89-57R

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

MULTIVARIATE SIMULTANEOUS GENERALIZED ARCH

BY

ROBERT F. ENGLE

AND

KENNETH F. KRONER

DISCUSSION PAPER 89-57R
JULY 1993

Multivariate Simultaneous Generalized ARCH*

by

Robert F. Engle
Department of Economics, 0508
UC San Diego
La Jolla, CA 92093-0508

and

Kenneth F. Kroner
Department of Economics
McClelland Hall, 401
University of Arizona
Tucson, AZ 85721

June 1993

Abstract

This paper presents theoretical results in the formulation and estimation of multivariate generalized ARCH models within simultaneous equations systems. A new parameterization of the multivariate ARCH process is proposed and equivalence relations are discussed for the various ARCH parameterizations. Constraints sufficient to guarantee the positive definiteness of the conditional covariance matrices are developed, and necessary and sufficient conditions for covariance stationarity are presented. Identification and maximum likelihood estimation of the parameters in the simultaneous equations context are also covered.

* This paper began as a synthesis of at least three UCSD Ph.D. dissertations on various aspects of multivariate ARCH modelling, by Yoshi Baba, Dennis Kraft and Ken Kroner. In fact, an early version of this paper was written by Baba, Engle, Kraft and Kroner, which led to the acronym (BEKK) used in this paper for the new parameterization presented. In the interests of continuity, we maintain the acronym BEKK even though two of the authors have gone on to other pursuits. In addition to Yoshi Baba and Dennis Kraft, we would also like to thank Tim Bollerslev, Doc Ghose, Jan Magnus, Ron Oaxaca, Hal White, and two thorough referees for fruitful discussion and comments, but of course we must accept full responsibility for all errors ourselves. Ken Kroner would like to acknowledge financial support from the Karl Eller Center at the University of Arizona, and Robert Engle, from NSF SES 89-10273.

I. Introduction

Although economists have long been interested in the analysis of behavior under uncertainty, econometricians have only recently begun developing an analytical framework to deal with uncertainty. A central feature of this framework is the modelling of second and possibly higher moments as well. One of the most prominent tools used to model the second moments is due to Engle [8]. Engle [8] suggested that these unobservable second moments could be modelled by specifying a functional form for the conditional variance and modelling the first and second moments jointly, giving what is called in the literature the Autoregressive Conditional Heteroskedasticity (ARCH) model. Of course, many different functional forms are possible, but Engle's [8] suggestion that the conditional variances depend on elements in the information set in an autoregressive manner has become perhaps the most common. This linear ARCH model was generalized by Bollerslev [2] in a manner analagous to the extension from AR to ARMA models in traditional times series by allowing past conditional variances to appear in the current conditional variance equation. The resulting model is called Generalized ARCH, or GARCH. These models have been applied extensively in the literature. See, for example, the survey by Bollerslev, Chou and Kroner [3].

Further extensions to multivariate models, which are usually analagous to the extension from ARMA to vector ARMA models, appear often in the literature, though usually without theoretical discussion. See, for example, Bollerslev, Engle and Wooldridge [4], Engel and Rodrigues [7], Engle, Granger and Kraft [9], Kaminsky and Peruga [12], Kroner and Claessens [14], Kroner and Lastrapes [15], or McCurdy and Morgan [19], among several others. Multivariate ARCH models allow the variances and covariances to depend on the information set in a vector ARMA manner, and are particulary useful in multivariate financial models (such as the CAPM or dynamic hedging models) which require the modelling of both variances and covariances. But while most applications of multivariate ARCH have been to financial modelling, several potential applications also exist in macroeconomics and in other areas of economics. For example, it is often conjectured that employment decreases with price level uncertainty (Friedman, [11]). A hypothesis like this could be tested with the following two–equation model:

$$\begin{aligned} Y &= f(P, X, \sigma_p) \\ P &= g(Y, X) \end{aligned} \tag{1.1}$$

where Y is employment, P is the price level, X are exogenous variables, and σ_p is the ARCH-measure of price uncertainty. A significantly negative coefficient on σ_p would provide support for the hypothesis.

The purpose of this paper is to examine the theoretical properties of multivariate generalized ARCH models, and to apply these models to systems of simultaneous equations where the second moments of the random variables may be regressors. The paper is organized as follows: Section 2 presents the models, discusses the positive definiteness of the covariance matrix and examines covariance stationarity of the model; Section 3 analyzes the multivariate GARCH-in-mean model in a simultaneous equations framework; Section 4 discusses estimation of the model; and Section 5 gives some concluding remarks.

2. The Models

2.1. Univariate GARCH.

The parameterization of the conditional variance used by Engle [8] to model the unobservable second moments allows the conditional variance to depend on the elements of the information set in an autoregressive manner. Letting $\Im_{t-1}$ be the sigma field generated by the past values of ϵ_t, with σ_t^2 measurable with respect to $\Im_{t-1}$, the linear univariate ARCH model can be written as

$$\begin{aligned} \epsilon_t \mid \Im_{t-1} \ &\sim \ N(0, \sigma_t^2) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2. \end{aligned}$$

This model is called ARCH of order p, or ARCH(p).

Bollerslev [2] generalizes the ARCH process by allowing past conditional variances to appear in the current conditional variance equation. His variance equation becomes

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2.$$

This process is called generalized ARCH of order (p, q), or GARCH(p, q). The simple GARCH(1,1) model often provides a parsimonious description of the data; see, for example, Bollerslev [2] or McCurdy and Morgan [18].

2.2 Multivariate GARCH.

The extension from a univariate GARCH model to an n-variate model requires allowing the conditional variance-covariance matrix of the n-dimensional zero mean random variables ϵ_t to depend on elements of the information set. Letting H_t be measurable with respect to $\Im_{t-1}$, the multivariate GARCH model can be written as

$$\epsilon_t \,|\, \Im_{t-1} \;\sim\; N(0, H_t).$$

The parameterization for H_t as a function of the information set $\Im_{t-1}$ chosen here allows each element of H_t to depend on q lagged values of the squares and crossproducts of ϵ_t as well as p lagged values of the elements of H_t, and a $J \times 1$ vector of weakly exogenous variables (as defined by Engle, Hendry and Richard [10]), x_t. So the elements of the covariance matrix follow a vector ARMAX process in squares and cross products of the residuals. We will assume x_t contains only current and lagged exogenous variables. Defining

$$\begin{aligned} h_t &= \mathrm{vec} H_t \\ \tilde{x}_t &= \mathrm{vec}(x_t x_t') \\ \eta_t &= \mathrm{vec}(\epsilon_t \epsilon_t'), \end{aligned}$$

where vec($\cdot$) is the vector operator which stacks the columns of the matrix, a parameterization can be written

$$h_t = C_0 + C_1 \tilde{x}_t + A_1 \eta_{t-1} + \ldots + A_q \eta_{t-q} + G_1 h_{t-1} + \ldots + G_p h_{t-p}$$

where C_0 is a $n^2 \times 1$ parameter vector, C_1 is a $n^2 \times J^2$ parameter matrix, and A_i and G_i are $n^2 \times n^2$ parameter matrices. In matrix notation, this becomes

$$\begin{aligned} h_t &= [C_0 \vdots C_1 \vdots A_1 \vdots \ldots \vdots A_q \vdots G_1 \vdots \ldots \vdots G_p] \begin{bmatrix} 1 \\ \tilde{x}_t \\ \eta_{t-1} \\ \vdots \\ h_{t-p} \end{bmatrix} \\ &= F z_t \\ &= (z_t' \otimes I) \mathrm{vec} F \\ &= Z_t \alpha \end{aligned} \tag{2.1}$$

where

$$
\begin{aligned}
z_t' &= (1, \tilde{x}_t', \eta_{t-1}', \ldots, \eta_{t-q}', h_{t-1}', \ldots, h_{t-p}') \\
F &= [C_0 \vdots C_1 \vdots A_1 \vdots \ldots \vdots A_q \vdots G_1 \vdots \ldots \vdots G_p] \\
\alpha &= \mathrm{vec} F \\
\text{and} \qquad Z_t &= (z_t' \otimes I).
\end{aligned}
$$

Equations (2.1) define a parameterization which we will call the *vec representation.*

To illustrate, consider a simple 2-equation GARCH(1,1) *vec* model without exogenous influences. Model (2.1) becomes

$$
h_t = \begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{01} \\ c_{02} \\ c_{03} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix}.
$$

Notice that we have omitted the equation for $h_{21,t}$ and have given no coefficient to $\epsilon_{2,t-1}\epsilon_{1,t-1}$ or $h_{21,t-1}$ as these are clearly redundant, leaving nine free parameters in each of the A_1 and G_1 matrices. Similar redundancies appear in the general n-variate GARCH(1,1) *vec* model. In particular, all the covariance equations appear twice – i.e. there is an equation for $h_{ij,t}$ as well as for $h_{ji,t}$ – and all the off-diagonal terms appear twice within each equation – i.e. both of the terms $\epsilon_{i,t-1}\epsilon_{j,t-1}$ and $\epsilon_{j,t-1}\epsilon_{i,t-1}$ and both of the terms $h_{ij,t-1}$ and $h_{ji,t-1}$ appear in each equation. The redundant terms can be eliminated without affecting the model, leaving a total of $\left(\frac{n(n+1)}{2}\right)^2$ unique parameters in each of the A_i and G_i matrices. In a direct formulation of (2.1) there appear to be n^4 parameters in each matrix, but many of these are superfluous.

For empirical implementation, it is desirable to further restrict this parameterization. A natural restriction which was first used in the ARCH context by Engle, Granger and Kraft [9] and in the GARCH context by Bollerslev, Engle and Wooldridge [4] is the *diagonal representation*, in which each element of the covariance matrix, $h_{jk,t}$, depends only on past values of itself and past values of $\epsilon_{j,t}\epsilon_{k,t}$. That is, variances depend solely on past own squared residuals, and covariances depend solely on past own cross products of residuals. This seems an intuitively plausible restriction since information about variances is usually revealed in squared residuals and if the variances are evolving slowly, then past squared residuals should be able to forecast future variances. A similar argument can be made for covariances. In the *vec* model, a *diagonal* representation is obtained if the matrices A_i and G_i are assumed to be diagonal.

