackle this problem. One difficulty in modeling the conditional second moment for multiple time series is that the model usually involves a huge number of parameters which is the same as in the VAR model. But there is one more complexity besides this which is the fact that every model must ensure that the estimated conditional covariance matrix is positive definite.

This paper starts with the *Vech* model of Bollerslev, Engle and Wooldridge (1988) and points out that the *Vech* model solves neither of the two difficulties above. A representation using Hadamard product in matrix algebra is then given to the Diagonal *Vech* model. Sufficient conditions on parameter matrices are provided to guarantee the conditional covariance matrices to be positive definite from this model. A sequence of simplified models is then presented and their properties discussed. Section 2 also presents another new class of multivariate ARCH model which we call Principal Component MARCH model. Two other models, namely the Constant Conditional Correlation model of Bollerslev(1992) and the BEKK model of Engle and Kroner (1995), are also discussed.

With so many possible models one can use, it is necessary to discriminate between them in application. Although varieties of different kinds of tests in analyzing univariate time series can be used directly in analyzing multiple time series, there are special properties which need to be considered in building conditional covariance matrices. Two sets of diagnostic tests are proposed in section 3. The tests are designed to detect different kinds of model misspecification. The first set of tests is essentially a moment test to check cross-variable and time series dependence. The second set of tests is an LM test to examine whether the conditional covariance matrix is right on average and also whether there is time series dependence. Section 4 gives a small Monte-Carlo study to compare the critical values from finite sample simulation with their asymptotic counterparts.

Section 5 gives an empirical example of estimation and testing. Different models are estimated for stock returns of five US companies. A comparison of the test statistic from different models is performed. This section offers guidance for selecting better models for a specific data set, especially when the system is large. Section 6 concludes the paper.

2. Varieties of Multivariate ARCH Models

In a regression context, the full model that will be considered in this paper is as follows,

$$\Gamma \mathbf{Y}_t + \Pi \mathbf{X}_t = \boldsymbol{\varepsilon}_t ,$$

$$\boldsymbol{\varepsilon}_t \mid \Psi_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t) ,$$

where Γ is an $N\times N$ parameter matrix, $\mathbf{Y}_t$ is an $N\times 1$ vector of the endogenous variables; Π is an $N\times K$ parameter matrix, $\mathbf{X}_t$ is a $K\times 1$ vector of exogenous variables in mean equation (including constant and ARMA term); $\boldsymbol{\varepsilon}_t$ is the residuals vector of $N\times 1$ from this regression model, Ψ_{t-1} is the information set containing all the information available up to time t–1, and $\mathbf{H}_t = (h_{ijt})$ is the conditional covariance matrix we will discuss below. Different specifications of $\mathbf{H}_t$ will give us different multivariate ARCH models.

The first and probably the most general multivariate ARCH model is Bollerslev, Engle and Wooldridge's *Vech* representation of the multivariate ARCH model. [see Bollerslev *et al.* (1988).] In *Vech* model, the conditional covariance matrix is specified as follows,

$$Vech(\mathbf{H}_t) = \mathbf{C} + \sum_{i=1}^{q} \mathbf{A}_i Vech(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}') + \sum_{j=1}^{p} \mathbf{B}_j Vech(\mathbf{H}_{t-j})\,. \tag{2.1}$$

Where *Vech* (.) denotes the column stacking operator of the lower portion of a symmetric matrix. So for an $N \times N$ covariance matrix $\mathbf{H}_t$, $Vech(\mathbf{H}_t)$ is an $N(N+1)/2 \times 1$ vector. In equation (2.1), $\mathbf{C}$ is also $N(N+1)/2 \times 1$, and $\mathbf{A}_i$, i =1,...,q and $\mathbf{B}_j$, j =1,...,p, are $N(N+1)/2 \times N(N+1)/2$ matrices.

Although the *Vech* representation is very general, it has two major disadvantages in applications. Firstly, the model involves (only for conditional covariance equation) $N(N+1)/2+(p+q)N^2(N+1)^2/4$ parameters. Even for a small size system, the number of parameters to be estimated is prohibitive. For ease of exposition, p and q will always be assumed to equal 1 throughout the paper so that only multivariate GARCH(1,1) models will be discussed. It is straightforward to generalize the model to p, q>1. For example, when N=5 and p=q =1 the number of parameters involved in the equation is 465 which will prove to be very costly to estimate if it is not impossible. For a relatively large system of N =10, the total parameters involved increases to 6105 which is merely impossible to implement. And secondly, the estimated conditional covariance matrices are not guaranteed to be positive definite. Every covariance matrix must be positive definite, but for this model it is probably impossible to give general restrictions on parameters to insure a positive definite covariance matrix. This shortcoming makes the model not reliable in applications.

To solve the first problem, Bollerslev *et al.* (1988) proposed the much simpler diagonal *Vech* model by restricting the coefficient matrices **A** and **B** in equation (2.1) to be diagonal so that each conditional covariance depends only on its own past values and surprises. That is, the conditional covariance is decided by the following equation,

$$h_{ijt} = \omega_{ij} + \alpha_{ij}\varepsilon_{it-1}\varepsilon_{jt-1} + \beta_{ij}h_{ijt-1}, \quad i,j = 1,\ldots,N. \tag{2.2}$$

Under this restriction only $(p+q+1)N(N+1)/2$ parameters are involved for an $N \times N$ system. For example, when p=q=1 and N=5, the total number of parameters to be estimated is 15×3 =45. When N =10, this number is 165 which is much smaller than that of a *Vech* model.

In this section, we will first give a sufficient condition for the diagonal *Vech* model to be positive definite. And based on this, a sequence of simplified models which all guarantee the conditional covariance matrix to be positive definite will then be proposed.

