

MPRA

Munich Personal RePEc Archive

A Multivariate Generalized Orthogonal Factor GARCH Model

Lanne, Markku and Saikkonen, Pentti
University of Helsinki, HECER

May 2005

Online at <http://mpra.ub.uni-muenchen.de/23714/>
MPRA Paper No. 23714, posted 07. July 2010 / 09:10



Helsinki
Center
of
Economic
Research

Discussion Papers

A Multivariate Generalized Orthogonal Factor GARCH Model

Markku Lanne
University of Jyväskylä, RUESG and HECER

and

Pentti Saikkonen
University of Helsinki, RUESG and HECER

Discussion Paper No. 63
May 2005

ISSN 1795-0562

A Multivariate Generalized Orthogonal Factor GARCH Model*

Abstract

The paper studies a factor GARCH model and develops test procedures which can be used to test the number of factors needed to model the conditional heteroskedasticity in the considered time series vector. Assuming normally distributed errors the parameters of the model can be straightforwardly estimated by the method of maximum likelihood. Inefficient but computationally simple preliminary estimates are first obtained and used as initial values to maximize the likelihood function. Maximum likelihood estimation with nonnormal errors is also straightforward. Motivated by the empirical application of the paper a mixture of normal distributions is considered. An interesting feature of the implied factor GARCH model is that some parameters of the conditional covariance matrix which are not identifiable in the case of normal errors become identifiable when the mixture distribution is used. As an empirical example we consider a system of four exchange rate return series.

JEL Classification: C32, C51, F31.

Keywords: Multivariate GARCH model, mixture of normal distributions, exchange rate.

Markku Lanne

University of Jyväskylä
School of Business and Economics
P.O. Box 35 (MaE)
FIN-40014 University of Jyväskylä
FINLAND

e-mail: markku.lanne@econ.jyu.fi

Pentti Saikkonen

University of Helsinki
Department of Statistics
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: pentti.saikkonen@helsinki.fi

* We are grateful to the Research Unit of Economic Structures and Growth (RUESG) in the University of Helsinki and the Yrjö Jahnsson Foundation for financial support. The second author also acknowledges financial support from the Academy of Finland.

1 Introduction

Even though the literature on volatility models is huge, only a relatively small fraction of it is devoted to developing and applying multivariate GARCH models. This is not due to the lack of interest or importance because understanding the comovements of financial returns is at the heart of empirical finance and models capable of describing the joint behavior of asset prices are in great demand in areas such as asset allocation and risk management. It is rather the practical problems inherent in most multivariate GARCH models that have retarded their widespread use. The problems with estimation are mainly caused by the typically quite rapid increase in the number of parameters with the dimension of the system and the restrictions required by the positive definiteness of the conditional covariance matrix. As one solution to these problems, several factor and orthogonal models have been introduced in the literature.

Factor models assume the data to be generated by a set of uncorrelated components and in orthogonal models these components are also (a subset of) the principal components of the data. According to the taxonomy of Bauwens *et al.* (2003), in orthogonal models the matrix by which the data are obtained from the components must be orthogonal, while in generalized orthogonal models its invertibility suffices. The notion of a factor model typically encompasses the idea that there are a relatively small number of common underlying variables, whereas (generalized) orthogonal models (e.g. van der Weide, 2002, Vrontos *et al.*, 2003) commonly do not have a reduced number of principal components. Thus, (generalized) orthogonal models are rather restrictive for financial data in that they allow for no idiosyncratic shocks.¹

In this paper we introduce a new kind of generalized orthogonal GARCH model that allows for a reduced number of conditionally heteroskedastic factors and, hence, idiosyncratic shocks. Therefore we call our model a generalized orthogonal factor

¹Orthogonal models with fewer principal components than time series have also been presented (see e.g. Alexander, 2001), but they are hampered by the fact that the conditional covariance matrix has a reduced rank.

GARCH model. Our model is related to the factor GARCH model of Engle *et al.* (1990), but it is more parsimonious and easier to estimate. Gaussian maximum likelihood (ML) estimates can be straightforwardly obtained and likelihood functions based on other distributions, such as the (multivariate) t distribution, can also be readily formulated. This is illustrated by the empirical application of the paper where, instead of the commonly used t distribution, a mixture of normal distributions is more appropriate and, therefore, applied. Interestingly, some parameters of the conditional covariance matrix, which are not identifiable in the Gaussian model, become identifiable in a model based on a mixture of normal distributions. In practice the first step of applying any factor GARCH model consists of selecting the unknown number of conditionally heteroskedastic common factors. In order to facilitate this selection two tests are developed in the paper.

The remainder of the paper is organized as follows. The model is introduced in Section 2. Parameter estimation and statistical inference are discussed in Section 3 assuming first a Gaussian likelihood and providing thereafter an extension to the case of a mixture of normal distributions. Section 4 develops tests for checking the number of conditionally heteroskedastic common factors. Section 5 presents an application to a data set of exchange rate return series. Finally, Section 6 concludes.