To illustrate in the bivariate case, the *diagonal* model is simply

$$h_t = \begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{01} \\ c_{02} \\ c_{03} \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} g_{11} & 0 & 0 \\ 0 & g_{22} & 0 \\ 0 & 0 & g_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix}$$

or

$$\begin{aligned} h_{11,t} &= c_{01} + a_{11}\epsilon_{1,t-1}^2 + g_{11}h_{11,t-1} \\ h_{12,t} &= c_{02} + a_{22}\epsilon_{1,t-1}\epsilon_{2,t-1} + g_{22}h_{12,t-1} \\ h_{22,t} &= c_{03} + a_{33}\epsilon_{2,t-1}^2 + g_{22}h_{22,t-1}. \end{aligned}$$

In the bivariate model illustrated here, there are three free parameters in each of the A_1 and G_1 matrices, and in the general n-variate *diagonal* model there are $\left(\frac{n(n+1)}{2}\right)$ free parameters in each matrix.

In order for any parameterization to be sensible, we require that H_t be positive definite for all values of ϵ_t and x_t in the sample space. In the *vec* representation, and even in the *diagonal* representation, this restriction can be difficult to check, let alone impose during estimation. We now propose a new parameterization which easily imposes these restrictions and which eliminates very few if any interesting models allowed by the *vec* representation.

Consider the following model:

$$H_t = C_0^{*\prime}C_0^* + \sum_{k=1}^{K} C_{1k}^{*\prime}x_t x_t' C_{1k}^* + \sum_{k=1}^{K}\sum_{i=1}^{q} A_{ik}^{*\prime}\epsilon_{t-i}\epsilon_{t-i}' A_{ik}^* + \sum_{k=1}^{K}\sum_{i=1}^{p} G_{ik}^{*\prime}H_{t-i}G_{ik}^* \tag{2.2}$$

where C_0^*, A_{ik}^* and G_{ik}^* are $n \times n$ parameter matrices with C_0^* triangular, C_{1k}^* are $J \times n$ parameter matrices, and the summation limit K determines the generality of the process. It should be clear that (2.2) will be positive definite under very weak conditions. Furthermore, this representation is sufficiently general that it includes all positive definite *diagonal* representations and nearly all positive definite *vec* representations. It will be shown to be a particularly convenient representation for estimation and for analysis of simultaneous equations systems. Throughout the paper we will refer to this representation as the *BEKK representation*[1].

To illustrate the *BEKK* model, consider first the the simple GARCH(1,1) model, with $K = 1$ and no exogenous influences[2]:

$$H_t = C_0^{*\prime}C_0^* + A_{11}^{*\prime}\epsilon_{t-1}\epsilon_{t-1}'A_{11}^* + G_{11}^{*\prime}H_{t-1}G_{11}^*. \tag{2.3}$$

In the bivariate case which is illustrated for both the *vec* and *diagonal* representations above, the *BEKK* model becomes

$$H_t = C_0^{*\prime}C_0^* + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix}' \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}\epsilon_{1,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} + \begin{bmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{bmatrix}' H_{t-1} \begin{bmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{bmatrix}.$$

or, suppressing the time subscripts and the GARCH terms,

$$\begin{aligned} h_{11} &= c_{11} + a_{11}^{*2}\epsilon_1^2 + 2a_{11}^*a_{21}^*\epsilon_1\epsilon_2 + a_{21}^{*2}\epsilon_2^2 \\ h_{12} &= c_{12} + a_{11}^*a_{12}^*\epsilon_1^2 + (a_{21}^*a_{12}^* + a_{11}^*a_{22}^*)\epsilon_1\epsilon_2 + a_{21}^*a_{22}^*\epsilon_2^2 \\ h_{22} &= c_{13} + a_{12}^{*2}\epsilon_1^2 + 2a_{12}^*a_{22}^*\epsilon_1\epsilon_2 + a_{22}^{*2}\epsilon_2^2. \end{aligned}$$

Comparing this model to the *vec* form of the model, we see that this model economizes on parameters by imposing restrictions both across and within equations. In fact, for $n = 2$ we see that this representation uses only eight parameters, compared to the 18 from the *vec* model (excluding constants).

Before formalizing the relationship between the *BEKK* and *vec* models, we first discuss the identification of the parameters in model (2.3). Proposition 2.1 shows that under simple and straightforward conditions, the parameters in this model, i.e. the parameters in the *BEKK* model with $K = 1$, are identified. Defining two representations to be equivalent if every sequence $\{\epsilon_t\}$ generates the same sequence $\{H_t\}$ for both representations, we have:

PROPOSITION 2.1. *Suppose that the diagonal elements in C_0^* are restricted to be positive and that a_{11}^* and g_{11}^* are also restricted to be positive. Then if $K = 1$ there exists no other C_0^*, A_1^* or G_1^* in model (2.3) which will give an equivalent representation.*

PROOF: All proofs are given in the Appendix.

In practice, nonnegativity restrictions on parameters are easy to impose, for example by estimating the square root of the restricted parameter, making identification of the parameters in (2.3) relatively easy for estimation. Also, it should be clear from the proof of Proposition 2.1 that the purpose of the restrictions is to eliminate all other observationally equivalent structures, and that there are several other sets of sufficient conditions which could be used in place of those given. For example, as relates to the term $A_1^{*\prime}\epsilon_{t-1}\epsilon_{t-1}^*A_1^*$, the only other observationally equivalent structure is obtained by replacing A_1^* with $-A_1^*$. The restriction that a_{11}^* be positive could be replaced with the condition that a^*_{ij} be positive for a given i and j, as this condition is also sufficient to eliminate $-A^*_1$ from the set of admissable structures. A final comment is that we decompose the constant matrix into $C^{*\prime}_0 C^*_0$ only to ensure positive definiteness; in practice the elements of C^*_0 are not of interest. We chose this decomposition because of its simplicity, but any other identifiable factorization of the constant matrix could be used.

The preceeding discussion and illustrations all deal with the *BEKK* model with $K = 1$. Clearly, however, setting $K = 1$ involves imposing restrictions on the model which might not be desirable in practice. The full generality of the *BEKK* representation can be recovered by simply adding more positive semidefinite terms to the variance equation, i.e. by letting $K > 1$ in (2.2), giving in the GARCH(1,1) case

$$H_t = C^{*\prime}_0 C^*_0 + \sum_{k=1}^{K} A^{*\prime}_{1k} \epsilon_{t-1} \epsilon^{\prime}_{t-1} A^*_{1k} + \sum_{k=1}^{K} G^{*\prime}_{1k} H_{t-1} G^*_{1k}. \tag{2.4}$$

The question of how large K must be for the *BEKK* representation to be "fully general" — i.e. to be equivalent to as many vec representations as possible — is answered in part by Proposition 2.2, which gives conditions on the A^*_{1k} and G^*_{1k} matrices which must be satisfied in order to eliminate all the unnecessary restrictions.

PROPOSITION 2.2. *In order for the BEKK model to achieve full generality, the following two necessary conditions must hold:*

a) Define $s = \frac{n(n+1)}{2}$. Then K must be large enough that there are a total of at least s^2 distinct parameters in the A^*_{1k} matrices.

b) Define $a^*_{ij,k}$ to be the ij^{th} element of A^*_{1k}. Then there must exist an A^*_{1k} matrix which contains either the pair of nonzero elements $(a^*_{il,k}, a^*_{jm,k})$, or the pair of nonzero elements $(a^*_{jl,k}, a^*_{im,k})$, for all i, j, l, m between 1 and n.

Similar restrictions hold for the G^*_{1k} matrices.

The first condition simply says that if there are fewer parameters in the *BEKK* model than in the *vec* model, then the *BEKK* model is implicitly imposing some unnecessary restrictions. The second condition says that certain pairs of parameters must appear together in an A^*_{1k} matrix for some k in order not to impose extra implicit restrictions. To illustrate, consider the case for $n = 2$.

The following set of A_{1k}^* matrices satisfies the conditions in Proposition 2.2:

$$A_{11}^* = \begin{bmatrix} a_{11,1} & a_{12,1} \\ 0 & a_{22,1} \end{bmatrix} \qquad A_{12}^* = \begin{bmatrix} a_{11,2} & a_{12,2} \\ a_{21,2} & 0 \end{bmatrix} \qquad A_{13}^* = \begin{bmatrix} a_{11,3} & 0 \\ a_{21,3} & a_{22,3} \end{bmatrix}$$

But the following matrices violate condition (b) of Proposition 2.2 and therefore cannot give a fully general parameterization:

$$A_{11}^* = \begin{bmatrix} a_{11,1} & a_{12,1} \\ 0 & a_{22,1} \end{bmatrix} \quad A_{12}^* = \begin{bmatrix} a_{11,2} & a_{12,2} \\ 0 & 0 \end{bmatrix} \quad A_{13}^* = \begin{bmatrix} a_{11,3} & 0 \\ a_{21,3} & 0 \end{bmatrix} \quad A_{14}^* = \begin{bmatrix} 0 & a_{12,4} \\ 0 & a_{22,4} \end{bmatrix}$$

because the pair a_{21} and a_{22} never appears together in any of the A_{1k}^* matrices. This restriction translates into the restriction that the term $\epsilon_{2,t-1}^2$ does not appear in the covariance equation.

Of course, Proposition 2.2 gives only necessary, and not sufficient, conditions for the full generality of the *BEKK* model. Many different sets of sufficient conditions are possible. For example, to look at an extreme case, one set of sufficient conditions is that $K = s$ and none of the A_{1k}^* matrices have any restrictions on its elements. However, this results in identification problems because there are now several equivalent models in the *BEKK* framework. For example, interchanging A_{11}^* and A_{12}^* will give an observationally equivalent structure. In general, an identification problem like this arises in the *BEKK* model whenever $K > 1$, and therefore restrictions must be imposed on the A_{1k}^* and G_{1k}^* matrices to eliminate other equivalent representations. Many different sets of restrictions could be used, but Proposition 2.3 gives a particularly convenient one because the model presented therein is also fully general[3].

PROPOSITION 2.3. *Suppose the diagonal elements of C_0^* are restricted to be positive. Consider the class of BEKK models in which $A_{1k_r}^*$, where $k_r = n(r-1)+1, \ldots, nr$ and $r = 1, \ldots, n$, is the matrix obtained by setting the first $r-1$ columns and the first $k_r - n(r-1) - 1$ rows to zero. Suppose also that $a_{nn,k_r}^* > 0 \quad \forall k_r$, and that similar restrictions are placed on the $G_{1k_r}^*$ matrices. Then a fully general BEKK model is obtained which has no other equivalent representations in this class.*

Notice that in this representation, $K = n^2$. To illustrate, if $n = 2$ then the following set of A_{1k}^* matrices will give a fully general *BEKK* model with no equivalent representations:

$$A_{11}^* = \begin{bmatrix} a_{11,1}^* & a_{12,1}^* \\ a_{21,1}^* & a_{22,1}^* \end{bmatrix} \quad A_{12}^* = \begin{bmatrix} 0 & 0 \\ a_{21,2}^* & a_{22,2}^* \end{bmatrix} \quad A_{13}^* = \begin{bmatrix} 0 & a_{12,3}^* \\ 0 & a_{22,3}^* \end{bmatrix} \quad A_{14}^* = \begin{bmatrix} 0 & 0 \\ 0 & a_{22,4}^* \end{bmatrix}.$$

One obvious corollary to this proposition is that for any *BEKK* model with $K = 2$, a sufficient condition to identify the model is that for some i, the i^{th} row of A_{12}^* contains only zeros, while the

i^{th} row of A_{11}^* contains nonzero elements. Of course, a similar restriction would have to be imposed on G_{12}^*, and the positivity restrictions in Proposition 2.3 would also have to be imposed.