It is not difficult to see that the diagonal *Vech* model defined by (2.2) is equivalent to the following representation:

$$\mathbf{H}_t = \mathbf{C} + \mathbf{A} \otimes \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \mathbf{B} \otimes \mathbf{H}_{t-1}, \tag{2.3}$$

where $\mathbf{H}_t$ ($N \times N$) is the conditional covariance matrix at time t, **C, A** and **B** are all $N \times N$ parameter matrices. Special attention should be paid to $\otimes$ which denotes the Hadamard product of two matrices. If $\mathbf{U} = (u_{ij})$ and $\mathbf{V} = (v_{ij})$ are each $m \times n$ matrices, then their Hadamard product is the $m \times n$ matrix of elementwise products

$$\mathbf{U} \otimes \mathbf{V} = (u_{ij}\, v_{ij}) .$$

Since $\mathbf{H}_t$ must be symmetric in equation (2.3), so must be the parameter matrices **C, A,** and **B**. Hence only the lower portion of matrices **C, A** and **B** need to be parameterized and estimated.

Some results for Hadamard product are needed before we present the conditions that will guarantee the diagonal *Vech* model to be positive definite. It will be assumed throughout the discussion in this paper that the usual matrix or vector multiplication will be carried out before the Hadamard product. So $\mathbf{AA}' \otimes \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'$ should be interpreted as $(\mathbf{AA}') \otimes (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}')$.

Lemma 2.1
If **U** *is a symmetric matrix of* $N \times N$ *,* **v** *is a nonzero vector of* $N \times 1$*, then* $\mathbf{U} \otimes \mathbf{vv}'$ *is positive semi-definite iff* **U** *is positive semi-definite.*
Proof:
By just writing them out element by element, one can easily see that

$$\mathbf{U} \otimes \mathbf{vv}' = diag\,[\mathbf{v}]\ \mathbf{U}\ diag\,[\mathbf{v}],$$

where $diag\,[\mathbf{v}]$ is a diagonal matrix with v_i as its i th diagonal element. But the right hand side is positive semi-definite iff **U** is positive semi-definite.

Theorem 2.2
If both **U** *and* **V** *are positive semi-definite matrices of* $N \times N$*, then so is* $\mathbf{U} \otimes \mathbf{V}$.
Proof:
Since **V** is positive semi-definite one can get a spectral decomposition of **V**,

$$\mathbf{V} = \sum_{i=1}^{N} \lambda_i \mathbf{v}_i \mathbf{v}_i'$$

such that $\lambda_i \geq 0$ for all i where λ_i's are the eigenvalues of **V** and $\mathbf{v}_i$'s are the corresponding eigenvectors. Hence

$$\begin{aligned} \mathbf{U} \otimes \mathbf{V} &= \mathbf{U} \otimes \sum_{i=1}^{N} \lambda_i \mathbf{v}_i \mathbf{v}_i' \\ &= \sum_{i=1}^{N} \lambda_i (\mathbf{U} \otimes \mathbf{v}_i \mathbf{v}_i') . \end{aligned}$$

By Lemma 2.1, $\mathbf{U} \otimes \mathbf{v}_i \mathbf{v}_i$ is positive semi-definite if **U** is positive semi-definite, and since $\lambda_i \geq 0$, $i = 1,\ldots,N$, so $\lambda_i (\mathbf{U} \otimes \mathbf{v}_i \mathbf{v}_i')$ is positive semi-definite. The result is proved since $\mathbf{U} \otimes \mathbf{V}$ is now the sum of N positive semi-definite matrices.

For a detailed discussion of Hadamard product, one is referred to Styan(1973). There a different proof is given to theorem 2.2 .

We now present four sequentially nested multivariate ARCH models from general to simple. They are all special cases of the diagonal *Vech* model and are all guaranteed to be positive semi-definite. It will be assumed that all the parameter matrices are either symmetric or lower triangle and so there are only $N(N+1)/2$ nonredundant parameters in a $N\times N$ parameter matrix.

Matrix-Diagonal Model

$$\mathbf{H}_t = \mathbf{CC}' + \mathbf{AA}' \otimes \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \mathbf{BB}' \otimes \mathbf{H}_{t-1} \tag{2.4}$$

Since $\mathbf{CC}'$, $\mathbf{AA}'$ and $\mathbf{BB}'$ are all positive semi-definite, so by theorem 2.2, $\mathbf{H}_t$ will be positive definite for all t as far as the initial covariance matrix $\mathbf{H}_0$ is positive definite. If sample covariance is used for $\mathbf{H}_0$ then $\mathbf{H}_t$ will always be positive definite. Let $\omega_{ij} = (\mathbf{CC}')_{ij}$, $\alpha_{ij} = (\mathbf{AA}')_{ij}$, $\beta_{ij} = (\mathbf{BB}')_{ij}$, then it is readily seen that,

$$h_{ijt} = \omega_{ij} + \alpha_{ij}\varepsilon_{it-1}\varepsilon_{jt-1} + \beta_{ij} h_{ijt-1}, \quad i,j = 1, \ldots, N. \tag{2.5}$$

So each conditional covariance depends on its own past values and surprises. The difference between this representation and Bollerslev *et al.* 's diagonal *Vech* representation is that the parametrization used here imposed restrictions implicitly among different parameters to assure that the parameter matrix is positive semi-definite, and which will further assure the conditional covariance matrices positive definite. By writing the parameter matrices in the form of $\mathbf{CC}'$, $\mathbf{AA}'$, and $\mathbf{BB}'$ instead of just $\mathbf{C}$, $\mathbf{A}$, and $\mathbf{B}$, the positive semi-definiteness is guaranteed in estimation without imposing any further restrictions.

The following three models are simplified versions of Matrix-Diagonal Model. By imposing extra conditions on the model we substantially reduce the parameters that need to be estimated while keeping the estimated conditional covariance matrices positive definite.

Vector-Diagonal Model

$$\mathbf{H}_t = \mathbf{CC}' + \mathbf{aa}' \otimes \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \mathbf{bb}' \otimes \mathbf{H}_{t-1} \tag{2.6}$$

where $\mathbf{a}$ and $\mathbf{b}$ are $N\times 1$ vectors. Here we impose restrictions on parameter matrix $\mathbf{A}$ and $\mathbf{B}$ to be only rank one.