2 Model

Consider an n -dimensional time series y_t ($t = 1, 2, \dots$) generated by

$$y_t = WH_t^{1/2}\varepsilon_t, \quad (1)$$

where W ($n \times n$) is a nonsingular parameter matrix, H_t ($n \times n$) is a stochastic (a.s.) positive definite diagonal matrix measurable with respect to the information set $\mathcal{F}_{t-1} = \{y_{t-j}, j \geq 1\}$, and ε_t ($n \times 1$) is a sequence of independent and identically distributed random vectors with zero mean and identity covariance matrix or, briefly, $\varepsilon_t \sim \text{i.i.d.}(0, I_n)$. It is also assumed that the distribution of ε_t has a (Lebesgue) density and that ε_t is independent of \mathcal{F}_{t-1} for all $t \geq 1$. Thus, $\{y_t, \mathcal{F}_t\}$ is a martingale

difference sequence and, given \mathcal{F}_{t-1} , the conditional covariance matrix of y_t is

$$\text{Cov}_{t-1}(y_t) = WH_tW' \stackrel{\text{def}}{=} \Sigma_t. \quad (2)$$

Equations (1) and (2) specify a general model for conditional heteroskedasticity of the time series vector y_t . To make this general model feasible in practice, the dependence of the diagonal elements of the matrix H_t on past values of y_t has to be specified.

A model similar to that defined by equations (1) and (2) has recently been considered by van der Weide (2002) and Vrontos *et al.* (2003) (see also Klaassen (2000) and Alexander (2001)). These authors do not explicitly discuss the case where some of the diagonal elements of the matrix H_t are constant, that is, independent of t . Because we feel that it is of interest to allow for this possibility we shall assume that

$$H_t = \text{diag}[V_t : I_{n-r}], \quad (3)$$

where $V_t = \text{diag}[v_{1t} \cdots v_{rt}]$ and $0 \leq r \leq n$. Thus, we have to specify the dependence of v_{it} ($i = 1, \dots, r$) on past values of y_t . Unless otherwise stated, it will henceforth be assumed that $r > 0$ so that the quantities v_{1t}, \dots, v_{rt} are all stochastic, and hence, time varying. In order to motivate the employed specification, write $W = [W_1 : W_2]$ and $W^{-1} = B' = [B_1 : B_2]'$ where the matrices W_1 and W_2 are of orders $n \times r$ and $n \times (n - r)$, respectively, and the matrix B' is partitioned conformably. From equation (1) one then obtains

$$B_1'y_t = V_t^{1/2} \varepsilon_{1t} \quad (4)$$

and

$$B_2'y_t = \varepsilon_{2t}, \quad (5)$$

where $\varepsilon_t = [\varepsilon'_{1t} \ \varepsilon'_{2t}]'$ is partitioned in an obvious way. Thus, the components of $B_1'y_t$ are uncorrelated univariate conditionally heteroskedastic processes whereas $B_2'y_t \sim \text{i.i.d.}(0, I_{n-r})$. As in van der Weide (2002) and Vrontos *et al.* (2003) we specify the diagonal elements of V_t as standard GARCH processes driven by squared lagged values of the components of $B_1'y_t$. For ease of exposition, we assume the GARCH(1,1) models

$$v_{it} = (1 - \alpha_i - \beta_i) + \beta_i v_{i,t-1} + \alpha_i (b'_{1i} y_{t-1})^2, \quad i = 1, \dots, r, \quad (6)$$

where b'_{1i} signifies the i th row of the matrix B'_1 and the other parameters satisfy $\alpha_i > 0$, $\beta_i \geq 0$ and $\beta_i + \alpha_i < 1$ for all $i = 1, \dots, r$. The inequality $\alpha_i > 0$ is due the assumption that v_{it} does not reduce to constant.

The imposed parameter restrictions imply that, under mild conditions about the density of ε_t , the vector process y_t defined by equations (1), (3) and (6) is strictly stationary and ergodic and also second order stationary (see Engle and Kroner (1995) and Comte and Lieberman (2003)). Note also that the intercept terms in (6) are normalized in such a way that the components of $B'_1 y_t$ have unit unconditional variance. These normalizations will be convenient in our subsequent developments. Combined with the assumption that the variances of the components of ε_{1t} are normalized to unity they ensure that the parameters in equations (4) and (6) are unique up to multiplying the columns of B_1 by minus one.

Because equation (5) can be premultiplied by any orthogonal $(n - r) \times (n - r)$ matrix without changing the second order properties of the model the parameter matrix B_2 is not identifiable without further assumptions. The special case $n - r = 1$ is an exception in that identifiability obtains up to multiplying (the vector) B_2 by minus one. Similar conclusions, of course, hold for the parameter matrix W_2 . In this section we shall only rely on second order properties of the series y_t and, therefore, identifiability of the parameter matrices B_2 and W_2 is not achieved. This means, in particular, that part of the subsequent discussion is only relevant when the error term ε_t is normally distributed or, more generally, has a spherically symmetric distribution such as a t distribution.² However, it will be seen later that with other distributions even the parameter matrices B_2 and W_2 may become identifiable (up to multiplying the columns by minus one).