We now turn to a formalization of the relationship between the *BEKK* and *vec* parameterizations. The mathematical relationship between the parameters of these two models can be found by vectorizing both sides of equation (2.4), recognizing that $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$:

$$h_t = (C_0^* \otimes C_0^*)'\text{vec}(I_n) + \sum_{k=1}^{K}(A_{1k}^* \otimes A_{1k}^*)'\text{vec}(\epsilon_{t-1}\epsilon_{t-1}') + \sum_{k=1}^{K}(G_{1k}^* \otimes G_{1k}^*)'\text{vec}(H_{t-1}).$$

Therefore,

$$A_1 = \sum_{k=1}^{K}(A_{1k}^* \otimes A_{1k}^*)' \qquad \text{and} \qquad G_1 = \sum_{k=1}^{K}(G_{1k}^* \otimes G_{1k}^*)',$$

which leads to the following proposition regarding the equivalence of the two models:

Proposition 2.4. *The vec and BEKK parameterizations are equivalent if and only if there exist C_0^*, A_{ik}^* and G_{ik}^* such that*

$$\begin{aligned} C_0 &= (C_0^* \otimes C_0^*)'\mathit{vec}(I_n) \\ A_i &= \sum_{k=1}^{K}(A_{ik}^* \otimes A_{ik}^*)' \\ G_i &= \sum_{k=1}^{K}(G_{ik}^* \otimes G_{ik}^*)'. \end{aligned} \tag{2.5}$$

One implication of Proposition 2.4 is that the *vec* model implied by any given *BEKK* model is unique, while the converse is not true. The transformation from a *vec* model to a *BEKK* model (when it exists) is not unique because for a given A_1 the choice of A_{1k}^* is not unique. This can be seen by recognizing that $(A_{1k}^* \otimes A_{1k}^*) = (-A_{1k}^* \otimes -A_{1k}^*)$, so while $A_1 = \sum_{k=1}^{K}(A_{ik}^* \otimes A_{ik}^*)'$ is unique, the choice of A_{ik}^* is not unique.

A second implication of Proposition 2.4 concerns the relationship between the *BEKK* model and the *diagonal* model. In particular, relations (2.5) make it clear that a *diagonal* model is returned from the *BEKK* parameterization if and only if each of the A_{ik}^* and G_{ik}^* matrices are diagonal. It will be shown later that further restrictions can be placed on the diagonal elements of A_{ik}^* and G_{ik}^* in order to obtain a *diagonal* model, but we will postpone that discussion until Proposition 2.6.

A third implication of Proposition 2.4 is the characterization of which *vec* models have *BEKK* representations and which do not. More specifically, we see that the *vec* models excluded from the fully general *BEKK* parameterization are those for which no C_0^*, A_{ik^*} and G_{ik}^* exist which satisfy relations (2.5). Proposition 2.5 demonstrates that this includes all non-positive definite *vec* parameterizations. In fact, this is arguably the key feature of the *BEKK* parameterization: positive definite covariance matrices are generated by essentially unrestricted parameterizations. More precisely,

PROPOSITION 2.5. *If $H_0, H_{-1}, \ldots, H_{-p+1}$ are all positive definite, then the parameterization of the GARCH equations given in (2.2) yields a positive definite H_t for all possible values of ϵ_t if the null space of C_0^* and the null spaces of G_{ik}^*, $i = 1, \ldots, p$ and $k = 1, \ldots, K$ all intersect only at the origin.*

Notice that a sufficient condition for this null space criterion to hold is that at least one of the C_0^* or G_{ik}^* be of full rank.

It can also be shown that the *BEKK* model eliminates few, if any, of the interesting positive definite models permitted by the *vec* model. In particular, all positive definite *diagonal vec* models can be written in the *BEKK* framework, so that if one restricts the focus to diagonal models, the *BEKK* model is equally as general as the *vec* model:

PROPOSITION 2.6. *In the vec model, suppose that the constant part of the covariance matrix is positive definite, so that $C_0 = vec(\Omega)$ where Ω is positive definite. Suppose also that A_i and G_i are diagonal. Then if H_t is positive definite for all possible realizations of ϵ_t, there exists a triangular matrix C_0^* and diagonal matrices A_{ik}^* and G_{ik}^*, $k = 1, \ldots, n$, such that*

$$
\begin{aligned}
C_0 &= vec(C_0^{*\prime} C_0^*) \\
A_i &= \sum_{k=1}^{n} (A_{ik}^* \otimes A_{ik}^*)' \\
G_i &= \sum_{k=1}^{n} (G_{ik}^* \otimes G_{ik}^*)'.
\end{aligned}
$$

This proposition says that the *BEKK* model includes as special cases all possible positive definite linear *diagonal* models, and is in this sense "general". The proof of Proposition 2.6 makes it clear that further restrictions can be placed on the A_{ik}^* and G_{ik}^* matrices, beyond just diagonality, without affecting the generality of the *diagonal BEKK* model. In particular, we can restrict each of the A^*_{ik} and G^*_{ik} matrices to be diagonal with the first $n-k$ elements on the diagonals set to zero. By Proposition 2.3, this restricted model will have no other equivalent representations in its class, meaning that we have a fully general linear diagonal model which is both identified and positive definite, making estimation relatively simple.

2.3. Covariance Stationarity.

Finally, we turn to a discussion of the necessary and sufficient conditions for covariance stationarity of the multivariate GARCH process, as given in the following proposition:

PROPOSITION 2.7. *Suppose the process $\{\epsilon_t\}$ begins arbitrarily far in the past, at time $t-\tau$. Suppose also that $\lim_{\tau\to\infty} H_{t-\tau}$ is positive definite and finite. If the vec model (2.1) is positive definite, then it is covariance stationary if and only if all the eigenvalues of*

$$\sum_{i=1}^{q} A_i + \sum_{i=1}^{p} G_i$$

are less than one in modulus. Also, the BEKK model (2.2) is covariance stationary if and only if all the eigenvalues of

$$\sum_{i=1}^{q}\sum_{k=1}^{K}(A^*_{ik}\otimes A^*_{ik}) + \sum_{i=1}^{p}\sum_{k=1}^{K}(G^*_{ik}\otimes G^*_{ik})$$

are less than one in modulus.

Focussing on the GARCH(1,1) case to keep the notation simple, we see from the proof to Proposition 2.7 that in the *vec* model, the unconditional covariance matrix, when it exists, is given by

$$\mathrm{E}(\eta_t) = [I - A_1 - G_1]^{-1}\mathrm{vec}C_0, \tag{2.6a}$$

and in the *BEKK* model with $K=1$, it is

$$\mathrm{E}(\eta_t) = [I - (A^*_{11}\otimes A^*_{11})' - (G^*_{11}\otimes G^*_{11})']^{-1}\mathrm{vec}C^{*\prime}_0 C^*_0. \tag{2.6b}$$

Several other implications are also apparent. For example, the *diagonal vec* model is stationary if and only if the sums $a_{ii}+g_{ii}$ are less than one for all i and the *diagonal BEKK* model is stationary if and only if $\sum_{k=1}^{n}(a^{*2}_{ii,k}+g^{*2}_{ii,k}) < 1$ for all i. However, it is important to recognize that it is only in the case of *diagonal* models that the stationarity properties are determined solely by the diagonal elements of the A^*_{ik} and G^*_{ik} (or A_i and G_i) matrices. In non-*diagonal BEKK* models, for example, it is possible to have diagonal elements exceeding one, yet the process be stationary. For example, in the GARCH(0,1) model with $K = 1$ and

$$A^*_{11} = \begin{bmatrix} 1.1 & 0.2 \\ -0.2 & 0.7 \end{bmatrix},$$

all four eigenvalues of $(A^*_{11} \otimes A^*_{11})$ are 0.81. By Proposition 2.7, this process is stationary even though the diagonal elements of A^*_{11} are not both less than one.

3. The Regression Model

Applying the *BEKK* representation (2.2) to a simultaneous regression model with the second moments appearing in the structural mean equations gives the multivariate GARCH-in-mean model[4]

$$\begin{aligned} \epsilon_t &= \Gamma y_t + B x_t + \Lambda \tilde{h}_t \\ \epsilon_t \mid \Im_{t-1} &\sim N(0, H_t) \\ H_t &= C_0^{*\prime} C_0^* + \sum_{k=1}^{K} A_{1k}^{*\prime} \epsilon_{t-1} \epsilon_{t-1}^{\prime} A_{1k}^* + \sum_{k=1}^{K} G_{1k}^{*\prime} H_{t-1} G_{1k}^* \end{aligned} \tag{3.1}$$

where $\Gamma_{n \times n}$, $B_{n \times J}$ and $\Lambda_{n \times s}$ are parameter matrices, y_t $(n \times 1)$ are endogenous variables, x_t $(J \times 1)$ are weakly exogenous and lagged dependent variables, and the second moments in the mean equation, $\tilde{h}_t = \text{vec}h(H_t)$, are predetermined but not weakly exogenous variables[5]. These moments, $\tilde{h}_t$, are not weakly exogenous for Γ, B and Λ because the information matrix is not block-diagonal between these parameters and the parameters in the marginal distribution of $\tilde{h}_t$. This can be seen by examining the simplest of cases — the GARCH(0,0) model with no intercepts in the mean equations. If $H_t = C_0^{*'} C_0^* \equiv \Delta$ then

$$\mathcal{I}_{\Delta\lambda} = (\text{vec}\Delta \otimes \Gamma'^{-1})\Delta^{-1}\Gamma^{-1}\Lambda(I \otimes \iota) \neq 0$$

where $\mathcal{I}_{\Delta\lambda}$ is the block of the information matrix which corresponds to the interactions between vec(Δ) and λ, ι is a $n^2 \times 1$ vector of ones and $\lambda \equiv \text{vec}(\Lambda)$. Alternatively, notice that in the general GARCH$(1,1)$ model, $\tilde{h}_t$ is a function of ϵ_{t-1}, and ϵ_{t-1} are functions of the same parameters as $\tilde{h}_t$. So the parameters in the mean equations cannot be estimated without estimating the variance equations. Hence there is no sequential cut (Engle, Hendry and Richard, [10]).

