Cholesky Factor GARCH

Hiroyuki Kawakatsu
Quantitative Micro Software
21 California Ave #302
Irvine, CA 92612
kawa@aris.ss.uci.edu

May 24, 2003

Abstract

In this paper, I propose a new class of multivariate GARCH models that specify the dynamics in terms of the cholesky factor of the conditional covariance. In the special case of a univariate model, this is equivalent to specifying the dynamics of the conditional standard deviation. The main advantage of the cholesky factor model is that it ensures positive definiteness of the conditional covariance without having to impose restrictions on the parameters of the model (except those to identify the model). The log-likelihood function of the model is particularly easy to evaluate. An empirical example illustrates estimation and inference using the cholesky factor model.

1 Introduction

As financial markets keep expanding with new instruments, so do the needs for tools to analyze volatility. One of the most successful volatility model is the family of generalized autoregressive conditional heteroskedasticity (GARCH) models, first introduced in Engle (1982). While the univariate family of GARCH models have been extensively analyzed, it is essential for risk analysis to extend the model in multivariate dimensions. A general family of multivariate GARCH models were explored in Engle and Kroner (1995) and Kroner and Ng (1998). There are two main difficulties in extending the univariate GARCH models in multivariate dimensions. The first is the "curse of dimensionality," in which the number of parameters to estimate grow rapidly with the dimension of the system k. In its most general form, the number of parameters is of order $O(k^4)$ and grow quadratically in k. Therefore, for practical use, it is essential to consider restricted models that reduce the number of parameters without overly restricting the dynamics of the system. The second is to find a specification that maintains positive definiteness of the conditional covariance matrix. This is a much harder condition to impose than simply ensuring that conditional variances remain positive in univariate models. Recent contributions to the literature tackle these difficulties by imposing a variety of simplifying assumptions (Tse and Tsui 2002, Engle 2002, Weide 2002, Ledoit, Santa-Clara and Wolf 2003).

One approach to ensuring positive definiteness is to transform the parameters of the model. For example, the well-known BEKK model in Engle and Kroner (1995) is parameterized so that all terms in the conditional variance equation are in a quadratic form. An alternative approach is to transform the covariance matrix and to specify the dynamics in the transformed space. The key to this approach is to find a transformation that frees us from the positive definiteness condition. In the univariate class of models, this is the approach taken by Nelson (1991) and Hentschel (1995). In this paper, I propose a new class of multivariate GARCH models that specify the dynamics in terms of the cholesky factor of the conditional covariance. In the univariate model, this is equivalent to specifying the dynamics of the conditional standard deviation. The main disadvantage of the cholesky factor model is that the parameters are difficult to interpret. Because the cholesky factor is a lower triangular matrix, the model imposes a recursive ordering in the system. However, the dynamics are no more restrictive than those of the other recent contributions which preclude cross-element dynamics.[1] As discussed below, the dynamics of the conditional covariance is highly non-linear in the parameters, the lagged conditional covariance, and the lagged squared innovations. If the researcher does not have prior information regarding the recursive structure of the system, a simple procedure based on the log-likelihood can be used to select an ordering that best fits the data. The main advantage of the cholesky factor model is that it ensures positive definite-

[1] By cross-element dynamics, I mean that the (i, j) element depends on lags of elements other than (i, j) in the dynamic specification.

ness of the conditional covariance without having to impose restrictions on the parameters of the model (except those to identify the model). The log-likelihood function of the model is particularly easy to evaluate and may be able to fit large dimensional models.

The plan of the paper is as follows. In the next section, I discuss specification and identification of the cholesky factor GARCH model and the properties of the model. In particular, I discuss several specifications that reduce the number of parameters in the model. Section 3 discusses estimation and inference. Section 4 compares the model with other multivariate GARCH models.

2 Specification and Identification

Assume the mean equation for the k-dimensional vector y_t is given by a general non-linear equation

$$\underset{k\times 1}{\epsilon_t} = f(y_t, h_t, \delta), \quad \epsilon_{t|t-1} \sim N(0, \underset{k\times k}{H_t}), \quad t = 1, \dots, T \tag{1}$$

where $f(\cdot)$ is a known function parameterized by the vector δ. We include elements of the conditional covariance matrix h_t to allow for multivariate GARCH-in-mean effects. While the presence of h_t in the mean equation renders the information matrix non-block diagonal between the conditional mean and variance equation parameters, it is important if we want to capture the risk-return tradeoffs in asset returns.

The proposal of this paper is to specify dynamics in the lower triangular Cholesky factor L_t of the conditional covariance matrix H_t such that $H_t = L_t L_t'$. The most general form of such a specification is to let L_t depend on its own past and lagged innovations:

$$\text{vech}(L_t) = \underset{k^*\times 1}{h_t} = c_0 + \sum_{i=1}^{p} A_i h_{t-i} + \sum_{j=1}^{q} B_j \epsilon_{t-j} \tag{2}$$

where $k^* = k(k+1)/2$ and c_0, A_i, B_j are parameter arrays of dimension $k^* \times 1$, $k^* \times k^*$, $k^* \times k$, respectively. This is the analogue of the vech representation given in Engle and

Kroner (1995), where vech$(\cdot)$ is an operator that stacks each column of the lower triangular part of a matrix on top of each other. The main advantage of this vech representation is that, unlike that of Engle and Kroner (1995), the positive definiteness of H_t is ensured without any restrictions on the parameters. As is well known, the Cholesky factor L_t of a positive definite matrix is not uniquely defined. One common way to uniquely identify L_t is to require all of its diagonal elements to be positive. Thus we have the following sufficient condition to identify the cholesky-factor-vech model.