Scalar-Diagonal or Two-parameter Model

$$\mathbf{H}_t = \mathbf{CC}' + \alpha\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \beta\mathbf{H}_{t-1} \tag{2.7}$$

where α, β are positive scalars. This model implies any linear combinations of the original series will have a GARCH representation with the same ARCH and GARCH parameter. This is obviously a very strong restriction. Each conditional covariance is as follows,

$$h_{ijt} = \omega_{ij} + \alpha\varepsilon_{it-1}\varepsilon_{jt-1} + \beta h_{ijt-1}, \quad i,j = 1, \ldots, N. \tag{2.8}$$

Integrated MARCH Model

$$\mathbf{H}_t = \alpha \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' + (1-\alpha)\mathbf{H}_{t-1} \tag{2.9}$$

This model is from the two-parameter model by assuming $\alpha + \beta = 1$ and **C**=**0**. Following the practice in univariate ARCH model we will call this Integrated Multivariate ARCH model.

A brief summary of these models is given in Table 2.1. These models are basically in the same group. One get more parsimonious models by imposing restrictions on the parameter matrices **A** and **B** in Matrix-Diagonal Model. One may find the restrictions are too strong in order to get a positive definite covariance matrix. In fact the above restrictions are sufficient but may not be necessary for the covariance matrix to be positive definite since **U** + **V** can be positive definite even if **V** is not. For more discussion on these and some other models, see Ding (1994).

For these models, a consistent estimator for the matrix **C** can be derived from the estimator for **A**, **B** and the sample covariance matrix if the system is covariance stationary. For example in the two-parameter model if $\alpha + \beta < 1$ then $E\mathbf{H}_t$, $E\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'$ exist and are both equal to the unconditional covariance matrix. So

$$\begin{aligned} E\mathbf{H}_t &= \mathbf{CC}' + \alpha E\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \beta E\mathbf{H}_{t-1} \\ &= \mathbf{CC}'/(1-\alpha-\beta)\,. \end{aligned} \tag{2.10}$$

But we know the sample covariance matrix is a consistent estimator of the unconditional covariance matrix when the system is stationary. So let

$$\mathbf{CC}'/(1-\alpha-\beta) = \mathbf{H}_0$$

then one gets

$$\mathbf{CC}' = \mathbf{H}_0(1-\alpha-\beta)\,.$$

Hence the two-parameter model becomes

$$\mathbf{H}_t = \mathbf{H}_0(1-\alpha-\beta) + \alpha\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + \beta\mathbf{H}_{t-1} \tag{2.11}$$

which is a *real* two-parameter model. The same arguments hold for the other models. This approach is called Variance Targeting by Engle and Mezrich(1996).

One other model in this class worth to be mentioned here is the Exponentially Weighted Moving Average model. It is defined as follows:

EWMA Model

$$\mathbf{H}_t = \frac{(1-\alpha)}{(1-\alpha^{t-1})}\sum_{i=1}^{t-1}\alpha^{i-1}\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}_{t-i}' \tag{2.12}$$

This model is widely used in industry, and is known as Exponentially Weighted Moving Average (EWMA) model. It is essentially the same as the Integrated Multivariate ARCH model. For this model, $\mathbf{H}_t$ is positive definite only when $t > N$.

Although the two-parameter model and the EWMA model are very simple and may impose too strong restrictions on parameters, they are found to be very useful and able to provide a reasonable approximation when one wants to build covariance forecast models for a large system.

Three other models will also be considered and be compared in this paper. They are all proposed to overcome the two disadvantages mentioned earlier for *Vech* model. They are the Constant Conditional Correlation (CCC) model of Bollerslev (1992), the BEKK model of Engle and Kroner (1995), and finally, the Principal Component (PRCOMP) ARCH model of Kahn (1992).

Constant Conditional Correlation Model

We first give the Constant Conditional Correlation Multivariate ARCH model. The model assumes that each individual series has a univariate ARCH structure but the conditional correlation is constant. i.e.

$$h_{iit} = \omega_i + \alpha_i \varepsilon_{it-1}^2 + \beta_i h_{iit-1}, \ i = 1, \ldots, N. \tag{2.13}$$

and

$$h_{ijt} = \rho_{ij} \sqrt{h_{iit}} \sqrt{h_{jjt}}, \ \ i = 1, \ldots, N, \ j = 1, \ldots, N. \tag{2.14}$$

Let $\mathbf{R} = (\rho_{ij})$ and $\mathbf{D}_t$ be a diagonal matrix with $\sqrt{h_{iit}}$, i=1,…,N on its diagonal entries, then the conditional covariance matrix is,

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t .$$

In this paper we will use Ding, Granger and Engle's Asymmetric Power ARCH model [see Ding, Granger, Engle (1993)] as a more generalized structure for each univariate series. So one has,

$$h_{iit} = s_{iit}^2, \quad i = 1, \ldots, N,$$

and

$$s_{iit}^{\delta_i} = \omega_i + \alpha_i (|\varepsilon_{it-1}| + \gamma_i \varepsilon_{it-1})^{\delta_i} + \beta_i s_{iit-1}^{\delta_i}, \ i = 1, \ldots, N.$$

By doing this seven other univariate ARCH models are nested in this model here.

BEKK Model

We next present the BEKK representation of the multivariate conditional covariance matrix. The model is first proposed by Baba, Engle, Kraft and Kroner (1991). The name BEKK is after the authorship of that paper. The latest published version of that paper is Engle and Kroner (1995). The BEKK representation embeds the covariance matrix in such a structure that the positive definiteness of the covariance matrix is guaranteed. The simplest BEKK model is as follows:

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}' \tag{2.15}$$

This model imposes restrictions over parameters across equations. More general versions of the model are available by putting more terms in the right hand side of the equation. As pointed out by Engle and Kroner (1995), this model includes all positive definite diagonal *Vech* models and nearly all positive definite *Vech* representations. The BEKK representation solves the positive definiteness problem successfully but the first disadvantage of the *Vech* model remains. Even for the simplest form given in (2.15), the number of parameters are the same order of magnitude that of any diagonal model.

If **A** and **B** are diagonal, Lemma 2.1 reveals that the BEKK model is simply the Vector Diagonal model defined above.