The reduced form of (3.1) will also have a multivariate GARCH representation because non-singular linear combinations of multivariate GARCH models are GARCH.

Proposition 3.1. *If ϵ_t is a multivariate GARCH process and P is a non-singular matrix, then $P\epsilon_t$ is also a multivariate GARCH process of the same order.*

One implication of this proposition is that if GARCH is placed on the structural errors as in (3.1) then letting $P = \Gamma^{-1}$, we see that the reduced form will also have GARCH errors with the same orders. In particular, we have

$$\begin{aligned} y_t &= -\Gamma^{-1}Bx_t - \Gamma^{-1}\Lambda\tilde{h}_t + \Gamma^{-1}\epsilon_t \\ &= \Pi_1 x_t + \Pi_2\tilde{h}_t + \nu_t \\ \nu_t \,|\, \Im_{t-1} &\sim N(0, \Gamma^{-1}H_t\Gamma^{-1\prime}). \end{aligned} \tag{3.2}$$

Furthermore, if the conditions for positive definiteness were satisfied in the structure they will be in the reduced form. Several other implications of this proposition become readily apparent from the following simple relations between the structural and reduced form parameters, which are developed in the proof to Proposition 3.1: For the *BEKK* parameterization, the relationships are

$$\begin{aligned} C^*_{rf,0}\Gamma' &= C^*_{s,0} \\ \Gamma A^*_{rf,ik} &= A^*_{s,ik}\Gamma \\ \Gamma G^*_{rf,ik} &= G^*_{s,ik}\Gamma \end{aligned}$$

and for the *vec* parameterization they are

$$\begin{aligned} (\Gamma \otimes \Gamma)C_{rf,0} &= C_{s,0} \\ (\Gamma \otimes \Gamma)A_{rf,i} &= A_{s,i}(\Gamma \otimes \Gamma) \\ (\Gamma \otimes \Gamma)G_{rf,i} &= G_{s,i}(\Gamma \otimes \Gamma) \end{aligned}$$

where the rf subscript refers to reduced form parameters and the s subscript refers to structural parameters. For example, one implication is that if the GARCH error process is placed on the reduced form model, then the structural model must have GARCH errors. Also, if any part of the structural (reduced form) covariance matrix follows a GARCH process then in general the complete reduced form (structural) covariance matrix will follow a GARCH process.

Identification of the structural coefficients in (3.1) using $\tilde{h}_t$ cannot proceed as in the standard simultaneous equations model both because linear combinations of the structural equations in the system will change the definition of $\tilde{h}_t$, and because h_t is unobservable and consistent estimates of h_t can only be obtained if the model is identified. But rewriting $\Lambda\tilde{h}_t$ in terms of the reduced form covariance matrix permits standard identification procedures[6]. To do this, we first note that because the parameters in the conditional covariance matrices are identified (see Section 2.2 above), we can get consistent estimates of the reduced form covariances $h_t^* = \text{vech}(\Gamma^{-1}H_{t-1}\Gamma^{-1\prime})$. Using

$$\begin{aligned}\tilde{h}_t = \text{vech}(H_t) = R\text{vec}(H_t) &= R(\Gamma\otimes\Gamma)\text{vec}(\Gamma^{-1}H_t\Gamma^{-1\prime}) \\ &= R(\Gamma\otimes\Gamma)Sh_t^*,\end{aligned}$$

where $R_{s\times n^2}$ is the selection matrix such that $\text{vech}(P) = R\text{vec}(P)$ and $S_{n^2\times s}$ is the expansion matrix such that $\text{vec}(P) = S\text{vech}(P)$ for any square matrix $P_{n\times n}$, gives

$$\Lambda\tilde{h}_t = \Lambda R(\Gamma\otimes\Gamma)Sh_t^* = \Lambda^* h_t^*. \tag{3.3}$$

This gives the new system

$$\Gamma y_t + Bx_t + \Lambda^* h_t^* = \epsilon_t \tag{3.4}$$

where h_t^*, though unobservable, is consistently estimable and is invariant with respect to linear combinations of the structural equations. Identification of the new system (3.4) now reduces to the standard problem of identification of a linear system (with variables y_t, x_t, h_t^*) which is nonlinear in the parameters. The nonlinearities arise because $\Lambda^* = \Lambda R(\Gamma\otimes\Gamma)S$ is a nonlinear function of Λ and Γ. So necessary and sufficient conditions for identification in this model must be given by conditions involving (Γ, B, Λ^*), not (Γ, B, Λ) as might at first seem natural. In other words, restrictions on Λ must first be transformed into restrictions involving Λ^* before they can be used in identification.

To illustrate the problems which can arise if $\Lambda\tilde{h}_t$ is used in identification instead of $\Lambda^* h^*$, consider the following bivariate system of equations (ignoring time subscripts):

$$\begin{aligned} y_1 &= \beta_1 y_2 + \lambda_{11}h_{11} + \lambda_{13}h_{22} + \epsilon_1 \\ y_2 &= \lambda_{21}h_{11} + \lambda_{22}h_{12} + \lambda_{23}h_{22} + \epsilon_2. \end{aligned} \tag{3.5}$$

Identification of this system means that premultiplying by any non-diagonal matrix does not give an observationally equivalent structure. However, premultiplying this system by the matrix

$$\Upsilon = \begin{bmatrix} 1 & v \\ 0 & 1 \end{bmatrix}$$

gives the new system

$$\begin{aligned} y_1 &= (\beta_1 - v)y_2 + (\lambda_{11} + v\lambda_{21})h_{11} + (\lambda_{13} + v\lambda_{23})h_{22} + v\lambda_{22}h_{12} + \epsilon_1 + v\epsilon_2 \\ y_2 &= \lambda_{21}h_{11} + \lambda_{22}h_{12} + \lambda_{23}h_{22} + \epsilon_2. \end{aligned} \tag{3.6}$$

When estimating system (3.6), the GARCH-in-mean terms will be measuring the conditional variances and covariances of the residuals in the transformed system (3.6), not those in the original system (3.5). The residual in the first transformed equation is $\epsilon_1 + v\epsilon_2$, which has conditional variance $h^{\diamond}_{11} = h_{11} + 2vh_{12} + v^2 h_{22}$, and the residual in the second transformed equation is still ϵ_2, which has conditional variance $h^{\diamond}_{22} = h_{22}$. The conditional covariance between the transformed residuals is $h^{\diamond}_{12} = h_{12} + vh_{22}$. Solving these three relationships for h_{11}, h_{12} and h_{22} then substituting into (3.6) gives

$$\begin{aligned} y_1 &= (\beta_1 - v)y_2 + (\lambda_{11} + v\lambda_{21})h^{\diamond}_{11} + v(\lambda_{22} - 2\lambda_{11} - 2v\lambda_{21})h^{\diamond}_{12} \\ &\qquad + [\lambda_{13} + v\lambda_{23} + v^2(\lambda_{11} - \lambda_{22}) + v^3\lambda_{21}]h^{\diamond}_{22} + \epsilon_1 + v\epsilon_2 \\ y_2 &= \lambda_{21}h^{\diamond}_{11} + (\lambda_{22} - 2v\lambda_{21})h^{\diamond}_{12} + (\lambda_{23} - v\lambda_{22} + v^2\lambda_{21})h^{\diamond}_{22} + \epsilon_2. \end{aligned}$$

Setting

$$v = \frac{\lambda_{22} - 2\lambda_{11}}{2\lambda_{21}}$$

gives an observationally equivalent structure to (3.5), indicating that the first equation in (3.5) is not identified. However, using traditional rank and order conditions on (3.5) would erroneously lead us to conclude that system (3.5) is identified. This happens because the definition of $\tilde{h}_t$ changes in different structures, so $\tilde{h}_t$ is no help in ruling out equivalent structures. But reduced form variances and covariances are invariant to different structures, so *a priori* restrictions on the reduced form variances and covariances in the mean equations can be used for identification purposes. Using (3.3) to rewrite (3.5) in terms of reduced form variances and covariances gives the new system

$$\begin{aligned} y_1 &= \beta_1 y_2 + \lambda_{11}h^*_{11} - 2\beta_1\lambda_{11}h^*_{12} + (\lambda_{13} + \beta_1^2\lambda_{11})h^*_{22} + \epsilon_1 \\ y_2 &= \lambda_{21}h^*_{11} + (\lambda_{22} - 2\beta_1\lambda_{21})h^*_{12} + (\lambda_{23} - \beta_1\lambda_{22} + \beta_1^2\lambda_{21})h^*_{22} + \epsilon_2, \end{aligned} \tag{3.7}$$

where h^*_{ij} are the reduced form variances and covariances. In this system, it is straightforward to show that the same linear combination of equations as above results in an observationally equivalent structure to (3.7), meaning the system (3.7) is not identified. However, it is clear that restrictions on the reduced form variances and covariances in system (3.7) can identify the system. For example, if it is known that the coefficient on h^*_{12} is zero in (3.7), then the system is identified. To conclude, rank and order conditions should not be used with *a priori* restrictions on Λ to identify the structural parameters. Instead, the model should be expressed in terms of h^*_t and Λ^* before identification is attempted[7]. Also, standard identification methods are appropriate if the *a priori* restrictions are on Λ^*.

4. System Estimation

If Λ equals zero in model (3.1) and the system is identified, the equations can be consistently estimated with two-stage or three-stage least squares, ignoring the GARCH error structure, because the reduced form estimates of y_t remain uncorrelated with ϵ_t. However, more efficient estimates can be obtained by accounting for the error structure, for example, by using full information maximum likelihood or an instrumental variables estimator. On the other hand, if $\Lambda \neq 0$ then full information methods are generally required to obtain efficient or often even consistent estimates of the models' parameters. Before discussing these methods further, we will first present the likelihood function.