Identification. *Let r denote the rows in h_t that correspond to the diagonal elements in L_t. Then the cholesky-factor-vech model (2) is identified if*

- *Rows r of A_i are positive in columns r and zeros elsewhere.*
- *Rows r of B_j are all zeros.*

Remark. This identification rule, which is simple to impose in practice, restricts the diagonal elements of the Cholesky factor L_t to depend only on its past diagonal elements and not on the lagged innovation vector. While this may appear to be a strong restriction, note that the diagonal elements of H_t are the sum of squares of each row of the lower triangular matrix L_t. Thus, even if the diagonal elements of L_t do not depend on past innovations, the conditional variances (the diagonal elements of H_t) will still depend on past innovations through the off-diagonal elements of L_t, except for the first row.[2] An alternative identified specification that allows the diagonal elements to depend on the lagged innovation vector is

$$h_t = c_0 + \sum_{i=1}^{p} A_i h_{t-i} + \sum_{j=1}^{q} B_j |\epsilon_{t-j}| \tag{3}$$

For this specification, the model is identified with the same restriction on A_i with B_j unrestricted. This specification restricts the effects of past shocks on the conditional variance (the

[2] For a univariate model, the conditional variance will not depend on past innovations and will follow a purely deterministic path.

news impact curve) to be symmetric in the sign of the shocks.

As with the vech specification of Engle and Kroner (1995), the main disadvantage of the cholesky-factor-vech model is the large number of parameters that need to be estimated. With the identification condition imposed, there are $k^* + p(k^2 + (k^* - k)k^*) + q(k^* - k)k$ free parameters in specification (2) and $k^* + p(k^2 + (k^* - k)k^*) + qk^*k$ free parameters in specification (3). Since k^* is quadratic in the number of GARCH components k, the number of parameters in the cholesky-factor-vech model is of order $O(k^4)$. Such a model is not usable in practice except for $k = 2$ or $k = 3$. Thus I consider restrictions that reduce the number of parameters to at most order $O(k^2)$, without overly restricting the cross-element dynamics of the conditional covariances. The natural step is to consider the analogue of the diagonal vech specification

$$L_t = C_0 + \sum_{i=1}^{p} A_i L_{t-i} + \sum_{j=1}^{q} \epsilon_{t-j}\beta_j'$$

where β_j is a $k \times 1$ parameter vector. The problem with this specification is that while restricting C_0 and A_i to be lower triangular ensures the first two terms to be lower triangular, the last term will not be lower triangular. The problem arises because the dimension of the 'square root' of the covariance matrix L_t $(k \times k)$ does not match that of the innovation vector ϵ_t $(k \times 1)$. One solution to this problem is to replace ϵ_t by the 'square root' of $\epsilon_t\epsilon_t'$.

$$L_t = C_0 + \sum_{i=1}^{p} A_i L_{t-i} + \sum_{j=1}^{q} B_j R_{t-j}, \quad R_{t-j}R_{t-j}' = \frac{1}{r}\sum_{i=0}^{r-1} \epsilon_{t-j-i}\epsilon_{t-j-i}' \tag{4}$$

where C_0, A_i, B_j are all $k \times k$ lower triangular parameter matrices. R_{t-j} is the cholesky factor of the sample covariance matrix $S_r = \frac{1}{r}\sum_{i=0}^{r-1} \epsilon_{t-j-i}\epsilon_{t-j-i}'$ of the past r innovations. Since S_r is a rank r matrix, we require $r \geq k$ for S_r to be positive definite and hence R_{t-1} to be well defined. To ensure that L_t is lower triangular with positive diagonals, we impose the following identification conditions.

Identification. *The cholesky-factor model (4) is identified if*

- *$r \geq k$.*

- *C_0, A_i, B_j are all $k \times k$ lower triangular matrices with positive diagonal elements.*

The number of free parameters in (4) with the identification condition imposed is $k^*(1 + p + q)$ and is of order $O(k^2)$. Thus specification (4) reduces the number of parameters while maintaining the lower triangular structure of L_t. However, there are several difficulties with this specification. First, there is a need to select an additional order $r \geq k$. A natural choice is $r = k$ but there is no reason not to select a higher order. Second, to evaluate the likelihood, the cholesky factor R_{t-j} needs to be computed for every observation and this is a costly operation. Third, the presence of the cholesky factor R_{t-j} makes the theoretical analysis of the model difficult. This is because while $E_t[R_{t+h}R'_{t+h}] = H_t = L_t L'_t$, $E_t[R_{t+h}]$ is not generally equal to L_t. Fourth, the extension to allow asymmetric effects in ϵ_t is not straightforward.[3]

An alternative specification that avoids these problems is

$$L_t = C_0 + \sum_{i=1}^{p} A_i L_{t-i} + \sum_{j=1}^{q} B_j |D_{t-j}|, \quad |D_{t-j}| = \begin{pmatrix} |\epsilon_{1,t-j}| & & & 0 \\ & |\epsilon_{2,t-j}| & & \\ & & \ddots & \\ 0 & & & |\epsilon_{k,t-j}| \end{pmatrix} \tag{5}$$

where C_0, A_i, B_j are all $k \times k$ lower triangular parameter matrices with positive diagonal elements. The reason for having $|D_{t-j}|$ a diagonal matrix with absolute values of the innovations ϵ_{t-j} on the main diagonal is to ensure that L_t is uniquely identified. This makes specification (5) restrictive in that the effects of ϵ_{t-j} on H_t is symmetric in the sign of ϵ_{t-j}. As summarized in Hentschel (1995) for univariate models, it is important to have a flexible specification that allows asymmetric effects of the innovations. A parsimonious generalization

[3] An alternative specification that avoids the first three problems is to set R_{t-j} to the lower triangular part of $\epsilon_{t-j}\epsilon'_{t-j}$ rather than its cholesky factor. Under this specification, the covariance matrix H_t will depend on the second and fourth moments of ϵ_t.