As in van der Weide (2002), we factorize the parameter matrix W but, instead of

²The distribution of a random vector x is spherically symmetric if x and Ox have the same distribution for any conformable orthogonal matrix O .

the singular value decomposition used in that paper, we use the polar decomposition

$$W = CR, \tag{7}$$

where C ($n \times n$) is a (symmetric) positive definite matrix and R ($n \times n$) is an orthogonal matrix. The nonsingularity of the matrix W ensures the positive definiteness of the matrix C and uniqueness of the decomposition (see Horn and Johnson (1985), p. 413). The normalizations used in (6) imply that $E(V_t) = I_{n-r}$. Thus, it follows from equations (2) and (3) that the covariance matrix of y_t satisfies

$$\text{Cov}(y_t) \stackrel{def}{=} \Omega = WW' = CC'. \tag{8}$$

Let $\Omega = P\Lambda P'$ be the spectral decomposition of the covariance matrix Ω so that Λ is a diagonal matrix containing the eigenvalues of Ω on the diagonal and P is an orthogonal matrix of corresponding eigenvectors. Thus, the matrix C is the unique (positive definite) square root of the covariance matrix Ω , that is, $C = P\Lambda^{1/2}P' = \Omega^{1/2}$, whereas $R = \Omega^{-1/2}W$. Because the theoretical covariance matrix Ω can be consistently estimated by the sample covariance matrix the continuity of the mapping $\Omega \rightarrow \Omega^{1/2}$ implies that the matrix C can be consistently estimated by using the spectral decomposition of the sample covariance matrix. Part of the parameter matrix W can therefore be consistently estimated in a very simple way.

The polar decomposition assumed in (7) is not the only possibility one could entertain. As already mentioned, van der Weide (2002) used the singular value decomposition and defined the matrix C as $C = P\Lambda^{1/2}$ with P and Λ as above. Unlike the polar decomposition, this factorization is unique only when the eigenvalues of Ω are distinct. Uniqueness is a useful property if the components of the factorization are estimated simultaneously. Instead of the polar decomposition, uniqueness can be achieved by an appropriate version of the QR decomposition (see Horn and Johnson (1985), p. 112-114). In what follows, the QR decomposition can be used instead of the polar decomposition. Then C is a lower triangular matrix which can be obtained from the Cholesky decomposition of the covariance matrix Ω .

In the model of Vrontos *et al.* (2003) the parameter matrix W is assumed to be lower triangular. This is a simplifying assumption which is not without loss of generality because, in general, the orthogonal matrix R cannot be dropped from (7) and because the application of the Cholesky decomposition to the conditional covariance matrix Σ_t does generally not imply a factorization of the form (2) with W lower triangular.³ In addition to being restrictive, the lower triangularity assumption of the matrix W involves the difficulty that a certain order is assumed for the components of the vector y_t .

By the above discussion we can write equation (1) as

$$y_t = W_1 V_t^{1/2} \varepsilon_{1t} + W_2 \varepsilon_{2t}. \quad (9)$$

This means that the model can be interpreted as a factor GARCH model in which the conditional heteroskedasticity is due to r common factors, the components of the vector $V_t^{1/2} \varepsilon_{1t}$. Alternatively, the model implies the existence of $n - r$ linearly independent homoskedastic linear combinations of the process y_t given by equation (5). Now, partition $R = [R_1 : R_2]$ conformably with the partition of W so that $W_1 = CR_1$ and $W_2 = CR_2$. Using equation (8) and the fact that $R_1 R_1' + R_2 R_2' = I_n$ it is easy to check that the conditional covariance matrix (2) can be written as

$$\Sigma_t = \Omega + CR_1 (V_t - I_r) R_1' C' \quad (10)$$

or, alternatively, as

$$\Sigma_t = \Omega^* + \sum_{i=1}^r v_{it} w_{1i} w_{1i}', \quad (11)$$

where $\Omega^* = W_2 W_2'$ and w_{1i} signifies the i th column of the matrix W_1 .

Equation (11) shows that in our model the conditional covariance matrix is similar to that in the factor GARCH model of Engle *et al.* (1990). A major difference between the conditional covariance matrices arising from these two models lies in their

³A multivariate GARCH model based on the Cholesky decomposition of the conditional covariance matrix has recently been considered by Kawakatsu (2003).

constant terms. In our model the constant term is closely related to the unconditional covariance matrix Ω whereas in the model of Engle *et al.* (1990) the constant term has no particular role. Note also that in the representation (11) the constant term Ω^* is singular and its rank is directly related to the number of common factors. In the model of Engle *et al.* (1990) the corresponding constant term is assumed to be positive semidefinite and it has typically been treated as a positive definite matrix. This implies that our model is generally more parsimonious than the model of Engle *et al.* (1990). Another difference between the two models concerns the correlation of the common factors. As typical in factor models, our model assumes uncorrelated common factors whereas in the model of Engle *et al.* (1990) the common factors are generally correlated although their conditional covariance is constant. These differences have important implications on the estimation of the parameters of the model. Above we already discussed the (preliminary) estimation of the covariance matrix Ω by using its sample analog and in the next section it will be seen that convenient (preliminary) estimators can also be obtained for other parameters of our model. Finally, note that Engle *et al.* (1990) list some attractive properties of the conditional covariance structure implied by their model. These properties, which include the positive (semi)definiteness of the conditional covariance matrix, are also shared by the conditional covariance matrix arising from our model.