Principal Component MARCH Model

The final model we will discuss here is first suggested by Ron Kohn of BARRA and we will call it the Principal Component Multivariate ARCH model. The main idea is to assume that there exists an orthogonal transformation of the original N time series to get N independent series which follow univariate ARCH processes. The formal model is as follows. Assume

$$\boldsymbol{\varepsilon}_t \mid \Psi_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t), \tag{2.16}$$

and

$$\mathbf{H}_t = \mathbf{P}\boldsymbol{\Lambda}_t\mathbf{P}', \tag{2.17}$$

where **P** is an orthonormal matrix with $\mathbf{P}' = \mathbf{P}^{-1}$, and $\boldsymbol{\Lambda}_t$ is a diagonal matrix changing over time, $\boldsymbol{\Lambda}_t =$ *diag* [$h_{11t}, h_{22t}, \ldots, h_{NNt}$]. Under this assumption we have,

$$\boldsymbol{\varepsilon}_t \mid \Psi_{t-1} \sim N(\mathbf{0}, \mathbf{P}\boldsymbol{\Lambda}_t\mathbf{P}'),$$

that is

$$\mathbf{e}_t = \mathbf{P}'\boldsymbol{\varepsilon}_t \mid \Psi_{t-1} \sim N(\mathbf{0}, \boldsymbol{\Lambda}_t).$$

The final assumption is that each e_{it} follows a univariate ARCH process with mean 0 and conditional variance h_{iit}. Having got the conditional variance estimation for the transformed time series one can easily get the conditional covariance matrix estimation of the original time series by simply transforming back using equation (2.28). Obviously it may not be reasonable to assume that **P** is constant over time here. But the model is still attractive because of its simplicity. Usually the largest two principal components will account for about 90% of the volatility of the whole system.

Table 2.1 gives a summary of the functional forms of the above models. The relationship between different models are clearly seen. Table 2.1 also gives the number of parameters needed in conditional covariance equation when there are *N* equations. Two special cases are given when *N*=5 and *N*=20.

Table 2.1 Summary of the multivariate ARCH models ($\boldsymbol{\Sigma}_{t-1} = \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'$)

model	functional form	#par	N=5	N=20
Matrix-Diagonal	$\mathbf{H}_t = \mathbf{CC}' + \mathbf{AA}' \otimes \boldsymbol{\Sigma}_{t-1} + \mathbf{BB}' \otimes \mathbf{H}_{t-1}$	3N(N+1)/2	45	630
Variance Targeting	$\mathbf{H}_t = \mathbf{H}_0 \otimes (\mathbf{ii}'\text{-}\mathbf{AA}'\text{-}\mathbf{BB}') + \mathbf{AA}' \otimes \boldsymbol{\Sigma}_{t-1} + \mathbf{BB}' \otimes \mathbf{H}_{t-1}$	N(N+1)	30	420
Vector-Diagonal	$\mathbf{H}_t = \mathbf{CC}' + \mathbf{aa}' \otimes \boldsymbol{\Sigma}_{t-1} + \mathbf{bb}' \otimes \mathbf{H}_{t-1}$	N(N+5)/2	25	250
Variance Targeting	$\mathbf{H}_t = \mathbf{H}_0 \otimes (\mathbf{ii}'\text{-}\mathbf{aa}'\text{-}\mathbf{bb}') + \mathbf{aa}' \otimes \boldsymbol{\Sigma}_{t-1} + \mathbf{bb}' \otimes \mathbf{H}_{t-1}$	2N	10	40
Two-parameter	$\mathbf{H}_t = \mathbf{CC}' + \alpha\, \boldsymbol{\Sigma}_{t-1} + \beta\, \mathbf{H}_{t-1}$	N(N+1)/2+2	17	212
Variance Targeting	$\mathbf{H}_t = \mathbf{H}_0 (1-\alpha-\beta) + \alpha\, \boldsymbol{\Sigma}_{t-1} + \beta\, \mathbf{H}_{t-1}$	2	2	2
Integrated MARCH	$\mathbf{H}_t = \alpha\, \boldsymbol{\Sigma}_{t-1} + (1-\alpha)\, \mathbf{H}_{t-1}$	1	1	1
EWMA	$\mathbf{H}_t = (1-\alpha)/(1-\alpha^{t-1}) \sum_{i=1}^{t-1} \alpha^{i-1} \boldsymbol{\Sigma}_{t-i}$	1	1	1
CCC	$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t$, $\mathbf{D}_t = diag\,[\mathbf{s}_t]$	N(N+9)/2	35	290
BEKK	$\mathbf{H}_t = \mathbf{CC}' + \mathbf{A}\boldsymbol{\Sigma}_{t-1}\mathbf{A}' + \mathbf{BH}_{t-1}\mathbf{B}'$	3N(N+1)/2	45	630
PRCOMP	$\mathbf{H}_t = \mathbf{P}\boldsymbol{\Lambda}_t\mathbf{P}'$, $\boldsymbol{\Lambda}_t = diag\,[\mathbf{h}_t]$	5N	25	100

3. Diagnostic Tests for Multivariate ARCH Models

Given all these possible multivariate ARCH models in section 2, it is important to design a set of multivariate diagnostic tests for assessing the general descriptive validity of these models. Compared to the huge body of diagnostic tests available for univariate models, there are relatively few such tests available in the literature for multivariate models. However, the general principles to carry out diagnostic tests for multivariate models are similar to those for univariate models. The ARCH model proposed by Engle (1982) has the property that although the residuals from the time series regression are not correlated themselves, their squared values are. The univariate ARCH test is thus designed to detect autocorrelation of squared residuals from a regression model. The idea can be easily carried over to do multivariate ARCH tests. In this section, two sets of diagnostic tests will be discussed. One is specifically for multivariate models, the other is a direct extension of the univariate ARCH test to the multivariate situation.