of (5) that allows for asymmetric effects is

$$L_t = C_0 + \sum_{i=1}^{p} A_i L_{t-i} + \sum_{j=1}^{q} \big(B_j |D_{t-j}| + F_j (|D_{t-j}| - D_{t-j}) \big) \tag{6}$$

where $C_0,\ A_i,\ B_j,\ F_j$ are all $k \times k$ lower triangular parameter matrices with non-negative diagonal elements. D_{t-j} is a diagonal matrix with ϵ_{t-j} on the main diagonal. $|D_{t-j}| - D_{t-j}$ is a diagonal matrix with non-negative diagonal elements and the model will be identified under the stated parameter restrictions. Note that (5) is a special case of (6) with $F_j = 0$ and this restriction can be tested for the presence of asymmetric effects. The number of parameters in (6) is $k^*(1 + p + 2q)$ and is of order $O(k^2)$.[4]

The dynamics of specifications (4) and (5)–(6) differ in the following sense. In specification (4), the innovation cross-terms $\epsilon_{i,t}\epsilon_{j,t}$ affect the conditional variance through R_t. However, in specifications (5), (6) there is no such interaction effect. $\epsilon_{i,t}$ does have an effect on $H_{ij,t}$ for $i \leq j$ through the coefficient matrices B_j and F_j. Restricting the coefficient matrices $A_i,\ B_j,\ F_j$ to be diagonal will further reduce the number of parameters to order $O(k)$ but will remove any direct spillover effects. Compared to other popular multivariate GARCH models, the dynamics of cholesky factor models have the feature that they not only depend on past cross-product terms of the innovations but also on its levels and interactions with the cholesky factor of past conditional covariances. It remains an empirical question whether such additional dynamics can provide a good fit to real data.

[4] As in Hentschel (1995) one can allow for rotation and shift in the news impact surface. However, since the multivariate GARCH model is already heavily parameterize, I restrict attention to the parsimonious specifcation (6).

3 Estimation and Inference

3.1 Maximum Likelihood Estimation

The gaussian log-likelihood function for (1) is given by

$$\ell = \sum_{t=1}^{T} -\frac{1}{2}\big(k\log(2\pi) + \log|H_t| + \epsilon_t' H_t^{-1}\epsilon_t\big)$$

One should view this objective function as a "surrogate" or quasi-likelihood if, as is usually the case with financial data, the gaussian assumption is considered tenuous. For the cholesky factor model, the log-likelihood evaluation simplifies considerably to

$$\ell = -\frac{Tk}{2}\log(2\pi) - \sum_{t=1}^{T}\sum_{j=1}^{k}\log(L_{jj,t}) - \frac{1}{2}\sum_{t=1}^{T} u_t' u_t \tag{7}$$

where $u_t = L_t^{-1}\epsilon_t$ are the standardized innovations. Note that there is no need to invert a matrix in evaluating this log-likelihood function. Since L_t is lower triangular, the standardized innovations can be obtained by forward substitution of the linear system $L_t u_t = \epsilon_t$. The only difference between specifications (4) and (6) is in the evaluation of L_t. As mentioned above, for specification (4), the computational "bottleneck" is the factorization required to obtain R_{t-j}, which depends only on the mean equation parameters if there are no GARCH-in-mean effects.

To complete the description of the log-likelihood evaluation, one must specify how to set the pre-sample data values in the recursions for L_t. For specification (4), one possibility is to use the values $L_t = R_t = L_0$ for $t < 1$ where

$$\begin{aligned} L_0 &= C_0 + \sum_{i=1}^{p} A_i L_0 + \sum_{j=1}^{q} B_j L_0 \\ &= (I_k - \sum_{i=1}^{p} A_i - \sum_{j=1}^{q} B_j)^{-1} C_0 \end{aligned} \tag{8}$$

This is based on the argument that one can view both L_t and R_t as cholesky factors of the same unconditional innovation covariance matrix. Note that since A_i and B_j are both lower triangular, L_0 can be obtained by forward substitution. For specification (6), one can set

$$
\begin{aligned}
L_0 &= C_0 + \sum_{i=1}^{p} A_i L_0 + \sum_{j=1}^{q} (B_j + F_j)|D_0| \\
&= (I_k - \sum_{i=1}^{p} A_i)^{-1}\big(C_0 + \sum_{j=1}^{q}(B_j + F_j)|D_0|\big) \qquad (9)
\end{aligned}
$$

The pre-sample value $|D_0|$ can be set to its unconditional expectation

$$
E[|D_0|] = \sum_{i=1}^{k} J_{ii} E[|\epsilon|] e_i' = \sum_{i=1}^{k} J_{ii} L_0 E[|u|] e_i' = \sqrt{\frac{2}{\pi}} \sum_{i=1}^{k} J_{ii} L_0 K_i'
$$

where e_i is the i-th column of the identity matrix I_k, J_{ii} is a $k \times k$ matrix of zeros except for the (i,i) element which is one and K_i is a $k \times k$ matrix of zeros except for the i-th column which is all ones. The last equality follows from the assumption that u is an independent standard unit gaussian variate. Then stacking both sides of (6), one can solve for L_0 as

$$
\text{vec}(L_0) = \Big(I_k \otimes (I_k - \textstyle\sum_i A_i) - \sqrt{\frac{2}{\pi}}\big(I_k \otimes \textstyle\sum_j (B_j + F_j)\big)\big(\textstyle\sum_i K_i \otimes J_{ii}\big)\Big)^{-1} \text{vec}(C_0)
$$

Alternatively, one can set $|D_0|_{jj} = \frac{1}{T}\sum_{t=1}^{T} |\widehat{\epsilon}_{j,t}|$ for some preliminary estimates $\widehat{\epsilon}_{j,t}$. For both specifications, a simple alternative is to set L_0 such that $L_0 L_0' = \frac{1}{T}\sum_{t=1}^{T} \widehat{\epsilon}_t \widehat{\epsilon}_t'$.