As with the factor GARCH model of Engle *et al.* (1990), our model can also be viewed as a special case of the BEKK model of Engle and Kroner (1995). Specifically, note that from equations (2) and (3) it follows that $v_{it} = b'_{1i}\Sigma_t b_{1i}$ which in conjunction with equations (6) and (11) yields

$$\Sigma_t = \Omega^{**} + \sum_{i=1}^r (\beta_i b'_{1i}\Sigma_{t-1} b_{1i} + \alpha_i b'_{1i} y_{t-1} y_{t-1}' b_{1i}) w_{1i} w'_{1i},$$

where $\Omega^{**} = \Omega^* + \sum_{i=1}^r (1 - \alpha_i - \beta_i) w_{1i} w'_{1i}$. This shows the aforementioned connection with the BEKK model. An implication of this is that the asymptotic estimation results obtained by Comte and Lieberman (2003) for Gaussian ML estimators in the BEKK model can be applied to our model.

3 Parameter Estimation

3.1 Gaussian ML Estimation

The parameters of the model introduced in the previous section can be estimated by Gaussian ML. Thus, assume that, conditional on \mathcal{F}_{t-1} , the distribution of y_t is normal with zero mean and covariance matrix Σ_t . The related conditional density is

$$f_{t-1}(y_t) = (2\pi)^{-n/2} \det(\Sigma_t)^{-1/2} \exp \left\{ -\frac{1}{2} y_t' \Sigma_t^{-1} y_t \right\}.$$

From equations (2), (3) and (8) it follows that $\det(\Sigma_t) = \det(\Omega) v_{1t} \cdots v_{rt}$ and $y_t' \Sigma_t^{-1} y_t = y_t' B H_t^{-1} B' y_t = y_t' B_1 V_t^{-1} B_1' y_t + y_t' B_2 B_2' y_t$ where $B_i = C^{-1'} R_i$ ($i = 1, 2$) by equation (7). Thus, because $R_1 R_1' + R_2 R_2' = I_n$ and $\Omega = C C'$, we can write $y_t' \Sigma_t^{-1} y_t = y_t' C^{-1'} R_1 (V_t^{-1} - I_r) R_1' C^{-1} y_t + y_t' \Omega^{-1} y_t$ and, furthermore,

$$\begin{aligned} f_{t-1}(y_t) &= (2\pi)^{-n/2} \det(\Omega)^{-1/2} \exp \left\{ -\frac{1}{2} y_t' \Omega^{-1} y_t \right\} \\ &\quad \times \prod_{i=1}^r v_{it}^{-1/2} \exp \left\{ -\frac{1}{2} y_t' C^{-1'} R_1 (V_t^{-1} - I_r) R_1' C^{-1} y_t \right\}. \end{aligned} \quad (12)$$

Equation (6) shows that, in addition to the GARCH parameters α_i and β_i ($i = 1, \dots, r$), the diagonal matrix V_t depends on the parameter matrix B_1 or on $C^{-1'} R_1$. Thus, since $\Omega = C C'$, the Gaussian likelihood depends on the parameters C , R_1 , α_i and β_i ($i = 1, \dots, r$). Instead of C it appears to be convenient to use its inverse on which the likelihood function only depends.

Assume that we have an observed data set y_0, \dots, y_T . The first observation y_0 is used as an initial value in the GARCH models (6) and the subsequent analysis will be conditional on it and also on v_{i0} ($i = 1, \dots, r$). Let ρ_{1i} signify the i th column of the matrix R_1 and set $\Phi = C^{-1}$ and $\delta = [\delta_1' \cdots \delta_r']'$ where $\delta_i = [\alpha_i \ \beta_i]'$ ($i = 1, \dots, r$). From the expression of the conditional density function (12) and the subsequent discussion it follows that the Gaussian log-likelihood function (without an additive constant)

can be written as

$$\begin{aligned}
l_T(\Phi, R_1, \delta) &= T \log \det(\Phi) - \frac{1}{2} \sum_{t=1}^T y_t' \Phi' \Phi y_t \\
&\quad - \sum_{i=1}^r \left(\frac{1}{2} \sum_{t=1}^T \log v_{it} + \frac{1}{2} \sum_{t=1}^T (v_{it}^{-1} - 1) (\rho_{1i}' \Phi y_t)^2 \right) \\
&\stackrel{def}{=} l_{0T}(\Phi) + \sum_{i=1}^r l_{iT}(\Phi, \rho_{1i}, \delta_i).
\end{aligned} \tag{13}$$

Here we have also made use of the fact that, in addition to the parameter δ_i , the conditional variance v_{it} only depends on the parameters ρ_{1i} and Φ (see equation (6)) and note that $b_{1i}' y_{t-1} = \rho_{1i}' \Phi y_{t-1}$. We need to maximize $l_T(\Phi, R_1, \delta)$ subject to the constraints

$$\rho_{1i}' \rho_{1j} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}. \tag{14}$$

The required maximization is obviously a highly nonlinear problem. Good initial values are therefore desirable for successful numerical optimization of the likelihood function.