As is assumed in section 2, the residuals from the simultaneous equation, $\boldsymbol{\varepsilon}_t$, follow a conditional multivariate normal distribution with mean zero and covariance matrix $\mathbf{H}_t$, i.e.

$$\boldsymbol{\varepsilon}_t \mid \Psi_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t), \tag{3.1}$$

where $\mathbf{H}_t = \mathbf{H}_t(\mathbf{z}_{t-1}, \theta)$, $\mathbf{z}_{t-1} \in \Psi_{t-1}$ is the lagged information available now and θ is the parameter vector of $l\times 1$ from the parameter space Θ, $\theta\in\Theta$. Various possible forms of $\mathbf{H}_t(\mathbf{z}_{t-1}, \theta)$ are discussed in section 2. If the model is correctly specified and the true values of parameter vector θ is known, then one has,

$$\mathbf{e}_t = \mathbf{H}_t^{-1/2}\boldsymbol{\varepsilon}_t \mid \Psi_{t-1} \sim N(\mathbf{0}, \mathbf{I}_N),$$

i.e. the normalized residuals are independently normally distributed with mean zero and covariance matrix $\mathbf{I}_N$. Let $\mathbf{e}_t^2 = (e_{1t}^2, \ . \ . \ . \ , e_{Nt}^2)'$ be the elementwise square of the vector $\mathbf{e}_t$, then three consequences of model correct specification are,

$A1$). $E(\mathbf{e}_t \mathbf{e}_t') = \mathbf{I}_N$,

$A2$). $cov(e_{it}^2, e_{jt}^2) = 0$, for all $i \neq j$,

and $A3$). $cov(e_{it}^2, e_{jt-k}^2) = 0$, for $k>0$.

Tests of A1 have no power for many important types of misspecifications. Tests of A2 have power to detect non-normality in the conditional distribution. Tests of A3 can be made robust to non-normality and are therefore capable of isolating failures in the dynamic structure of $\mathbf{H}$. A3 are necessary conditions for the correct specification of the covariance equation. As in the univariate ARCH case, it is no longer enough just looking at whether e_{it} and e_{jt} ($i\neq j$) are correlated or not when the conditional constant second moment assumption is removed. Conditions A2 and A3 examine cross-section and time series independence of the normalized residuals. As an illustration, let us look at the following example. If the true conditional distribution of $\boldsymbol{\varepsilon}_t$ is (3.1) with unconditional covariance matrix $\mathbf{H}$, $E\mathbf{H}_t = \mathbf{H}$, but one misspecifies the distribution of $\boldsymbol{\varepsilon}_t$ without heteroskedasticity as,

$$\boldsymbol{\varepsilon}_t \mid \Psi_{t-1} \sim N(\mathbf{0}, \mathbf{H})$$

then the normalized residuals of the misspecified model are:

$$\tilde{\boldsymbol{e}}_t = \mathbf{H}^{-1/2}\boldsymbol{\varepsilon}_t = \mathbf{H}^{-1/2}\mathbf{H}_t^{1/2}\mathbf{e}_t$$

and,

$$
\begin{aligned}
E(\tilde{\boldsymbol{e}}_t\,\tilde{\boldsymbol{e}}_t') &= E\,(\mathbf{H}^{-1/2}\,\mathbf{H}_t^{1/2}\,\mathbf{e}_t\,\mathbf{e}_t'\,\mathbf{H}_t^{1/2}\,\mathbf{H}^{-1/2})\\
&= E\,\{E_{t-1}[\,\mathbf{H}^{-1/2}\,\mathbf{H}_t^{1/2}\;\mathbf{e}_t\,\mathbf{e}_t'\,\mathbf{H}_t^{1/2}\,\mathbf{H}^{-1/2}\,|\Psi_{t-1}]\}\\
&= E\,\{\mathbf{H}^{-1/2}\,\mathbf{H}_t^{1/2}\,E_{t-1}[\,\mathbf{e}_t\,\mathbf{e}_t'\,|\Psi_{t-1}]\,\mathbf{H}_t^{1/2}\,\mathbf{H}^{-1/2}\}\\
&= E\,\{\,\mathbf{H}^{-1/2}\,\mathbf{H}_t^{1/2}\,\mathbf{I}_N\;\mathbf{H}_t^{1/2}\,\mathbf{H}^{-1/2}\}\\
&= E\{\,\mathbf{H}^{-1/2}\,\mathbf{H}_t\,\mathbf{H}^{-1/2}\}\\
&= \mathbf{H}^{-1/2}\,E\mathbf{H}_t\;\mathbf{H}^{-1/2}\\
&= \mathbf{H}^{-1/2}\,\mathbf{H}\,\mathbf{H}^{-1/2}\\
&= \mathbf{I}_N
\end{aligned}
$$

So condition A1 is still satisfied as far as the conditional mean equation is correctly specified. But condition A2 will not hold in general since usually

$$cov(\tilde{e}_{it}^2,\,\tilde{e}_{jt}^2) \neq 0\,.$$

So the misspecification of the conditional covariance equation can result in the violation of condition A2. Thus a test of only A1 will have no power against the misspecification of the conditional covariance matrix. It is also possible that the conditional covariance equation is correctly specified but the conditional normality assumption is violated or both the conditional covariance equation is misspecified and the conditional normality assumption is violated.

If the true conditional distribution is multivariate *t* instead of multivariate normal, i.e.

$$\boldsymbol{\varepsilon}_t\,|\,\Psi_{t-1} \sim t\,(v,\,\mathbf{0},\,\mathbf{H}_t)$$

where $t\,(v\,,\,\mathbf{0},\,\mathbf{H}_t)$ is a multivariate *t* distribution with covariance matrix $\mathbf{H}_t$ and *v* degrees of freedom [see Fang, *et al.* (1990), Johnson and Kotz (1992)]. The density function of the multivariate *t* distribution used here is defined as follows,

$$f(\mathbf{x}) = \frac{\Gamma((v+N)/2)}{(\pi v)^{N/2}\,\Gamma(v/2)\left|\Sigma\right|^{1/2}}\left(1+\frac{\mathbf{x}'\Sigma^{-1}\,\mathbf{x}}{(v-2)}\right)^{-(v+N)/2}$$

where Σ is the covariance matrix and *N* is the number of variables. For this distribution, one has,

$$\mathbf{e}_t = \mathbf{H}_t^{-1/2}\,\boldsymbol{\varepsilon}_t\,|\,\psi_{t-1} \sim\; t\,(v,\,\mathbf{0},\,\mathbf{I}_N).$$

So,

$$E\,\mathbf{e}_t\,\mathbf{e}_t' = \frac{v}{(v-2)}\,\mathbf{I}_N\,. \qquad (3.2)$$

i.e.

$$cov\,(\,e_{it}\;,\,e_{jt}\,) = 0 \text{ for } i \neq j\,.$$

But

$$E(\mathbf{e}_t^2 - E\,\mathbf{e}_t^2)(\mathbf{e}_t^2 - E\,\mathbf{e}_t^2)' = \frac{2v^2}{(v-4)(v-2)}[\,\mathbf{I}_N + \frac{1}{(v-2)}\mathbf{ii}'\,],$$

i.e.

$$cov\,(e_{it}^2, e_{jt}^2) = \frac{2v^2}{(v-4)(v-2)^2}\text{, for } i \neq j\text{ ,} \tag{3.3}$$

which is nonzero. If one normalizes $\mathbf{e}_t$ to have unit variance then condition A1 still holds under conditional t distribution but condition A2 no longer holds. When $v \to \infty$ the multivariate t distribution goes to multivariate normal distribution and equations (3.2) and (3.3) become the same as conditions A1 and A2.