3.2 Forecasting

The primary use of a GARCH model in risk analysis is to forecast volatility. For the cholesky factor model, an unbiased predictor of the conditional covariance H_t is not available since the recursion is in its cholesky factor L_t and H_t is not linear in L_t. I consider approximate forecasts of H_t which are based on $\widehat{H}_{t+h} = \widehat{L}_{t+h}\widehat{L}_{t+h}'$. For specification (4), one can use the approximation $\widehat{L}_{t+h} = \widehat{R}_{t+h}$ so that

$$\begin{aligned}\widehat{L}_{t+h} &= C_0 + \sum_{i=1}^{p} A_i \widehat{L}_{t+h-i} + \sum_{j=1}^{q} B_j \widehat{R}_{t+h-j} \\ &= C_0 + \sum_{i=1}^{p} A_i \widehat{L}_{t+h-i} + \sum_{j=1}^{q} B_j \widehat{L}_{t+h-j}\end{aligned}$$

For the leading case $p = q = 1$, the h-step ahead forecast can be written as[5]

$$\begin{aligned}\widehat{L}_{t+h} &= C_0 + A_1 \widehat{L}_{t+h-1} + B_1 \widehat{R}_{t+h-1} \\ &= \sum_{i=0}^{h-1} (A_1 + B_1)^i C_0 + (A_1 + B_1)^h L_t\end{aligned}$$

Alternatively, $\widehat{R}_{t+h}$ can be updated based on the approximation

$$\widehat{R}_{t+h-j} \widehat{R}'_{t+h-j} = \frac{1}{r} \sum_{i=0}^{r-1} \widehat{\epsilon_{t+h-j-i} \epsilon'_{t+h-j-i}} = \frac{1}{r} \sum_{i=0}^{r-1} \widehat{L}_{t+h-j-i} \widehat{L}'_{t+h-j-i}$$

For the asymmetric specification (6), $E_t[D_{t+h}] = 0$ for $h > 0$ and we have the forecast recursion

$$\begin{aligned}\widehat{L}_{t+h} &= C_0 + \sum_{i=1}^{p} A_i \widehat{L}_{t+h-i} + \sum_{j=1}^{q} (B_j + F_j) |\widehat{D_{t+h-j}}| \\ |\widehat{D_{t+h-j}}| &= \sqrt{\frac{2}{\pi}} \sum_{i=1}^{k} J_{ii} \widehat{L}_{t+h-j} K'_i\end{aligned}$$

where the second recursion is based on the gaussian assumption $E[|u_i|] = \sqrt{\frac{2}{\pi}}$.

[5]For the leading case $p = q = 1$, the expression for the approximate h-step forecast suggests that we can check for stability of the cholesky factor process by checking the eigenvalues of $A_1 + B_1$. Since both A_1 and B_1 are lower triangular matrices, the eigenvalues of $A_1 + B_1$ are simply its diagonal elements.

3.3 News Impact Surface and Impulse Response

The news impact surface is a mapping from $\epsilon_{i,t}$ to $H_{jk,t+1}$, holding constant all other elements. It is the one-period response of the conditional covariance to an innovation (news) in the conditional mean equation. For non-linear models, especially asymmetric models, the news impact surface gives you can indication of how positive and negative shocks impact risk as the size of the shock varies. As discussed extensively in the vector autoregression literature, the interpretation of a shock in multivariate models is not unambiguous as they are typically contemporaneously correlated. For the cholesky factor model, it is natural to consider the standardized innovation $u_t = L_t^{-1}\epsilon_t$ as the set of independent shocks to the system. Therefore, define the news impact surface

$$NIS(u, \mathcal{I}_t) = H_{t+1}(u, \mathcal{I}_t) - H_{t+1}(0, \mathcal{I}_t) \tag{10}$$

where $\mathcal{I}_t = \{L_t, L_{t-1}, \dots, \epsilon_t, \epsilon_{t-1}, \dots\}$ is the history up to period t. The news impact surface is defined relative to a baseline $H_{t+1}(0, \mathcal{I}_t)$, under which no shock occurs. The news impact surface is history dependent if the effect of $\mathcal{I}_t$ does not cancel out. The news impact surface is asymmetric if $NIS(u, \mathcal{I}_t) \neq NIS(-u, \mathcal{I}_t)$. For the cholesky factor model, the news impact surface is generally history dependent. For the asymmetric specification (6), the news impact surface is asymmetric.

To evaluate the news impact surface, one must select a realization of the history $\mathcal{I}_t$. One possibility is to set L_{t-j} and $\epsilon_{t-j}\epsilon_{t-j}'$ to one of the pre-sample values L_0 as in (8) or (9). Then for specification (4), $H_{t+1}(u, \mathcal{I}_t) = L_{t+1}L_{t+1}'$ where

$$L_{t+1} = C_0 + \sum_{i=1}^{p} A_i L_0 + \sum_{j=2}^{q} B_j L_0 + B_1 R_0$$

$$R_0 R_0' = \frac{1}{r}\big(L_0 u u' L_0' + (r-1) L_0 L_0'\big) = L_0 L_0' + \frac{1}{r} L_0 (u u' - I_k) L_0'$$

For the asymmetric specification (6),

$$L_{t+1} = C_0 + \sum_{i=1}^{p} A_i L_0 + \sum_{j=2}^{q} (B_j + F_j)|D_0| + (B_1 + F_1)|D_u| - F_1 D_u$$

where D_u is a $k \times k$ diagonal matrix with u on its main diagonal.

4 Comparison with other multivariate GARCH models

5 Empirical Illustration

As an illustration of cholesky factor garch modeling, I fit a trivariate garch model to weekly returns from three market indices, UK FTSE 100, US S&P 500, and Nikkei 225 (Japan). The data cover the period from April 1984 to December 2002 for 967 observations.[6] The weekly return is computed as the log difference of the weekly index (closing price on the last trading day of the week). The weekly frequency was chosen partly to bypass the issue of handling weekend effects in daily data and to use daily data to obtain integrated volatility measures at the weekly frequency (Ledoit et al. 2003). The time series of the weekly returns data are plotted in Figure 1.