As discussed in the previous section, a consistent estimator of Ω can be obtained from the (uncentered) sample covariance matrix

$$\tilde{\Omega} = T^{-1} \sum_{t=1}^T y_t y_t'.$$

A consistent estimator of the matrix Φ , denoted by $\tilde{\Phi}$, can be obtained from $\tilde{\Omega}$. The estimator $\tilde{\Phi}$ can be either the (unique) square root of $\tilde{\Omega}^{-1}$ or its Cholesky factor depending on which one of the two alternative specifications is adopted. Note that the estimator $\tilde{\Phi}$ can be obtained by maximizing the first component of the log-likelihood function, that is, $l_{0T}(\Phi)$. Although consistent, the estimator $\tilde{\Phi}$ is therefore inefficient because it ignores the second component of the log-likelihood function which is due to conditional heteroskedasticity (cf. van der Weide (2002)).

Obtaining initial estimates for the parameters R_1 and δ is more complicated. However, a simplifying fact is that the estimation of the parameters ρ_{1i} and δ_i can be

carried out separately for $i = 1, \dots, r$. Define the set $\Delta_1 = \{(\rho_{11}, \delta_1) : \rho'_{11}\rho_{11} = 1\}$ and, for $i = 2, \dots, r$, $\Delta_i = \{(\rho_{1i}, \delta_i) : \rho'_{1i}\rho_{1i} = 1, \rho'_{1i}\rho_{1j} = 0, 1 \leq j \leq i-1\}$. Then consider estimators defined by

$$(\tilde{\rho}_{11}, \tilde{\delta}_1) = \arg \max_{(\rho_{11}, \delta_1) \in \Delta_1} l_{1T}(\tilde{\Phi}, \rho_{11}, \delta_1) \quad (15)$$

$$(\tilde{\rho}_{1i}, \tilde{\delta}_i) = \arg \max_{(\rho_{1i}, \delta_i) \in \tilde{\Delta}_i} l_{iT}(\tilde{\Phi}, \rho_{1i}, \delta_i), \quad i = 2, \dots, r,$$

where $\tilde{\Delta}_i$ is defined by replacing ρ_{1j} in the definition of Δ_i by $\tilde{\rho}_{1j}$. It is straightforward to see that solving these r separate maximization problems is equivalent to maximizing the function $l_T(\tilde{\Phi}, R_1, \delta)$ subject to the constraints (14). Thus, because the estimator $\tilde{\Phi}$ is consistent the consistency of the estimators \tilde{R} and $\tilde{\delta}$ formed from $\tilde{\rho}_{1i}$ and $\tilde{\delta}_i$ ($i = 1, \dots, r$) is expected to hold under appropriate regularity conditions.

Once the initial estimators $\tilde{\Phi}$, \tilde{R} and $\tilde{\delta}$ are available one can use numerical optimization methods to maximize the likelihood function $l_T(\Phi, R_1, \delta)$ subject to the constraints (14). This optimization yields Gaussian ML estimates for the parameters Φ , R_1 and δ denoted by $\hat{\Phi}$, \hat{R}_1 and $\hat{\delta}$, respectively. Gaussian ML estimates of other parameters can be formed from these estimates by transformations. For instance, the Gaussian ML estimate of Ω is $\hat{\Omega} = \hat{\Phi}^{-1}\hat{\Phi}^{-1'}$ and, since $W_1 = \Phi^{-1}R_1$ and $B_1 = \Phi'R_1$, the Gaussian ML estimates of W_1 and B_1 are $\hat{W}_1 = \hat{\Phi}^{-1}\hat{R}_1$ and $\hat{B}_1 = \hat{\Phi}'\hat{R}_1$, respectively. One can also estimate the parameters W_2 and B_2 but first an estimate for R_2 has to be obtained. For any $k \times l$ matrix A of full column rank let A_\perp be its orthogonal complement, that is, any $k \times (l - k)$ matrix of full column rank and such that $A'A_\perp = 0$. Then a Gaussian ML estimate of R_2 is given by $\hat{R}_2 = \hat{R}_{1\perp}(\hat{R}'_{1\perp}\hat{R}_{1\perp})^{-1/2}$ and Gaussian ML estimates of W_2 and B_2 are $\hat{W}_2 = \hat{\Phi}^{-1}\hat{R}_2$ and $\hat{B}_2 = \hat{\Phi}'\hat{R}_2$, respectively. These estimates are not unique (and hence not consistent) but they can be used to provide unique (and consistent) estimates for some derived parameters. For instance, $\hat{\Omega}^* = \hat{W}_2\hat{W}'_2$ is the unique Gaussian ML estimate of the parameter Ω^* in equation (11).

3.2 Limiting Distributions of Gaussian ML Estimators

Because our model is a special of the BEKK model asymptotic properties of the Gaussian ML estimators $\widehat{\Phi}$, \widehat{R}_1 and $\widehat{\delta}$ can be inferred from the results of Comte and Lieberman (2003). Under regularity conditions these results hold even if the true likelihood is not Gaussian.