Although the error vectors ϵ_t are conditionally multivariate-normally distributed, their outer products are not independent, and hence the joint distribution of $(\epsilon_1, \epsilon_2, \ldots, \epsilon_T)$, where T is the number of observations, need not be multivariate-normally distributed. But the joint density is the product of all the conditional densities, so the log likelihood function of the joint distribution is the sum of all the log likelihood functions of the conditional distributions, i.e. the sum of the logs of the multivariate-normal distribution. So if we are estimating the reduced form model (3.2), then letting L_t be the log likelihood of observation t, and L be the joint log likelihood, gives

$$\begin{aligned} L &= \sum_{t=1}^{T} L_t \\ L_t &= \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|H_{rf,t}| - \frac{1}{2}\nu_t' H_{rf,t}^{-1}\nu_t. \end{aligned} \tag{4.1}$$

Proposition 3.1 says the the structural model will also follow a multivariate GARCH process, so applying the transformation $\epsilon_t = \Gamma\nu_t$ gives

$$L_t = \frac{n}{2}\ln(2\pi) + \ln|\Gamma| - \frac{1}{2}\ln|H_t| - \frac{1}{2}\epsilon_t' H_t^{-1}\epsilon_t \tag{4.2}$$

Of course, these likelihood functions are not fully specified until we make assumptions on initial conditions. A reasonable set of assumptions is that all presample data be fixed at its unconditional expectation. So, for example, $\epsilon_0\epsilon_0'$ is assumed to equal its unconditional expectation, given in equation (2.6).

Note that because no reference is made in (4.1) or (4.2) to the functional form chosen for the conditional covariance matrix, the results of this section apply whether the *vec* or *BEKK* parameterization is chosen. In either case, however, the models are large and complex, leading one to question how flat the likelihood function is with respect to many of the parameters in the covariance equations. A reasonable procedure is to estimate a restricted model such as the diagonal model or the *BEKK* model with $K = 1$, then use a Lagrange multiplier test to examine the validity of the restriction.

Define Ξ to be the non-redundant parameters in the covariance equations, so that

$$\Xi' = [(RC_0)', (RA_1S)', \ldots, (RA_qS)', (RG_1S)', \ldots, (RG_pS)']$$

for the *vec* representation or

$$\Xi' = [C_0^{*\prime}, A_{11}^{*\prime}, \ldots, A_{qK}^{*\prime}, G_{11}^{*\prime}, \ldots, G_{pK}^{*\prime}]$$

for the *BEKK* representation. Then the maximum likelihood estimator for the parameters in the structural model, Γ, B, Λ and Ξ, is found by maximizing (4.2) with respect to these parameters. Unfortunately, the properties of MLEs in GARCH models are still open to debate, so little can confidently be said about the asymptotic properties of this estimator. Furthermore, the properties of maximum likelihood estimates of GARCH-in-mean models ($\Lambda \neq 0$) have not even been addressed in the literature. Instead of attempting to close the book on these issues here, we refer the reader to the papers by Bollerslev and Wooldridge [5], Lee and Hansen [16], Lumsdaine [17], Pagan and Sabau [20], or Weiss [21], which all address (with varying degrees of success) the properties of maximum likelihood and quasi-maximum likelihood estimates in univariate ARCH models.

Letting $\theta' = [(\mathrm{vec}\Gamma)', (\mathrm{vec}B)', (\mathrm{vec}\Lambda)', (\mathrm{vec}\Xi)']$ and differentiating (4.2) with respect to θ gives

$$\frac{\partial L_t}{\partial \theta} = \frac{1}{2}\left(\frac{\partial h_t}{\partial \theta}\right)' \left(H_t^{-1} \otimes H_t^{-1}\right) \mathrm{vec}(\epsilon_t \epsilon_t' - H_t) - \left(\frac{\partial \mathrm{vec}\Gamma}{\partial \theta}\right)' \mathrm{vec}(\Gamma^{-1\prime}) - \left(\frac{\partial \epsilon_t}{\partial \theta}\right)' H_t^{-1} \epsilon_t.$$

Noting, for example, that[8]

$$\frac{\partial h_t}{\partial \mathrm{vec}B} = \sum_{i=1}^{q} A_i[(I \otimes \epsilon_{t-i}) + (\epsilon_{t-i} \otimes I)]\frac{\partial \epsilon_{t-i}}{\partial \mathrm{vec}B} + \sum_{i=1}^{p} G_i \frac{\partial h_{t-i}}{\partial \mathrm{vec}B}$$

and that

$$\frac{\partial \epsilon_t}{\partial \mathrm{vec}B} = (x_t' \otimes I) + \Lambda \frac{\partial h_t}{\partial \mathrm{vec}B}$$

we see that calculation of $\frac{\partial h_t}{\partial \theta}$ is complicated by the fact that $\frac{\partial h_t}{\partial \theta}$ depends on $\frac{\partial h_{t-1}}{\partial \theta}$ if $G_i \neq 0$ for any i or if $\Lambda \neq 0$. Hence, the use of analytical derivatives would require calculating $\frac{\partial h_t}{\partial \theta}$ recursively, with $\frac{\partial h_0}{\partial \theta}$ assumed to be independent of θ. Because of this problem and because the derivatives $\frac{\partial h_t}{\partial \theta}$ and $\frac{\partial \epsilon_t}{\partial \theta}$ are so cumbersome, nonlinear maximization methods and numerical derivatives seem appropriate. Many nonlinear maximization methods are available, but a particularly convenient one and one that the authors have found useful in practice is the Berndt, Hall, Hall and Hausman [1] algorithm, which is an iterative method of calculating the optimal parameters in which the updating term is found by a regression of a vector of one's on the scores $\frac{\partial L}{\partial \theta}$:

$$\theta^{i+1} = \theta^i + \tau_i (S'S)^{-1} S' \iota.$$

Here, $[S]_{tk} = \frac{\partial L_t}{\partial \theta_k}$, i represents the iteration number, ι is the vector of ones, and τ_i is the step length which is calculated at each iteration by a line search. There are three features of the BHHH algorithm which make it particularly advantageous in these models. First, it is easy to use in practice, because its use requires little more than a subroutine to compute numerical derivatives and a subroutine to compute OLS regression parameters. Second, under normality, the $(S'S)^{-1}$ from the final iteration can be used as a consistent estimate of the variance-covariance matrix of the parameters. And third, Lagrange multiplier statistics are easily computed as T times the R^2 of the regression in the first step of the BHHH iteration, starting at the estimates under the null. This provides a particularly easy way to test the validity of any restrictions that might have been imposed during estimation, such as diagonality, $p = q = 1$ or $\Lambda = 0$, and suggests the following model building strategy: Begin by estimating a *diagonal BEKK* model with $K = 1$, then use the Lagrange multiplier test described above to examine whether the diagonality restriction is valid. If not, then either additional factors can be added or additional terms added to the first factor. Alternatively, if the *vech* model is chosen, then one could begin by estimating a diagonal *vech* model, then use the above LM test to examine the validity of the restriction. If rejected, then the appropriate non-diagonal terms should be added. This "bottom-up" model building procedure is easy to use in practice because diagonal models are not difficult to estimate, and once they are estimated, the LM tests are very easy to compute using the procedure described above. Furthermore, if the restrictions are rejected, then the diagonal model provides an obvious set of starting values to use in the nonlinear estimation of the unrestricted model. However, as usual, this model building procedure is only guaranteed to be consistent in certain special cases.

Of course, BHHH is only one of several possible optimization algorithms, each of which has its advantages. For example, it is widely recognized that as one gets closer to the optimum, the benefits to calculating the expected value of the Hessian become enormous, relative to using the outer product of the scores (as in BHHH). This is both because the convergence rate tends to increase enough to offset the extra effort required, and because the estimated standard errors tend to become more accurate.

Note that if one is estimating the reduced form model and if $\Lambda = 0$, then the information matrix is block-diagonal between the parameters in the mean equations and the parameters in the covariance equations (see Kraft and Engle, [13]). This means that efficient estimates of Π_1 can be calculated independently of Ξ, given only $\sqrt{T}$-consistent estimates of Ξ. Similarly, efficient estimates of Ξ can be calculated independently of Π_1, given only $\sqrt{T}$-consistent estimates of Π_1. This suggests the following estimation procedure: first obtain consistent estimates of Π_1 (call them $\tilde{\Pi}_1$) by using a seemingly unrelated regression[9] Then maximize the likelihood function (4.1) with respect to $\mathrm{vec}\Xi$, given $\tilde{\Pi}_1$, to get an estimate of $\mathrm{vec}\Xi$, say $\mathrm{vec}\hat{\Xi}$. Finally, maximize (4.1) again, this time with respect to Π_1 (given $\mathrm{vec}\hat{\Xi}$), imposing the structural restrictions, to get the estimates $\hat{\Pi}_1$. $(\hat{\Pi}_1, \mathrm{vec}\hat{\Xi})$ are asymptotically equivalent to the maximum likelihood estimates. If this procedure is iterated and converges, the estimates will solve the first order conditions for full information maximum likelihood estimation.

5. Conclusion

This paper extends Engle's [8] ARCH model and Bollerslev's [2] GARCH model to a multivariate setting. To parameterize the multivariate process so that positive definiteness is insured, a new formulation is presented and compared with that used in much of the existing multivariate ARCH literature. Equivalence relations between these parameterizations are derived, conditions for covariance stationarity are presented, and the relationship between the reduced form and structural models is analyzed. Maximum likelihood estimation of the system is then discussed, though we have little to say about the properties of this estimator. This area is perhaps one of the most important areas for future research in multivariate ARCH modelling. Very little is currently known about the properties of maximum likelihood estimators in *univariate* GARCH models, let alone in *multivariate* GARCH-*in-mean* models, despite the fact that this estimator permeates the multivariate GARCH-in-mean literature.

Appendix - Proofs

PROPOSITION 2.1.

First, we show that if the diagonal elements of C_0^* are restricted to be positive, then the C_0^* matrix is identified: If the diagonal elements of C_0^* are positive, then we know that $C_0^{*\prime}C_0^*$ is a positive definite matrix. But by Proposition 58 and Remark 34 of Dhrymes ([6], pp. 68-69), the decomposition of a positive definite matrix into the product of a triangular matrix and its transpose always exists, and this decomposition is unique if the diagonal terms are restricted to be positive.