5.1 Estimation

The first step in estimating cholesky factor GARCH models is to specify the ordering of the variables. If there is no prior information regarding the ordering of the variables, I propose to select an ordering based on the log-likelihood value. For a k-dimensional system, one estimates all $k!$ possible orderings and selects the ordering with the highest log-likelihood value. I restrict my search to the 'canonical' model with only a constant μ in the mean

[6] The choice of the sample period was constrained by data availability. The data were downloaded from Yahoo! Finance at *http://finance.yahoo.com* on 12/28/2002. The longest sample available at the time for the FTSE 100 index started in April 1984.

equation and lag orders $p = q = 1$ and $r = k$ for the covariance equation. For fixed p, q, r, the number of estimated parameters do not change for different cholesky orderings, so the log-likelihood criterion is equivalent to selection based on information criteria.

To obtain maximum likelihood estimates for each cholesky ordering, I numerically maximize the log-likelihood function (7). The pre-sample values are set to $L_0 L_0' = \frac{1}{T}\sum_{t=1}^{T} \widehat{\epsilon}_t \widehat{\epsilon}_t'$ and $|D_0|_{jj} = \frac{1}{T}\sum_{t=1}^{T} |\widehat{\epsilon}_{j,t}|$ where $\widehat{\epsilon}_t = y_t - \widehat{\mu}$. There are two important factors for estimation of cholesky factor models: how to set the starting parameter values and how to impose the non-negativity constraints on the diagonals of the lower triangular parameter matrices. Since there does not appear to be an obvious choice for the starting parameter values, I arbitrarily set them to some fixed values. For this data set, the following set of starting values were used: $\mu = 0$ for the mean equation and $C_0 = 0.1 I_k$, $A_i = 0.8 I_k$, $B_j = F_j = 0.1 I_k$. I have observed, however, that the optimization routine fails to converge for certain cholesky orderings. For such difficult cases, I have perturbed the starting values until convergence is obtained.

To impose the non-negativity constraints, I have tried two approaches. One approach is to use a transformation of unconstrained parameters so that the diagonals remain non-negative. For this approach, I use the exponential function for the transformation together with the unconstrained trust region BFGS code by Gay (1983). There are a number of cases in which the converged parameter values on the main diagonal are near the boundary of zero. To check whether these cases arise from the transformation, I have also tried a constrained optimizer. The code I used is the limited memory BFGS for bound constrained problems (L-BFGS-B) developed by Zhu, Byrd, Lu and Nocedal (1997).

To determine the cholesky ordering, I estimated all cholesky orderings using both algorithms.[7] For model (4), L-BFGS-B performed better (i.e., generally returned a higher log-likelihood from the same starting values) than the transformation method based on the

[7] The codes for both algorithms were obtained from *www.netlib.org/toms*. For the unconstrained trust region BFGS, I use subroutine `SMSNO` (numerical gradients) with the default convergence tolerance. For L-BFGS-B, I numerically evaluate the gradients by central differences and set the two tolerance inputs for termination test to `factr` $= 10^1$ (for 'extremely high accuracy') and `pgtol` $= 10^{-7}$.

trust region BFGS.[8] The results are reported in Table 1. For model (6), the transformation method based on the trust region BFGS performed better than L-BFGS-B.[9] The results are reported in Table 2. The estimated parameters for the ordering selected for each model are reported in Tables 3–4. I do not report standard errors for each individual parameter for two reasons. First, it is difficult to interpret each individual parameter separately and hence there is no obvious hypothesis of interest. Second, since the parameters on the main diagonal are restricted to be positive, an issue arises when the parameter is near or at the boundary. This is the case for C_0 in Table 3. This problem has been recently addressed by Andrews (2001) and I leave a proper treatment of this problem for the cholesky factor model for future work. In Tables 3–4, I report Wald statistics for the joint significance of the *strict* lower part (which is unrestricted) in each triangular parameter matrix.[10] These statistics may provide an indication of the significance of the cross-element effects in the conditional variance dynamics. An exception is for the parameter matrix F_j in specification (6). The hypothesis $F_j = 0$ is a test for the presence of asymmetric effects in the news impact surface. Thus for F_j, I report Wald statistics for the hypothesis $F_j = 0$ (ignoring the fact that the diagonals were constrained to be positive). Table 4 indicates that there is statistically significant asymmetry in the conditional variance dynamics for this data set. Another interesting feature of the estimates in Tables 3–4 is that the parameter A_1 in both models have large eigenvalues compared to those of B_1 and F_1.[11] This may be the analogue of the empirical regularity found in the univariate GARCH literature in which the GARCH process for asset returns is typically highly persistent, or nearly integrated.

Figure 2 plots the in-sample fitted conditional covariance series (elements of H_t) using the parameter estimates reported in Tables 3–4. The fitted covariance series obtained from

[8] `SMSNO` returns 'singular convergence' for two out of the six cases.

[9] L-BFGS-B returns 'abnormal termination in line search' for two out of the six cases.

[10] The Wald statistics were computed using an estimate of the information matrix based on the outer-product of the (numerical) gradients.

[11] Recall that the eigenvalues of a lower triangular matrix are given by its diagonal elements.

specification (4) is much smoother than that obtained from the asymmetric specification (6). This is not surprising given that the innovation second moments in specification (4) is averaged over $r \geq k$ periods. The off-diagonal elements in Figure 2 remain mostly positive for the entire sample period, indicating strong positive co-movement among the three major stock exchanges. In particular, all conditional covariance series have a large spike around observation 200 (the stock market crash in October 1987).