Denote $\theta = [\text{vec}(R_1)' \text{vech}(\Phi)' \delta'_1 \cdots \delta'_r]'$ and let $\widehat{\theta}$ signify the corresponding vector of Gaussian ML estimators. Here vec and vech signify the usual vectorization and half vectorization operators, respectively (see Chapter 7 in Lütkepohl (1996)). The parameter vector θ satisfies the identifying constraints (14) which can be expressed in matrix form as

$$g(\theta) \stackrel{def}{=} L_r \text{vec}(R'_1 R_1) - \text{vech}(I_r) = 0.$$

Here L_r is the $\frac{1}{2}r(r+1) \times r^2$ elimination matrix defined by the equation $\text{vech}(A) = L_r \text{vec}(A)$ where A is an $r \times r$ matrix. We also introduce the $r^2 \times r^2$ commutation matrix K_{rr} defined by $K_{rr} \text{vec}(A) = \text{vec}(A')$ with A as above. Then,

$$\partial g(\theta) / \partial \text{vec}(R_1)' = [L_r (I_{r^2} + K_{rr}) (I_r \otimes R'_1) : 0] \stackrel{def}{=} G,$$

where \otimes denotes the Kronecker product (cf. result 10.5.1(2a) in Lütkepohl (1996)).

The limiting distribution of the estimator $\widehat{\theta}$ can be obtained by using Theorem 4 of Comte and Lieberman (2003) and well known results of constrained (quasi) ML estimation based on the use of Lagrange multipliers. Note, however, that the usual procedure described in Aitchison and Silvey (1958) and Davidson (2000, p. 289-290) needs to be modified because the (Gaussian) information matrix is singular. This feature can be dealt with by using the modification described in Silvey (1959). To this end, let $\bar{l}_t(\theta)$ signify the Gaussian log-likelihood of the t th observation, that is, $\bar{l}_t(\theta) = \log f_{t-1}(y_t)$ with $f_{t-1}(y_t)$ defined in (12) and interpreted as a function of the parameter vector θ . The idea is to replace the information matrix of a single observation, that is, $-\text{E}(\partial^2 \bar{l}_t(\theta) / \partial \theta \partial \theta')$ by $J \stackrel{def}{=} G'G - \text{E}(\partial^2 \bar{l}_t(\theta) / \partial \theta \partial \theta')$. Thus, proceeding as in the aforementioned references we can conclude that, even when the true

likelihood is not Gaussian but appropriate regularity conditions hold,

$$T^{1/2}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, Q_1 Q_0 Q_1), \quad (16)$$

where $Q_0 = E((\partial \bar{l}_t(\theta) / \partial \theta)(\partial \bar{l}_t(\theta) / \partial \theta'))$ and $Q_1 = J^{-1} - J^{-1} G' (G J^{-1} G')^{-1} G J^{-1}$. In the case of Gaussian likelihood the limiting distribution simplifies because then $Q_0 = J - G' G$ and, consequently, $Q_1 Q_0 Q_1 = Q_1$. A consistent estimator of the matrix Q_0 is given by $\widehat{Q}_0 = T^{-1} \sum_{t=1}^T (\partial \bar{l}_t(\widehat{\theta}) / \partial \theta)(\partial \bar{l}_t(\widehat{\theta}) / \partial \theta')$ and a consistent estimator of the matrix Q_1 is obtained by replacing the matrices J and G in the definition of Q_1 by consistent estimators. A consistent estimator of the matrix J is $\widehat{J} = \widehat{G}' \widehat{G} - T^{-1} \sum_{t=1}^T E_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial \theta \partial \theta')|_{\theta=\widehat{\theta}}$ where $\widehat{G} = [L_r (I_{r^2} + K_{rr}) (I_r \otimes \widehat{R}'_1) : 0]$ is a consistent estimator of G . Here $E_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial \theta \partial \theta')|_{\theta=\widehat{\theta}}$ could be replaced by $\partial^2 \bar{l}_t(\widehat{\theta}) / \partial \theta \partial \theta'$ but, according to Hafner and Herwartz (2003), the given form is preferable. As advocated by these authors, we use analytical derivatives instead of numerical ones to estimate the limiting covariance matrix in (16). Approximate standard errors for the components of the estimator vector $\widehat{\theta}$ can be obtained by taking square roots of the diagonal elements of the matrix $\widehat{Q}_1 \widehat{Q}_0 \widehat{Q}_1$.

From (16) it is straightforward to derive limiting distributions for other estimators of which \widehat{W}_1 and \widehat{B}_1 are of special interest. Suppose the parameter matrix Φ is symmetric. Then, it is straightforward to check that $\widehat{W}_1 - W_1 = \Phi^{-1}(\widehat{R}_1 - R_1) - \Phi^{-1}(\widehat{\Phi} - \Phi)\Phi^{-1}R_1 + o_p(T^{-1/2})$ and, upon vectorizing and using the fact $\Phi^{-1}R_1 = W_1$, one obtains

$$\begin{aligned} \text{vec}(\widehat{W}_1) - \text{vec}(W_1) &= (I_r \otimes \Phi^{-1}) \left(\text{vec}(\widehat{R}_1) - \text{vec}(R_1) \right) \\ &\quad - (W'_1 \otimes \Phi^{-1}) D_n \left(\text{vech}(\widehat{\Phi}) - \text{vech}(\Phi) \right) + o_p(T^{-1/2}), \end{aligned}$$