Next, we show that $a_{11}^* > 0$ is sufficient to identify the A_{11}^* matrix. Straightforward algebra reveals that the lm^{th} element of H_t is

$$h_{lm,t} = c + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{il}^* a_{jm}^* \epsilon_{i,t-1}\epsilon_{j,t-1} + \sum_{i=1}^{n}\sum_{j=1}^{n} g_{il}^* g_{jm}^* h_{ij,t-1}.$$

So, for example, ignoring the constant and the GARCH terms, the (1,1) element of H_t is

$$h_{11,t} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i1}^* a_{j1}^* \epsilon_{i,t-1}\epsilon_{j,t-1}.$$

So the coefficient attached to $\epsilon_{1,t-1}^2$ is a_{11}^{*2}, meaning that a_{11}^* is identified, up to its sign. Restricting it to be positive identifies this term. Next, the coefficient attached to $\epsilon_{1,t-1}\epsilon_{j,t-1}$ in this equation is $(a_{11}^* a_{j1}^* + a_{j1}^* a_{11}^*) = 2a_{11}^* a_{j1}^*$. So a_{j1}^* is identified for all j because a_{11}^* is identified. So we have that the first column of A_{11} is identified. Next, the $(1,2)$ element of H_t is, again ignoring the constant and the GARCH terms,

$$h_{12,t} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i1}^* a_{j2}^* \epsilon_{i,t-1}\epsilon_{j,t-1}.$$

The coefficient attached to $\epsilon_{1,t-1}\epsilon_{1,t-1}$ in this equation is $(a_{11}^* a_{12}^* + a_{11}^* a_{12}^*)$. Because a_{11}^* is identified, a_{12}^* must also be identified. Similar logic can be used to show that all the other terms in A_{11}^* are identified, completing the proof that the matrix A_{11} is identified.

The identification of the G_{11} matrix follows an identical argument as above, and is therefore not presented here.

Q.E.D.

Proposition 2.2.

a) The proof here is trivial. Ignoring the constants and GARCH terms, there are a total of s^2 variables on the right-hand side of the covariance equations, so if we have fewer than s^2 parameters, some restrictions are implicitly being imposed.

b) Elementary algebra reveals that the lm^{th} element of H_t in the *BEKK* model with $K > 1$ is

$$h_{lm,t} = c + \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} a^*_{il,k} a^*_{jm,k} \epsilon_{i,t-1}\epsilon_{j,t-1} + \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} g^*_{il,k} g^*_{jm,k} h_{ij,t-1}.$$

So the coefficient attached to $\epsilon_{i,t-1}\epsilon_{j,t-1}$ in the $h_{lm,t}$ equation is

$$\sum_{k=1}^{K} a^*_{il,k} a^*_{jm,k} + a^*_{jl,k} a^*_{im,k}$$

if $i \neq j$ and

$$\sum_{k=1}^{K} a^*_{il,k} a^*_{im,k}$$

if $i = j$. Clearly, the term $\epsilon_{i,t-1}\epsilon_{j,t-1}$ will drop out of the $h_{lm,t}$ equation if there is no matrix A^*_{1k} which contains either both of $a^*_{il,k}$ and $a^*_{jm,k}$ or both of $a^*_{jl,k}$ and $a^*_{im,k}$. Hence, for the model to be fully general, we require that either the pair $(a^*_{il,k}, a^*_{jm,k})$ or the pair $(a^*_{jl,k}, a^*_{im,k})$ appear in a A_{1k} matrix for some k.

Q.E.D.

Proposition 2.3.

First, in the proof to Proposition 2.1 we show that the conditions on C^*_0 are sufficient to ensure that the elements of C^*_0 are identified. Also, as in the proof to Proposition 2.1, we focus our proof on the ARCH terms because the proof for the GARCH terms is directly analagous. The proof proceeds in a manner similar to the proof of Proposition 2.1.

From the proof to Proposition 2.2, we have

$$h_{lm,t} = c + \sum_{k_r=1}^{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} a^*_{il,k_r} a^*_{jm,k_r} \epsilon_{i,t-1}\epsilon_{j,t-1} + \sum_{k_r=1}^{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g^*_{il,k_r} g^*_{jm,k_r} h_{ij,t-1},$$

and that the coefficient on $\epsilon^2_{1,t-1}$ in the $h_{11,t}$ equation is $\sum_{k_r=1}^{n^2} a^{*2}_{11,k_r}$, which is just $a^{*2}_{11,1}$ because the restrictions on the $A^*_{1k_r}$ matrices ensure that $a^*_{11,k_r} = 0$ for all $k_r > 1$. Therefore, $a^*_{11,1}$ is identified up to its sign. Next, the coefficient on $\epsilon_{1,t-1}\epsilon_{j,t-1}$ ($j \neq 1$) in the $h_{11,t}$ equation is $2\sum_{k_r=1}^{n^2} a^*_{11,k_r} a^*_{j1,k_r}$, which is $2a^*_{11,1}a^*_{j1,1}$ because the restrictions on the $A^*_{1k_r}$ matrices ensure that $a^*_{11,k_r} = 0$ for all $k_r > 1$. Therefore, $a^*_{j1,1}$, $j = 2, \ldots, n$, are identified (up to their sign) because $a^*_{11,1}$ is identified. The coefficient on $\epsilon^2_{2,t-1}$ in the $h_{11,t}$ equation is $\sum_{k_r=1}^{n^2} a^{*2}_{21,k_r}$, which becomes $(a^{*2}_{21,1} + a^{*2}_{21,2})$ because the zero restrictions on the other A_{1k_r} matrices imply that $a^*_{21,k_r} = 0$ for all $k_r > 2$. Therefore, $a^*_{21,2}$ is identified (up to its sign) because $a^2_{21,1}$ was previously identified. Proceeding similarly, the coefficient on $\epsilon_{2,t-1}\epsilon_{j,t-1}$ (if $j \neq 2$) is $2\sum_{k_r=1}^{n^2} a^*_{21,k_r} a^*_{j1,k_r}$, which becomes $2(a^*_{21,1}a^*_{j1,1} + a^*_{21,2}a^*_{j1,2})$ because the zero restrictions on the other $A^*_{1k_r}$ matrices imply that $a^*_{21,k_r} = 0$ for all $k_r > 2$. So $a^*_{j1,2}$, $j = 3, 4, \ldots, n$ are identified (up to sign), because all the other terms in $2(a^*_{21,1}a^*_{j1,1} + a^*_{21,2}a^*_{j1,2})$ were already identified. Similar logic can be used to identify a^*_{j1,k_r}, $k_r = 3, \ldots, n$. So we have that the first columns of the $A^*_{1k_r}$ matrices, $k_r = 1, \ldots, n$, are identified (except that each column can be multiplied by -1 without changing the model).

To identify the second columns of the $A^*_{1k_r}$ matrices, $k_r = 1, \ldots, n$, we look at the $h_{12,t}$ equation. The coefficient on $\epsilon^2_{1,t-1}$ is $\sum_{k_r=1}^{n^2} a^*_{11,k_r} a^*_{12,k_r}$, which equals $a^*_{11,1}a^*_{12,1}$ because the zero restrictions on the $A^*_{1k_r}$ matrices ensures that $a^*_{11,k_r} = 0$ for all $k_r > 1$. This identifies $a^*_{12,1}$ because $a^*_{11,1}$ is identified. The coefficient on $\epsilon_{1,t-1}\epsilon_{j,t-1}$, after accounting for the zero restrictions, is $a^*_{11,1}a^*_{j2,1} + a^*_{j1,1}a^*_{12,1}$, which identifies $a^*_{j2,1}$, $j = 2, \ldots, n$, because the other terms are identified. So the second column of A^*_{11} is identified. The coefficient on $\epsilon^2_{2,t-1}$ is $a^*_{21,1}a^*_{22,1} + a^*_{21,2}a^*_{22,2}$ which identifies $a^*_{22,2}$, and the coefficient on $\epsilon_{2,t-1}\epsilon_{j,t-1}$, $j = 3, \ldots, n$, is $a^*_{21,1}a^*_{j2,1} + a^*_{j1,1}a^*_{22,1} + a^*_{21,2}a^*_{j2,2} + a^*_{22,2}a^*_{j1,2}$, which identifies $a^*_{j2,2}$, $j = 3, \ldots, n$. Therefore, the second column of A^*_{12} is identified, and continuing along these lines will identify the second columns of the remaining A_{1k_r} matrices, $k_r = 1, \ldots, n$.

Proceeding similarly will identify the remaining columns of the matrices $A^*_{11}, \ldots, A^*_{1n}$ up to sign. The restriction that $a^*_{nn,k_r} > 0$ will eliminate the sign problem, and these matrices are now identified. Identification of the remaining matrices proceeds analagously. For example, the elements of the $h_{22,t}$ equation will be used to identify the second column of the matrices $A^*_{1,n+1}, \ldots, A^*_{1,2n}$, and the elements of the $h_{23,t}$ equation will be used to identify the third column of these matrices.

Finally, we see from the above proof that every coefficient in every equation has a free parameter, and therefore each coefficient can take on any value permitted by the *BEKK* model. Therefore, adding other $A^*_{1k_r}$ matrices cannot give any models which were previously being precluded. So the model is fully general.

Q.E.D.

PROPOSITION 2.4.

For simplicity, let $p = q = 1$. Then the *vec* parameterization becomes

$$h_t = [C_0 \vdots A_1 \vdots G_1] \begin{bmatrix} 1 \\ \eta_{t-1} \\ h_{t-1} \end{bmatrix} \tag{A.1}$$

and the *BEKK* parameterization becomes

$$H_t = C_0^{*\prime} C_0^* + \sum_{k=1}^{K} A_{1k}^{*\prime} \epsilon_{t-1} \epsilon_{t-1}^{\prime} A_{1k}^* + \sum_{k=1}^{K} G_{1k}^{*\prime} H_{t-1} G_{1k}^*. \tag{A.2}$$

Vectorizing,

$$\begin{aligned} h_t &= \mathrm{vec}(C_0^{*\prime} C_0^*) + \mathrm{vec} \sum_{k=1}^{K} A_{1k}^{*\prime} \epsilon_{t-1} \epsilon_{t-1}^{\prime} A_{1k}^* + \mathrm{vec} \sum_{k=1}^{K} G_{1k}^{*\prime} H_{t-1} G_{1k}^* \\ &= \mathrm{vec}(C_0^{*\prime} C_0^*) + \sum_{k=1}^{K} (A_{1k}^* \otimes A_{1k}^*)^{\prime} \mathrm{vec}(\epsilon_{t-1} \epsilon_{t-1}^{\prime}) + \sum_{k=1}^{K} (G_{1k}^* \otimes G_{1k}^*)^{\prime} \mathrm{vec}(H_{t-1}) \\ &= (C_0^* \otimes C_0^*)^{\prime} \mathrm{vec}(I_n) + \sum_{k=1}^{K} (A_{1k}^* \otimes A_{1k}^*)^{\prime} \eta_{t-1} + \sum_{k=1}^{K} (G_{1k}^* \otimes G_{1k}^*)^{\prime} h_{t-1}. \end{aligned} \tag{A.3}$$

Now if relations (2.5) hold, then (A.3) becomes

$$h_t = C_0 + A_1 \eta_{t-1} + G_1 h_{t-1}$$

which is (A.1), proving sufficiency. Necessity can be shown by noting that (A.1) and (A.3) must hold for all ϵ_{t-1}, so by appropriate choice of ϵ_{t-1}, each column of A_1 can be equated individually with each column of $\sum_{k=1}^{K} (A_{1k}^* \otimes A_{1k}^*)^{\prime}$. For example, letting $\epsilon_{t-1}^{\prime} = (1, 0, \ldots, 0)$ establishes equality of the first column of A_1 with the first column of $\sum_{k=1}^{K} (A_{1k}^* \otimes A_{1k}^*)^{\prime}$. Necessity of the rest of the relations (2.5) can be shown in the same way.