In conclusion, the test given above can either detect the misspecification of the conditional covariance equation or departures from the conditional normality assumption or both.

If the true values of the parameter vector in the system are known exactly, then, the residuals $\boldsymbol{\varepsilon}_t$ and conditional covariance matrix $\mathbf{H}_t$ are also known exactly. Hence the standardized residuals $\mathbf{e}_t$ which is *iid* with multivariate normal distribution. It is then very easy to derive the asymptotic distribution of the sample covariance of $\mathbf{e}_t^2$ by the law of large numbers for *iid* variables. The sample correlation coefficient will usually follow a normal distribution with mean zero and variance $1/T$, where T is the sample size. So T times the squared correlation coefficient will have a $\chi^2(1)$ distribution. But this is not the situation we are dealing with here.

In our case all the parameters are unknown and need to be estimated from the sample data and are thus stochastic themselves. The asymptotic distribution of the sample covariance of the standardized residuals from the maximum likelihood estimated model will be different. Let $\mathbf{m}_t$ be a vector of $N(N\!-\!1)/2\times 1$ with $(e_{it}^2 - 1)(e_{jt}^2 - 1),\ i \neq j$, as its elements, then condition A2 is equivalent to the following moment condition,

$$E\,\mathbf{m}_t = \mathbf{0}. \tag{3.4}$$

The sample moments $\mathbf{m}_T(\hat{\boldsymbol{\theta}}_T) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{m}_t(\hat{\boldsymbol{\theta}}_T)$ should be close to zero in large sample if equation (3.4) is satisfied. Thus, the result for conditional moment test of Newey (1985) and Tauchen (1985) can be adopted here. Let $\mathbf{s}_t = \mathbf{s}_t(\hat{\boldsymbol{\theta}}_T)$ be the score vector of $l \times 1$ at data point t and assume certain regularity conditions satisfied, then by Lemma 2.1 and Theorem 2.2 of Newey (1985), the M statistic for testing model misspecification is simply

$$d_T = TR^2, \tag{3.5}$$

where R^2 is the uncentered R-squared from a regression of 1 on $[\mathbf{m}_t', \mathbf{s}_t']$. d_T will have a asymptotic χ^2 distribution with $N(N+1)/2$ degrees of freedom.

More specification tests can be carried out by just adding more relevant moment conditions as desired.

4. A Monte-Carlo Study

To examine the performance of these two sets of test, a small Monte-Carlo experiment was carried out. The ideal way to do this would be first to generate data from some specific multivariate ARCH model, and then estimate a multivariate ARCH model (either correctly specified or not) and do the moment tests for standardized residuals as described in section 3. In this way one can investigate both the size and the power of the tests. But since on the one hand, there are too many possible models that can be used and no one is known to be very representative. On the other hand, it takes too much computer time even to estimate one large multivariate ARCH model (say 20×20). It is merely impractical to do a Monte-Carlo study in this way. So a simplification is adopted here.

We first generate a 5×1 vector $\mathbf{e}_t$, t =1,...,1000 from a multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_5)$, so e_{it}, e_{jt-k} is independent to each other when $i \neq j$ or $k \neq 0$. $\mathbf{e}_t$ is then used as standardized residuals from the multivariate ARCH estimation. The auxiliary regression is then run without the score function. Five sets of moment conditions will be used in the simulation study. They are:

1). C-test

We use $N(N-1)/2$ sample covariances as regressors in the moment test,

$$m^0_{ijt} = (e^2_{it} - \bar{e}^2_i)(e^2_{jt} - \bar{e}^2_j)$$

for i =1, ..., N −1, and j =i +1 , . . . , N. We will refer to this test as Covariance test (C-test). The test statistic is simply T times the R^2 from the auxiliary regression of 1 on m^0_{ijt}. The test statistic is asymptotically equivalent to T times the sum of $N(N-1)/2$ squared sample correlation coefficients between e^2_{it} and e^2_{jt} for $i \neq j$ which, by the standard law of large numbers for *iid* variables, will have a $\chi^2(N(N-1)/2)$ distribution.

2). CC-test

We use the sum of N ($N(N-1)/2$ sample covariances in C-test as one moment condition:

$$m^0_t = \sum_{ij} m^0_{ijt}$$

This test will be referred to as Composite Covariance test (CC-test). T times the R^2 of the regression 1 on m^0_t will have a $\chi^2(1)$ distribution.

3). LC-test.

We use N^2 first order lagged sample covariances as regressors in moment test,

$$m^1_{ijt} = (e^2_{it} - \bar{e}^2_i)(e^2_{jt-k} - \bar{e}^2_j)$$

for i, j=1, ..., N. We will refer to this test as Lagged Covariance test (LC-test). The test is designed to detect time dependency of multivariate time series. TR^2 from the auxiliary regression of 1 on m^1_{ijt} will have a $\chi^2(N^2)$ distribution.

4). CLC-test

We use the sum of N^2 lagged sample covariances in LC-test as one moment condition:

$$m_t^1 = \sum_{ij} m_{ijt}^1$$

This test will be referred to as Composite Lagged Covariance test (CLC-test). T times the R^2 of the regression 1 on m_t^1 will have a $\chi^2(1)$ distribution.

5). AC-test

The last test we will introduce is an additive composite test to check 2) and 4) simultaneously. TR^2 of the regression 1 on m_t^0, m_t^1 will have a $\chi^2(2)$ distribution.