Figure 3 plots the news impact curves for an orthogonalized shock in the series that comes first in the cholesky ordering. Because the two specifications are estimated with a different cholesky ordering, these correspond to shocks in a different series. They are plotted together to indicate the relative size of the news impact curves. The size of the impulse on the horizontal axis is roughly in standard deviation units of the original series. Thus the figure plots the response to up to five standard deviation shocks. There are several interesting features in these news impact curves. First, the conditional covariance in specification (4) is much less responsive to a shock compared to specification (6). This is consistent with the conditional covariance series from specification (4) being much smoother than that from specification (6) in Figure 2. Second, the asymmetry in the response to positive and negative shocks in specification (6) is quite pronounced. All variance and covariance series increase in response to a negative shock (bad news), while most decrease in response to a positive shock (good news).[12] Recall that, by construction, the response in specification (4) is symmetric to positive and negative shocks.

6 Concluding Remarks

In this paper I have proposed a new class of multivariate GARCH models that specifies the dynamics of the cholesky factor of the conditional covariance matrix. The model can be

[12] Recall that our definition (10) of the news impact curve is relative to the base line case of no shock. Thus the *relative* response in the variance can become negative even though the variance itself cannot be negative.

thought of as a middle ground between a fully speficied multivariate GARCH model that is over-parameterized and a diagonal model that overly restricts cross-element dynamics among the elements of the conditional covariance matrix. It is hoped that the model will find empirical use in applications such as risk management.

There are several theoretical loose ends that need to be addressed in future work. Among others, these include the stationarity conditions for the cholesky factor model and inference when some of the parameters on the main diagonal are near or at the boundary of zero.

Order	Log-likelihood	Feval	Active
JP,UK,US	7049.848635	16633	2
JP,US,UK	7052.547459	4627	2
UK,JP,US	7053.801803*	17805	3
UK,US,JP	7047.040837	8386	2
US,JP,UK	7051.338951	5941	1
US,UK,JP	7042.142284	4319	1

Table 1: Determination of cholesky ordering for model (4) with $p = q = 1$ and $r = k = 3$ for a total of 21 parameters. Feval is the number of function evaluations. Active is the number of binding constraints at the optimum in the L-BFGS-B algorithm. For all cases, convergence was achieved based on the relative reduction of the objective function.

Order	Log-likelihood	Feval	Geval
JP,UK,US	7136.921945	353	8893
JP,US,UK	7149.481680	317	7849
UK,JP,US	7150.350339	333	7814
UK,US,JP	7150.983160	321	7441
US,JP,UK	7163.825200	313	7225
US,UK,JP	7171.612269*	351	8491

Table 2: Determination of cholesky ordering for model (6) with $p = q = 1$ for a total of 27 parameters. Feval is the number of function evaluations, excluding those made only for computing the gradients. Geval is the number of function evaluations made only for computing the gradients. The total number of function evaluations is the sum of Feval and Geval. The diagonals were constrained to be non-negative by the exponential transformation. For all cases, `SMSNO` returned relative function convergence.

μ	C_0			A_1			B_1		
0.0018	0.0000			0.8759			0.0899		
0.0012	0.0007	0.0000		0.0777	0.8021		−0.0818	0.1461	
0.0024	−0.0004	0.0002	0.0000	−0.0516	−0.0488	0.9781	0.0534	0.0294	0.0166
	$\chi^2(3) = 3.4590$			$\chi^2(3) = 12.1774$			$\chi^2(3) = 18.3441$		
	$p = 0.3261$			$p = 0.0068$			$p = 0.0004$		

Table 3: Maximum likelihood estimates for model (4) with ordering UK, JP, US. $\chi^2(3)$ is the Wald statistic for the joint significance of the parameters below the main diagonal and p is the corresponding p-value.

μ	C_0			A_1			B_1			F_1		
0.0017	0.0010			0.8621			0.0521			0.0640		
0.0012	−0.0043	0.0083		0.5304	0.4411		−0.0667	0.0544		0.0535	0.0457	
0.0002	−0.0044	0.0095	0.0015	0.7008	−0.6195	0.8731	−0.1951	0.0508	0.0159	0.0772	0.0661	0.0619
	$\chi^2(3) = 14.5132$			$\chi^2(3) = 16.2434$			$\chi^2(3) = 12.4701$			$\chi^2(6) = 87.9048$		
	$p = 0.0023$			$p = 0.0010$			$p = 0.0059$			$p = 0.0000$		

Table 4: Maximum likelihood estimates for model (6) with ordering US, UK, JP. $\chi^2(3)$ is the Wald statistic for the joint significance of the parameters below the main diagonal. $\chi^2(6)$ is the Wald statistic for the joint significance of the full lower triangular parameter matrix. p is the p-value for the corresponding Wald statistic.