Q.E.D.

PROPOSITION 2.5.

For simplicity, we look at GARCH(1,1) models, but the generalization to GARCH(p, q) models is obvious. The *BEKK* parameterization is

$$H_t = C_0^{*\prime} C_0^* + \sum_{k=1}^{K} A_{1k}^{*\prime} \epsilon_{t-1} \epsilon_{t-1}^{\prime} A_{1k}^* + \sum_{k=1}^{K} G_{1k}^{*\prime} H_{t-1} G_{1k}^*.$$

The proof proceeds by induction. First, H_t is positive definite for $t = 1$: Clearly, the term $\sum_{k=1}^{K} A_{1k}^{*\prime}\epsilon_{t-1}\epsilon_{t-1}^{\prime}A_{1k}^{*}$ is positive semidefinite because $\epsilon_0\epsilon_0^{\prime}$ is positive semidefinite. Also, if the null space condition holds, then

$$C_0^{*\prime}C_0^{*} + \sum_{k=1}^{K} G_{1k}^{*\prime}H_0G_{1k}^{*} \tag{A.4}$$

is positive definite. This is clearly true if C_0^* or any G_{1k}^* has full rank, but to show that the null space condition is sufficient, note that the expression (A.4) is positive definite if and only if

$$x^{\prime}\left(C_0^{*\prime}C_0^{*} + \sum_{k=1}^{K} G_{1k}^{*\prime}H_0G_{1k}^{*}\right)x > 0 \qquad \forall x \neq 0$$

or

$$(C_0^{*}x)^{\prime}(C_0^{*}x) + \sum_{k=1}^{K}(H_0^{\frac{1}{2}}G_{ik}^{*}x)^{\prime}(H_0^{\frac{1}{2}}G_{ik}^{*}x) > 0 \qquad \forall x \neq 0 \tag{A.5}$$

where $H_0 = H_0^{\frac{1}{2}\prime}H_0^{\frac{1}{2}}$ and $H_0^{\frac{1}{2}}$ is full rank. But defining $N[P]$ to be the null space of the matrix P, (A.5) is true if and only if

$$N[C_0^*] \cap N[H_0^{\frac{1}{2}}G_{11}^*] \cap \ldots \cap N[H_0^{\frac{1}{2}}G_{1K}^*] = \emptyset.$$

Noting that $N[H_0^{\frac{1}{2}}G_{1k}^*]$ is the same as $N[G_{1k}^*]$ because $H_0^{\frac{1}{2}}$ is full rank gives the desired result — (A.4) is positive definite if and only if $N[C_0^*] \cap N[G_{11}^*] \cap \ldots \cap N[G_{1K}^*] = \emptyset$.

Now suppose that the statement is true for $n = \tau$. I.e. suppose

$$H_\tau = C_0^{*\prime}C_0^{*} + \sum_{k=1}^{K} A_{1k}^{*\prime}\epsilon_{\tau-1}\epsilon_{\tau-1}^{\prime}A_{1k}^{*} + \sum_{k=1}^{K} G_{1k}^{*\prime}H_{\tau-1}G_{1k}^{*} \tag{A.6}$$

is positive definite. Then

$$H_{\tau+1} = C_0^{*\prime}C_0^{*} + \sum_{k=1}^{K} A_{1k}^{*\prime}\epsilon_{\tau}\epsilon_{\tau}^{\prime}A_{1k}^{*} + \sum_{k=1}^{K} G_{1k}^{*\prime}H_{\tau}G_{1k}^{*}$$

is positive definite: First, following a similar line of argument as above, the term $\sum_{k=1}^{K} A_{1k}^{*\prime}\epsilon_{\tau}\epsilon_{\tau}^{\prime}A_{1k}^{*}$ is positive semidefinite. Second, the term

$$C_0^{*\prime}C_0^{*} + \sum_{k=1}^{K} G_{1k}^{*\prime}H_{\tau}G_{1k}^{*}$$

is positive definite if and only if the null space condition holds, because H_τ is positive definite by the induction assumption. This can be shown by following the identical steps as in the above paragraph, replacing H_0 with H_τ. So

$$H_{\tau+1} = C_0^{*\prime}C_0^{*} + \sum_{k=1}^{K} A_{1k}^{*\prime}\epsilon_{\tau}\epsilon_{\tau}^{\prime}A_{1k}^{*} + \sum_{k=1}^{K} G_{1k}^{*\prime}H_{\tau}G_{1k}^{*}$$

is positive definite, meaning H_τ is positive definite for all τ.

Q.E.D.

PROPOSITION 2.6.

First, suppose $C_0 = \mathrm{vec}(\Omega)$ where Ω is positive definite. Then $\Omega = C_0^{*'} C_0^*$ for some triangular C_0^*. Therefore, $C_0 = \mathrm{vec}(\Omega) = \mathrm{vec}(C_0^{*'} C_0^*)$.

To prove the remaining equations, we again assume $p = q = 1$. The extension to higher order GARCH models is obvious. Also, we only show that the equations relating A_i and A_{ik}^* hold; the proof for the equations relating G_i and G_{ik}^* is directly analagous. First, consider the second term in the *diagonal vec* model, which we will label $\tilde{H}_t$. This term can be written as

$$\tilde{H}_t = \begin{bmatrix} \tilde{a}_{11}\epsilon_{1,t-1}^2 & \tilde{a}_{12}\epsilon_{1,t-1}\epsilon_{2,t-1} & \dots & \tilde{a}_{1n}\epsilon_{1,t-1}\epsilon_{n,t-1} \\ \tilde{a}_{21}\epsilon_{2,t-1}\epsilon_{1,t-1} & \tilde{a}_{22}\epsilon_{2,t-1}^2 & \dots & \tilde{a}_{2n}\epsilon_{2,t-1}\epsilon_{n,t-1} \\ \vdots & \vdots & \dots & \vdots \\ \tilde{a}_{n1}\epsilon_{n,t-1}\epsilon_{1,t-1} & \tilde{a}_{n2}\epsilon_{n,t-1}\epsilon_{2,t-1} & \dots & \tilde{a}_{nn}\epsilon_{n,t-1}^2 \end{bmatrix}. \tag{A.7}$$

If H_t is positive definite for all realizations of ϵ_t, then $\tilde{H}_t$ must be positive semidefinite for all realizations of ϵ_t. But $\tilde{H}_t$ is positive semidefinite if and only if all its principal minors are nonnegative. Also, the principal minors of (A.7) have the same sign as the principal minors of the matrix

$$\tilde{A} = \begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \dots & \tilde{a}_{1n} \\ \tilde{a}_{21} & & & \tilde{a}_{2n} \\ \vdots & & & \vdots \\ \tilde{a}_{n1} & & \dots & \tilde{a}_{nn} \end{bmatrix},$$

implying that if $\tilde{H}_t$ is positive semidefinite for all realizations of ϵ_t, then $\tilde{A}$ must be positive semidefinite. But if $\tilde{A}$ is positive semidefinite, then it can always be decomposed into $\tilde{A} = B'B$ with B triangular. Define B to be

$$B = \begin{bmatrix} b_{11,n} & b_{11,n-1} & \dots & b_{11,1} \\ 0 & b_{22,n-1} & \dots & b_{22,1} \\ \vdots & & & \vdots \\ 0 & 0 & \dots & b_{nn,1} \end{bmatrix}.$$

Then

$$\tilde{A} = B'B = \begin{bmatrix} \sum_{k=1}^{n} b_{11,k}^2 & \sum_{k=1}^{n-1} b_{11,k}b_{22,k} & \dots & b_{11,1}b_{nn,1} \\ \cdot & \sum_{k=1}^{n-1} b_{22,k}^2 & \dots & b_{22,1}b_{nn,1} \\ \vdots & & & \vdots \\ \cdot & & & b_{nn,1}^2 \end{bmatrix}.$$

So if H_t is positive definite, then there exist $b_{ij,k}$ such that

$$\tilde{H}_t = \begin{bmatrix} \sum_{k=1}^{n} b_{11,k}^2\epsilon_{1,t-1}^2 & \sum_{k=1}^{n-1} b_{11,k}b_{22,k}\epsilon_{1,t-1}\epsilon_{2,t-1} & \dots & b_{11,1}b_{nn,1}\epsilon_{1,t-1}\epsilon_{n,t-1} \\ \cdot & \sum_{k=1}^{n-1} b_{22,k}^2\epsilon_{2,t-1}^2 & \dots & b_{22,1}b_{nn,1}\epsilon_{2,t-1}\epsilon_{n,t-1} \\ \vdots & & & \vdots \\ \cdot & & & b_{nn,1}^2\epsilon_{n,t-1}^2 \end{bmatrix}. \tag{A.8}$$

But notice that in the *diagonal BEKK* model, if we define

$$A_{ik}^{*} = \begin{bmatrix} a_{11,k}^{*} & & & & & & \\ & a_{22,k}^{*} & & & & 0 & \\ & & \ddots & & & & \\ & & & a_{rr,k}^{*} & & & \\ & & & & 0 & & \\ & 0 & & & & \ddots & \\ & & & & & & 0 \end{bmatrix},$$

where $r = n - k + 1$; i.e. if we define $A_{1k}^{*} = \text{diag}\{a_{11,k}^{*}, a_{22,k}^{*}, \ldots, a_{rr,k}^{*}, 0, \ldots, 0\}$, then evaluation of $H_t = \sum_{k=1}^{n} A_{1k}^{*\prime}\epsilon_{t-1}\epsilon_{t-1}^{\prime}A_{1k}^{*}$ gives

$$H_t = \begin{bmatrix} \sum_{k=1}^{n} a_{11,k}^{*2}\epsilon_{1,t-1}^{2} & \sum_{k=1}^{n-1} a_{11,k}^{*}a_{22,k}^{*}\epsilon_{1,t-1}\epsilon_{2,t-1} & \cdots & a_{11,1}^{*}a_{nn,1}^{*}\epsilon_{1,t-1}\epsilon_{n,t-1} \\ \cdot & \sum_{k=1}^{n-1} a_{22,k}^{*2}\epsilon_{2,t-1}^{2} & \cdots & a_{22,1}^{*}a_{nn,1}^{*}\epsilon_{2,t-1}\epsilon_{n,t-1} \\ \vdots & & & \vdots \\ \cdot & & & a_{nn,1}^{*2}\epsilon_{n,t-1}^{2} \end{bmatrix}.$$

Comparing this with (A.8), we see that if H_t is positive definite then we can always choose each $a_{ii,k}^{*}$ in the diagonal A_{1k}^{*} matrices equal to be equal to $b_{ii,k}$ from the decomposition of $\tilde{A}$, implying that if H_t is positive definite then it can always be written in the *BEKK* framework.