Table 4.1 gives the simulated 1%, 5%, 10% level critical values for the tests discussed above. 1000 replications were performed. For convenience of comparison, table 4.1 also gives the asymptotic critical values for these tests. It is seen that the small sample distribution is quite close to their asymptotic one. As the data set we will analyze in section 5 have 7420 observations, it will be quite reasonable to use the asymptotic value in doing the multivariate ARCH tests.

Table 4.1 Simulated and asymptotic critical values for moment tests

test	level	C	CC	LC	CLC	AC
small sample	1%	23.14	6.82	44.13	5.72	8.75
T=1000	5%	18.83	4.18	38.68	3.58	6.10
N=5	10%	16.46	2.95	35.64	2.56	4.49
asymptotic	1%	23.21	6.64	44.31	6.64	9.21
$T\to\infty$	5%	18.31	3.84	37.65	3.84	5.99
N=5	10%	15.99	2.71	34.38	2.71	4.61

5. An Empirical Example

Because of the difficulties encountered in estimating multivariate ARCH models, most of the empirical examples in the literature only deal with two or three variable problems. However, in practice it is often needed to estimate the covariance matrix for a much larger system. For example, BARRA, a financial consulting firm at Berkeley, built a 68×68 factor covariance matrix using EWMA method for their US equity model. However, the weight they used is *ad hoc* - simply decided by assuming the half life of the system to be 36 months. For a large system like this, only a small portion of the models discussed in section 2 are applicable. Most useful the Integrated MARCH model, the two parameter MARCH model, the Exponentially Weighted Moving Average model, the Constant Conditional Correlation model and the Principal Component model. All the other models are merely impossible to estimate. Although we have done estimation and comparison for large systems (17×17 and 68×68) for the five simple models mentioned above, we will give a 5×5 example in this paper in order to give illustrative guidance in estimating and selecting best models for multivariate systems.

The five series we will use are daily individual stock returns for five US companies, 1). Amoco, 2). Ford, 3). HP, 4). IBM, and 5). Merck. The data are drawn from CRSP data file. They all start from

July 3, 1962 and end at December 31, 1991 with 7420 observations. Table 5.1 gives the summary statistics of these five series. It can be seen that all these five companies have a positive mean returns over this 30 year period. They are all leptokurtic and the Jarque-Bera normality test statistic show that the returns are far from normal. The Ljung-Box test statistic for autocorrelation in returns, $Q_{12}(r)$, show that there are statistically significant autocorrelations in return processes. However, the Ljung-Box test statistics for autocorrelation in absolute returns, $Q_{12}(|r|)$, and in squared returns, $Q_{12}(r^2)$, show that there are substantially more autocorrelations in absolute returns and squared returns which is one would expect for stock market return series.

Table 5.1 Summary statistics of five stock returns

data	mean	std	skewness	kurtosis	normality test	$Q_{12}(r)$	$Q_{12}(\|r\|)$	$Q_{12}(r^2)$
AMOCO	.0007	.0149	-.180	11.75	23680	121.80	1633.4	367.3
FORD	.0005	.0164	.272	7.86	7400	28.35	1362.9	680.4
HP	.0008	.0208	-.026	6.74	4330	37.28	989.0	733.0
IBM	.0004	.0137	-.340	16.56	56985	22.70	1074.8	512.5
MERCK	.0008	.0143	.133	6.28	3350	99.33	722.8	950.1

In order to focus our attention on the higher moments, we did some preprocessing before we start to fit any multivariate ARCH models. The first and second order autocorrelations in the return are corrected and the mean is subtracted. The summary statistics for adjusted returns are shown in table 5.2. From the Ljung-Box test statistic for the adjusted returns it can be seen that there are much less autocorrelations left in the new transformed series. However, the autocorrelations in the absolute returns and squared returns are still very significant.

Table 5.2 Summary statistics of five adjusted returns

data	mean	std	skewness	kurtosis	normality test	$Q_{12}(r)$	$Q_{12}(\|r\|)$	$Q_{12}(r^2)$
AMOCO	.0000	.0148	-.150	11.63	23060	18.51	1814.8	532.3
FORD	.0002	.0164	.282	7.74	7032	6.48	1440.3	726.0
HP	-.0001	.0208	-.035	6.72	4283	16.46	982.6	764.0
IBM	.0000	.0137	-.340	16.56	56980	22.70	1074.8	512.5
MERCK	.0001	.0142	.114	6.25	3280	17.31	702.4	954.5

Table 5.3 shows the sample covariance matrix for the five adjusted returns, and table 5.4 shows the sample correlation matrix. It is seen that these five series are significantly positively correlated with each other. That is, the individual stock returns tend to move in the same direction with each other.

Table 5.3 Sample covariance ($\times$10000) matrix

	AMOCO	FORD	HP	IBM	MERCK
AMOCO	2.19				
FORD	.71	2.68			
HP	.91	1.24	4.32		
IBM	.73	1.01	1.33	1.89	
MERCK	.61	.79	1.03	.81	2.03

Table 5.4 Sample correlation matrix

	AMOCO	FORD	HP	IBM	MERCK
AMOCO	1				
FORD	.294	1			
HP	.295	.365	1		
IBM	.359	.450	.468	1	
MERCK	.289	.338	.347	.416	1

Table 5.5 shows the correlation matrix for squared sample covariance normalized residuals. If the conditional covariance matrix is constant over time and the normality is a good assumption for the data, then the off diagonal elements in Table 5.5 should be very close to zero and $T\sum\rho_{ij}^2$ should follow a $\chi^2(10)$ distribution. But in table 5.5 at least half of the correlations are too big to be ignored and $T\sum\rho_{ij}^2 = 1978$ which is significant almost in any statistical level.

Table 5.5 Sample correlation matrix for squared sample covariance normalized residuals

	AMOCO	FORD	HP	IBM	MERCK
AMOCO	1				
FORD	.246	1			
HP	.069	.097	1		
IBM	.359	.156	.132	1	
MERCK	.032	.047	.074	.112	1

For many multivariate ARCH models, estimation is usually computationally time consuming because a lot of matrix inverse operation must be performed when calculating the likelihood function in each iteration. For some of them, for example the BEKK model, it is difficult to estimate without proper initial values for the parameters. The program designed for this study is able to set the proper initial values automatically according to different data set [The estimation procedure used in this study is now available in S-Plus GARCH, see Martin, *et al*. (1996)].