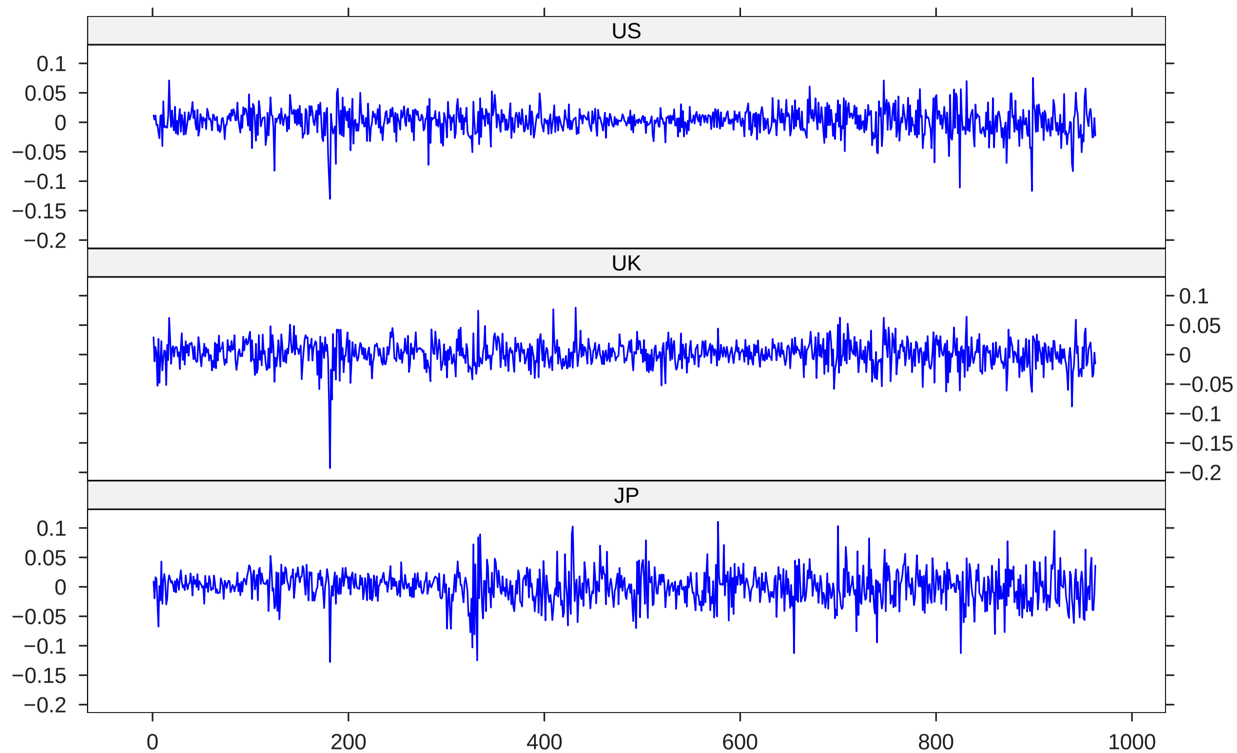

Figure 1: Weekly return series (April 1984 to December 2002)

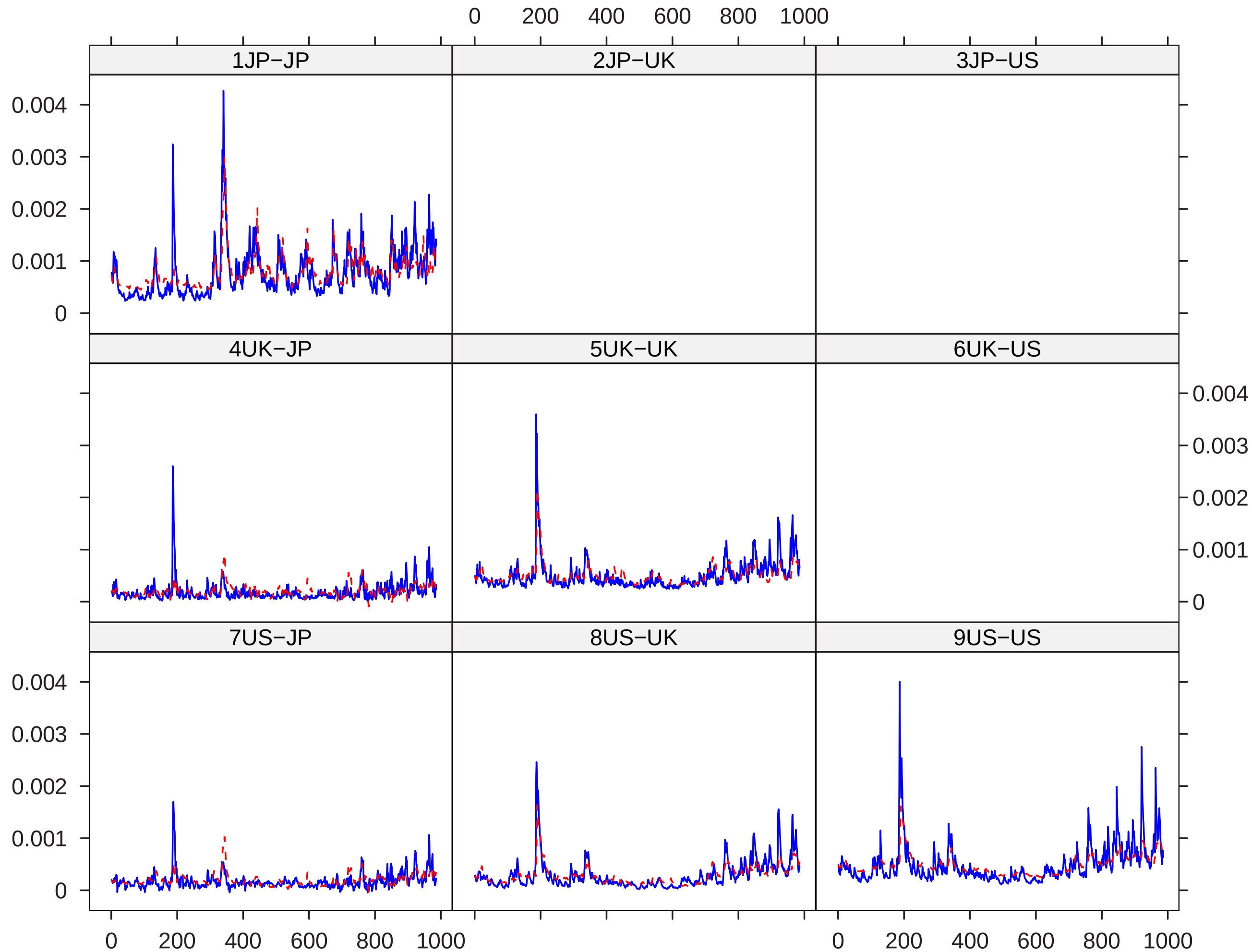

Figure 2: Filtered conditional covariance series. The red dashed line is from specification (4), while the blue solid line is from the asymmetric specification (6). The entry UK-JP indicates the conditional covariance between the UK and JP return series and so on.