where D_n is the $n^2 \times n(n+1)/2$ duplication matrix defined by $\text{vec}(A) = D_n \text{vech}(A)$ for any symmetric $n \times n$ matrix A . From this and (16) one obtains

$$T^{1/2} \left(\text{vec}(\widehat{W}_1) - \text{vec}(W_1) \right) \xrightarrow{d} N(0, G_2 Q_1 Q_0 Q_1 G'_2), \quad (17)$$

where $G_2 = [(I_r \otimes \Phi^{-1}) : -(W'_1 \otimes \Phi^{-1}) D_n : 0]$. The covariance matrix of the limiting distribution can be consistently estimated by $\widehat{G}_2 \widehat{Q}_1 \widehat{Q}_0 \widehat{Q}_1 \widehat{G}'_2$ with \widehat{G}_2 defined by

replacing the parameter matrices Φ and W_1 in the definition of G_2 by $\widehat{\Phi}$ and \widehat{W}_1 , respectively. Here we assumed a symmetric specification for the parameter matrix Φ and used the relation $\text{vec}(\Phi) = D_n \text{vech}(\Phi)$. If Φ is assumed lower triangular one can use a similar reasoning with the identity $\text{vec}(\Phi) = L'_n \text{vech}(\Phi)$ (see result 9.6.3(1)(b) in Lütkepohl (1996)). Thus, in (17) and the subsequent discussion the matrix G_2 should be redefined by replacing the duplication matrix D_n by the transposed elimination matrix L'_n .

As for the estimator \widehat{B}_1 , we have $\widehat{B}_1 - B_1 = \Phi'(\widehat{R}_1 - R_1) + (\widehat{\Phi}' - \Phi')R_1 + o_p(T^{-1/2})$ from which it follows that

$$\begin{aligned} \text{vec}(\widehat{B}_1) - \text{vec}(B_1) &= (I_r \otimes \Phi') \left(\text{vec}(\widehat{R}_1) - \text{vec}(R_1) \right) \\ &\quad + (R'_1 \otimes I_n) \left(\text{vec}(\widehat{\Phi}') - \text{vec}(\Phi') \right) + o_p(T^{-1/2}). \end{aligned}$$

Assume again first that the parameter matrix Φ is symmetric and note that then $\text{vec}(\Phi') = K_{nn}D_n \text{vech}(\Phi)$ and $K_{nn}D_n = D_n$ (see results 7.2.3(a) and 9.2.3(2) in Lütkepohl (1996)). Hence, from the preceding equation and (16) we find that

$$T^{1/2} \left(\text{vec}(\widehat{B}_1) - \text{vec}(B_1) \right) \xrightarrow{d} N(0, G_3 Q_1 Q_0 Q_1 G'_3), \quad (18)$$

where $G_3 = [(I_r \otimes \Phi') : (R'_1 \otimes I_n) D_n : 0]$ and a consistent estimator of the covariance matrix of the limiting distribution is $\widehat{G}_3 \widehat{Q}_1 \widehat{Q}_0 \widehat{Q}_1 \widehat{G}'_3$ with \widehat{G}_3 defined in an obvious way. If the parameter matrix Φ is assumed lower triangular we have $\text{vec}(\Phi') = K_{nn}L'_n \text{vech}(\Phi)$. Thus, (18) and the subsequent discussion apply with the duplication matrix D_n replaced by $K_{nn}L'_n$. Note that here, as well as in the corresponding previous cases, the product $\widehat{Q}_1 \widehat{Q}_0 \widehat{Q}_1$ can be replaced by \widehat{Q}_1 if Gaussian likelihood is assumed.

In addition to approximate standard errors, the asymptotic results given in (16), (17) and (18) can also be used to construct Wald tests for smooth constraints on the involved parameters. It should be noted, however, that the limiting distributions in (16), (17) and (18) are singular and therefore care is needed to make sure that the conventional limit theory applies. For instance, in the case of the linear hypothesis $\text{Avec}(B_1) = 0$ it is not sufficient to assume that the matrix A is of full row rank. In

addition to this, the matrix A should be such that the matrix $AG_3Q_1Q_0Q_1G_3'A'$ is positive definite.

The preceding estimation and testing results were based on Gaussian likelihood. In the application of GARCH models it is frequently found that the normal distribution is not optimal and a more leptokurtic distribution, such as the t distribution, should be used. It is not difficult to set up the likelihood function based on the t distribution or any other spherical distribution and maximize it by numerical methods. For nonnormal distributions no factorization similar to that in (12) is possible, however, and no formal proof of the asymptotic properties of the obtained estimators and test statistics seems to be available.

Instead of distributions discussed above we shall consider a certain mixture of normal distributions. This is motivated by our empirical applications where the t distribution turned out to be inappropriate whereas a mixture distribution appeared more reasonable.