Table 5.6 Number of parameters and log-likelihood functions

	# of parameters	log-likelihood function
Matrix-Diagonal	45	106789.9
Vector-Diagonal	25	106638.1
Two-parameter	17	106608.3
Variance Targeting	2	106589.6
Integrated	1	106342.8
EWMA	1	106060.5
CCC	25	106731.3
BEKK	45	106715.7
PRCOMP	15	106452.6

Table 5.6 gives the log likelihood functions for different estimated models and the parameters used in the final preferred parameter specifications for each different model. The convergence criteria used for the estimation is that the R^2 from the auxiliary regression of 1 on score vector be less than 0.0001. The whole estimation results will not be reported here because of the space limit. Most parameters in the final preferred model are significant. Usually the residuals from these ARCH filters are much closer to normal compared to the residuals normalized just by sample covariance. From table 5.6, it is seen that going from Two-parameter model to Vector-Diagonal model to Matrix-Diagonal model, i.e. giving more flexibility to the GARCH parameter does not improve the likelihood function a lot. This could possibly mean that a lot of parameters in **B** matrix are redundant. On the other hand, the likelihood value decreases a lot if we impose more restriction on the ARCH coefficient matrix **A** (results not shown here). So the ARCH parameters should be more flexible for each shock. Overall, the Matrix-Diagonal model gives the highest likelihood value with the maximum number of parameters. However, the constant correlation matrix does quite a good job with only 25 parameters. It should also be noted that the Two-parameter Variance Targeting model has a likelihood value of 106589.6 which is also fairly good and is potentially very useful in modeling large systems.

Table 5.7 shows the results of the five tests discussed in section 4 for each estimated model. Since the test statistic is T times the R^2 from the regression of 1 on the moment conditions and scores, the more parameters one used, the more punishment there will be at least if the optimizer has not fully reached the maximum. From table 5.7 it is seen that the models with less parameters, such as the Two-parameter Variance Targeting, Integrated MARCH and EWMA, perform better overall than other models in these moment tests even though their likelihood values from these models are smaller. The constant correlation model also performs very well and is one of the top candidates due to its simplicity in estimation. For this particular data set, the less desired models are the BEKK model, Vector-Diagonal model and the Principal Component model. On the other hand, compared to the theoretical critical values for these tests, no model can pass all these tests. As mentioned before, this may suggest that all these models are misspecified or that the conditional distribution is not multivariate normal.

Table 5.7 M-test of squared residuals orthogonality

Model	C	CC	LC	CLC	AC
Matrix-Diagonal	96.30	35.00	100.66	33.56	33.80
Vector-Diagonal	99.38	38.47	165.43	61.60	61.95
Two-parameter	100.26	24.12	152.71	54.08	54.22
Variance Targeting	69.63	8.71	126.26	29.19	29.65
Integrated MARCH	61.85	4.00	117.29	19.11	20.01
EWMA	44.63	3.44	96.88	12.75	14.43
CCC	75.41	28.41	80.62	27.13	27.14
BEKK	118.40	81.84	152.99	60.07	62.13
PRCOMP	78.64	19.76	138.29	44.52	46.90

6. Conclusion

This paper proposes several new Multivariate ARCH models that are simple and guarantee the positive definiteness of the estimated conditional covariance matrices. Two sets of diagnostic tests are designed for detecting various model misspecification. The new proposed models, together with the Constant Conditional Correlation model and the BEKK model, are estimated for asset returns of five US companies. The test results show that no single model can pass all the diagnostic tests proposed which suggests either the models are misspecified or the conditional multivariate normal distribution is not a good assumption. Nevertheless, our feeling is that the one and two parameter models and the constant conditional correlation models are potentially useful in modeling large scale conditional covariance matrix because of their simplicity in estimation and their fairly good performance.

References

Baba, Y., R. F. Engle, D. F. Kraft and K. F. Kroner (1991), "Multivariate simultaneous generalized ARCH, Unpublished manuscript, Department of Economics, University of California, San Diego.

Bollerslev, T. (1990), "Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *The Review of Economics and Statistic*s, Vol. 74, No. 3, 498-505.

Bollerslev, T., R. F. Engle and J. M. Wooldridge (1988), "A Capital Asset Pricing Model With Time-Varying Covariance," *Journal of Political Econom*y, Vol. 96, No. 1, 116-131.

Ding, Zhuanxin, C. W. J. Granger and R. F. Engle (1993), "A long memory property of stock market returns and a new model," *Journal of Empirical Financ*e, 1, 83-106.

Ding, Zhuanxin (1994), "Time series analysis of speculative returns," Ph.D. dissertation (University of California, San Diego, CA).

Engle, R. F. (1982),"Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometric*a, 50, 987-1008.

Engle, R. F. and K. F. Kroner (1995), "Multivariate simultaneous generalized ARCH," *Econometric Theor*y, 11, 122-150.

Engle, R.F. and Joe Mezrich (1996), "GARCH for Groups," Risk Magazine, pp36-40.

Fang, Kai-Tai, S. Kotz and Kai Wang Ng (1990), *Symmetric Multivariate and Related Distribution*s, New York, Chapman and Hall.

Johnson, N. L. and S. Kotz, (1972), *Distribution in Statistics: Continuous Multivariate Distribution*s, New York, John Wiley & Sons, INC.

Kahn, Ron (1992), Private conversation.

Martin, R. Douglas, Hong-Ye Gao, Yihui Zhan and Zhuanxin Ding (1996), *S-Plus GARCH User's Manual*.

Newey, W. K. (1985), "Maximum likelihood specification testing and conditional moment tests," *Econometrica*, Vol. 53, No. 5, 1047-1070.

Styan, G. P. H. (1973), "Hadamard products and multivariate statistical analysis," *Linear Algebra and Its Application*s, 6, 217-240.

Tauchen, G. (1985), "Diagnostic testing and evaluation of maximum likelihood models," *Journal of Econometric*s, 30, 415-443.