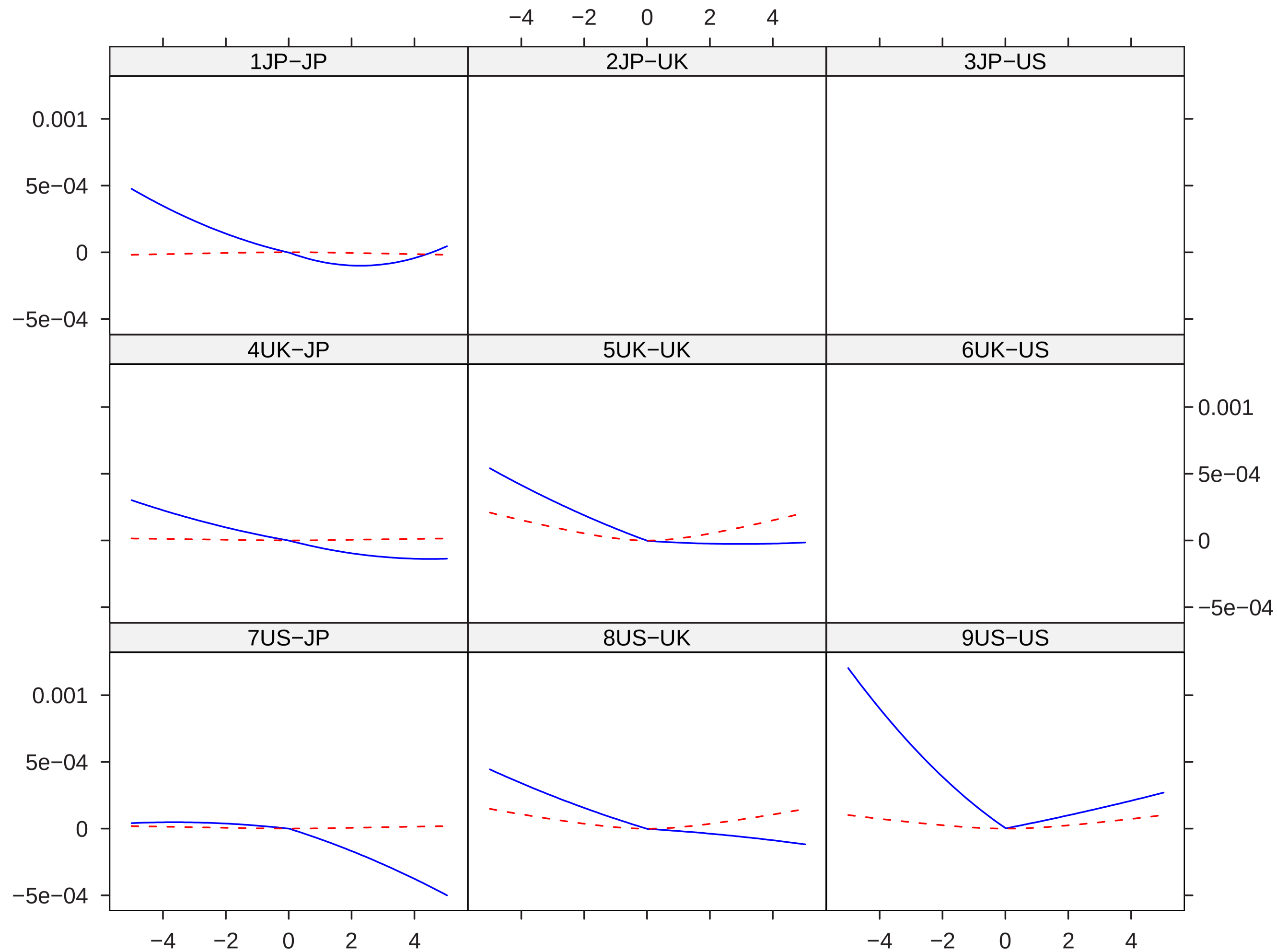

Figure 3: News impact curves. These are the responses to an orthogonalized shock to the variable that comes first in the cholesky ordering. The red dashed line is for specification (4) with ordering UK, JP, US. The blue solid line is for specification (6) with ordering US, UK, JP.

References

Andrews, Donald W. K., "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 2001, *69*, 683–734.

Engle, Robert F., "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 1982, *50*, 987–1007.

——, "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business & Economic Statistics*, 2002, *20*, 339–350.

—— **and Kenneth F. Kroner**, "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 1995, *11*, 122–150.

Gay, David M., "Algorithm 611: Subroutines for unconstrained minimization using a model/trust-region approach," *ACM Transactions on Mathematical Software*, 1983, *9*, 503–524.

Hentschel, Ludger, "All in the family: Nesting symmetric and asymmetric GARCH models," *Journal of Financial Economics*, 1995, *39*, 71–104.

Kroner, Kenneth F. and Victor K. Ng, "Modeling Asymmetric Comovements of Asset Returns," *Review of Financial Studies*, 1998, *11*, 817–844.

Ledoit, Olivier, Pedro Santa-Clara, and Michael Wolf, "Flexible Multivariate GARCH Modeling with an Application to International Stock Markets," 2003. forthcoming Review of Economics and Statistics.

Nelson, Daniel B., "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 1991, *59*, 347–370.

Tse, Y. K. and Albert K. C. Tsui, "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations," *Journal of Business & Economic Statistics*, 2002, *20*, 352–362.

Weide, Roy Van Der, "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model," *Journal of Applied Econometrics*, 2002, *17*, 549–564.

Zhu, Ciyou, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software*, 1997, *23*, 550–560.