Q.E.D.

PROPOSITION 2.7.

Again, we show this for a GARCH(1,1) model, but the extension to GARCH(p, q) models is trivial. Also, we prove this only for the *vec* model. The proof for the *BEKK* model can be obtained by substituting relations (2.5) into the following proof.

Let E_{t-1} be the expectations operator, conditioned on the information set $\Im_{t-1}$. Then

$$\begin{aligned}
h_t = \mathrm{E}_{t-1}\eta_t &= C_0 + A_1\eta_{t-1} + G_1 h_{t-1} \\
\mathrm{E}_{t-2}\eta_t &= C_0 + A_1\mathrm{E}_{t-2}\eta_{t-1} + G_1\mathrm{E}_{t-2}h_{t-1} \\
&= C_0 + [A_1 + G_1]h_{t-1} \\
\mathrm{E}_{t-3}\eta_t &= C_0 + A_1\mathrm{E}_{t-3}\eta_{t-1} + G_1\mathrm{E}_{t-3}h_{t-1} \\
&= C_0 + A_1\mathrm{E}_{t-3}\eta_{t-1} + G_1\mathrm{E}_{t-3}[C_0 + A_1\eta_{t-2} + G_1 h_{t-2}] \\
&= C_0 + A_1\mathrm{E}_{t-3}\eta_{t-1} + G_1 C_0 + G_1 A_1 h_{t-2} + G_1^2 h_{t-2} \\
&= (I + G_1)C_0 + A_1\mathrm{E}_{t-3}\eta_{t-1} + G_1(A_1 + G_1)h_{t-2} \\
&= (I + G_1)C_0 + A_1[C_0 + (A_1 + G_1)h_{t-2}] + G_1(A_1 + G_1)h_{t-2} \\
&= [I + (A_1 + G_1)]C_0 + (A_1 + G_1)^2 h_{t-2} \\
&\;\vdots \\
\mathrm{E}_{t-\tau}\eta_t &= [I + (A_1 + G_1) + \cdots + (A_1 + G_1)^{\tau-2}]C_0 + (A_1 + G_1)^{\tau-1}h_{t-\tau+1}
\end{aligned}$$

It is widely known that for any square matrix Z, $Z^\tau \longrightarrow 0$ as $\tau \to \infty$ if and only if all the eigenvalues of Z are less than one in modulus, and that the eigenvalues of Z are less than one in modulus if and only if $[I + Z + Z^2 + \cdots] \longrightarrow (I - Z)^{-1}$. Therefore, if $\lim_{\tau\to\infty} h_{t-\tau}$ is finite, $\mathrm{E}_{t-\tau}\eta_t$ converges in probability (as $\tau \to \infty$) to $[I - A_1 - G_1]^{-1}C_0$ if and only if the eigenvalues of $(A_1 + G_1)$ are all less than one in modulus. Also, by the law of iterated expectations, $\mathrm{E}(\epsilon_t\epsilon'_{t+\gamma}) = \mathrm{E}[\mathrm{E}_t(\epsilon_t\epsilon_{t+\gamma})] = 0$ for all $\gamma \neq 0$. Therefore, $\mathrm{E}(\epsilon_t\epsilon_{t+\gamma})$ exists and depends only on γ for all integers $t \geq 1$ and $\gamma \geq t+1$.

Q.E.D.

Proposition 3.1.

Again, the proof is presented for the *vec* model, but an application of relations (2.5) reveals that the proof also applies for the *BEKK* model. Suppose

$$\begin{aligned}
\epsilon_t \,|\, \Im_{t-1} &\sim N(0, H_t) \\
h_t &= C_0 + \sum_{i=1}^{q} A_i\eta_{t-i} + \sum_{i=1}^{p} G_i h_{t-i}.
\end{aligned}$$

Define $\epsilon_t^* \equiv P\epsilon_t$, $h_t^* \equiv \mathrm{vec}PH_tP'$ and $\Psi \equiv (P \otimes P)$. Then, ignoring the summation limits,

$\epsilon_t^* \,|\, \Im_{t-1} \;\sim\; N(0, PH_tP')$ with

$$
\begin{aligned}
h_t^* &= \mathrm{vec}PH_tP' \\
&= (P \otimes P)\mathrm{vec}H_t \\
&= \Psi h_t \\
&= \Psi C_0 + \Psi \sum A_i \eta_{t-i} + \Psi \sum G_i h_{t-i} \\
&= \Psi C_0 + \Psi \sum A_i \mathrm{vec}(\epsilon_{t-i}\epsilon_{t-i}') + \Psi \sum G_i \mathrm{vec}H_{t-i} \\
&= \Psi C_0 + \Psi \sum A_i \mathrm{vec}(P^{-1}P\epsilon_{t-i}\epsilon_{t-i}'P'P^{-1\prime}) + \Psi \sum G_i \mathrm{vec}(P^{-1}PH_{t-i}P'P^{-1\prime}) \\
&= \Psi C_0 + \Psi \sum A_i (P^{-1} \otimes P^{-1})\mathrm{vec}(P\epsilon_{t-i}\epsilon_{t-i}'P') + \Psi \sum G_i (P^{-1} \otimes P^{-1})\mathrm{vec}(PH_{t-i}P') \\
&= \Psi C_0 + \Psi \sum A_i (P^{-1} \otimes P^{-1})\mathrm{vec}(\epsilon_{t-i}^*\epsilon_{t-i}^{*\prime}) + \Psi \sum G_i (P^{-1} \otimes P^{-1}) h_{t-i}^* \\
&= \Psi C_0 + \Psi \sum A_i (P^{-1} \otimes P^{-1})\eta_{t-i}^* + \Psi \sum G_i (P^{-1} \otimes P^{-1}) h_{t-i}^* \\
&= \Psi C_0 + \sum \Psi A_i \Psi^{-1} \eta_{t-i}^* + \sum \Psi G_i \Psi^{-1} h_{t-i}^*.
\end{aligned}
$$

Hence, $P\epsilon_t$ follows a GARCH process of the same order as ϵ_t.

Q.E.D.

Bibliography

1. Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* 3/4 (1974): 653-665.

2. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (1986): 307-328.

3. Bollerslev, T., R.Y. Chou and K.F. Kroner. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52 (1992): 5-59.

4. Bollerslev, T., R.F. Engle and J.M. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy* 96 (1988): 116-131.

5. Bollerslev, T. and J.M. Wooldridge. Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*, forthcoming, 1991.

6. Dhrymes, P.J. *Mathematics for econometrics*, 2^{nd} edition. Springer-Verlag: New York, 1984.

7. Engel, C. and A.P. Rodrigues. Tests of international CAPM with time varying covariances. *Journal of Applied Econometrics* 4 (1989): 119-128.

8. Engle, R.F. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (1982): 987-1007.

9. Engle, R.F., C.W.J. Granger and D.F. Kraft. Combining competing forecasts of inflation using a bivariate ARCH model. *Journal of Economic Dynamics and Control* 8 (1984): 151-165.

10. Engle, R.F., D.F. Hendry and J.-F. Richard. Exogeneity. *Econometrica* 51 (1983): 277-304.

11. Friedman, M. Nobel Lecture: Inflation and unemployment. *Journal of Political Economy* 85 (1977): 451-472.

12. Kaminsky, G.L. and R. Peruga. Can a time varying risk premium explain excess returns in the forward market for foreign exchange? *Journal of International Economics* 28 (1990): 47-70.

13. Kraft, D.F. and R.F. Engle. Autoregressive conditional heteroskedasticity in multiple time series models. unpublished manuscript, Department of Economics, UC San Diego, 1983.

14. Kroner, K.F. and S. Claessens. Optimal dynamic hedging portfolios and the currency composition of external debt. *Journal of International Money and Finance* 10 (1991): 131-148.

15. Kroner, K.F. and W.D. Lastrapes. The impact of exchange rate volatility on international trade: Estimates using the GARCH-M model. *Journal of International Money and Finance* forthcoming, 1993.

16. Lee, S.-W. and B.E. Hansen. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, forthcoming, 1993.

17. Lumsdaine, R.L. Asymptotic properties of the quasi-maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. unpublished manuscript, Princeton University and NBER, 1991.

18. McCurdy, T.H. and I.G. Morgan. Testing the martingale hypothesis in Deutschmark futures with models specifying the form of heteroskedasticity. *Journal of Applied Econometrics* 3 (1988): 187-202.

19. McCurdy, T.H. and I.G. Morgan. Tests for systematic risk components in deviations from uncovered interest rate parity. *Review of Economic Studies* 58 (1991): 587-602.

20. Pagan, A.R. and H.C.L. Sabau. On the inconsistency of the MLE in certain heteroskedastic regression models. unpublished manuscript, Department of Economics, University of Rochester, 1987.

21. Weiss, A.A. Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2 (1986): 107-131.

Footnotes

1. This acronym comes from an earlier version of this paper which synthesized the work on multivariate ARCH models by Yoshi Baba, Rob Engle, Dennis Kraft and Ken Kroner.

2. For the remainder of this section of the paper, the terms representing the exogenous influences will be dropped in order to shorten some of the equations and proofs. From the structure of the proofs, it should be clear that the extension would involve simply repeating some of the steps of the proofs or adding another term to the proofs.

3. To keep the notation as simple as possible, we focus on GARCH(1,1) models. The generalization to higher order models is trivial.

4. Again, we focus here on the GARCH(1,1) model, both because it keeps the notation simple and because most empirical work finds that $p = q = 1$ is a reasonable restriction to impose. Also, we focus only on the model with variances and covariances in the mean equation. The implications of using standard deviations or other functions of the variances, as in Kroner and Lastrapes [15], will be the same.

5. The vec*h* operator is the vector-half operator, which stacks the lower triangular portion of a matrix.

6. We would like to acknowledge an anonymous referee for suggesting this solution.

7. Notice that since the unconditional distribution is not normal, information in the higher moments and in the information matrix can be used to help identify the model.

8. This derivative assumes the *vec* model was used. A similar expression holds for the *BEKK* model.

9. See equation (2.6) for the formula for the unconditional covariance matrix.