3.3 ML Estimation Based on a Mixture of Normal Distributions

First suppose that the error term ε_t is a mixture of two components e_{1t} and e_{2t} such that $e_{1t} \sim N(0, I_n)$ with probability p and $e_{2t} \sim N(0, \Psi)$ with probability $1 - p$ ($0 < p < 1$). Here $\Psi = \text{diag}[\psi_1 \cdots \psi_n]$ is assumed to be a nonzero diagonal matrix with positive diagonal elements. This assumption is made for ease of exposition and could be relaxed by modifying the subsequent arguments in a way similar to that used to obtain the Gaussian likelihood function. Thus, the distribution of the error term is a mixture of the $N(0, I_n)$ distribution and the $N(0, \Psi)$ distribution, and characterized by the density $p(2\pi)^{-n/2} \exp\{-\frac{1}{2}\varepsilon_t'\varepsilon_t\} + (1-p)(2\pi)^{-n/2} \det(\Psi)^{-1/2} \exp\{-\frac{1}{2}\varepsilon_t'\Psi^{-1}\varepsilon_t\}$. The implied covariance matrix of the error term is $pI_n + (1-p)\Psi$ and, since we wish to have an error term with identity covariance matrix, we make the required standardization $\varepsilon_t \rightarrow (pI_n + (1-p)\Psi)^{-1/2} \varepsilon_t$. Thus, we assume the error term to

have the density

$$\begin{aligned}\phi_\varepsilon(\varepsilon_t) &= p(2\pi)^{-n/2} \det(\Psi_1(p))^{1/2} \exp\left\{-\frac{1}{2}\varepsilon_t' \Psi_1(p) \varepsilon_t\right\} \\ &\quad + (1-p)(2\pi)^{-n/2} \det(\Psi_2(p))^{1/2} \exp\left\{-\frac{1}{2}\varepsilon_t' \Psi_2(p) \varepsilon_t\right\}\end{aligned}$$

where $\Psi_1(p) = pI_n + (1-p)\Psi$ and $\Psi_2(p) = \Psi_1(p)\Psi^{-1} = p\Psi^{-1} + (1-p)I_n$.

From the preceding discussion and (1) it follows that the conditional distribution of y_t has density

$$\begin{aligned}f_{t-1}(y_t) &= p(2\pi)^{-n/2} \det(\Sigma_{1t})^{-1/2} \exp\left\{-\frac{1}{2}y_t' \Sigma_{1t}^{-1} y_t\right\} \\ &\quad + (1-p)(2\pi)^{-n/2} \det(\Sigma_{2t})^{-1/2} \exp\left\{-\frac{1}{2}y_t' \Sigma_{2t}^{-1} y_t\right\},\end{aligned}$$

where $\Sigma_{it} = WH_t\Psi_i(p)^{-1}W'$ ($i = 1, 2$). Clearly, $\text{Cov}_{t-1}(y_t) = p\Sigma_{1t} + (1-p)\Sigma_{2t}$. An interesting feature in this model is that, if the last $n-r$ diagonal elements of the matrix Ψ are distinct, even the parameter matrix W_2 , and hence also B_2 , is identifiable (up to multiplying columns by minus one). This can be readily seen from the form of the density $f_{t-1}(y_t)$ and the definitions of the conditional covariance matrices Σ_{1t} and Σ_{2t} . In what follows, we shall assume this condition and only briefly discuss the case where the last $n-r$ diagonal elements of the matrix Ψ are not distinct.

The unknown parameters of the model are collected in the parameter vector $\vartheta = [b' \delta' \psi' p]'$ where $b = \text{vec}(B)$. From the definitions it follows that $\det(\Sigma_{it})^{-1/2} = \det(B) \det(H_t)^{-1/2} \det(\Psi_i(p))^{1/2}$. Thus, the log-likelihood function can be written as

$$l_T(\vartheta) = T \log \det(B) - \frac{1}{2} \sum_{i=1}^r \sum_{t=1}^T \log v_{it} + \sum_{t=1}^T \log L_t(\vartheta),$$

where $L_t(\vartheta) = pL_{1t}(\vartheta) + (1-p)L_{2t}(\vartheta)$ with

$$L_{it}(\vartheta) = \det(\Psi_i(p))^{1/2} \exp\left\{-\frac{1}{2}y_t' BH_t^{-1}\Psi_i(p)B'y_t\right\}.$$

Because now all the parameters are identifiable the log-likelihood function can be maximized directly without any constraints used in the Gaussian case. Initial values

for the parameters that also enter the Gaussian likelihood can be obtained in the same way as in the Gaussian case or, if Gaussian ML estimates are available, they can be used as initial values. Limiting distributions of the ML estimators can be derived in the usual way although no formal proof on this seems to be available. Thus, if $\widehat{\vartheta}$ signifies the ML estimator of ϑ and a correct model specification is assumed we have

$$T^{1/2}(\widehat{\vartheta} - \vartheta) \xrightarrow{d} N\left(0, (-E(\partial^2 \bar{l}_t(\vartheta) / \partial \vartheta \partial \vartheta')^{-1})\right). \quad (19)$$

A consistent estimator of the information matrix is $-T^{-1} \sum_{t=1}^T E_{t-1}(\partial^2 \bar{l}_t(\vartheta) / \partial \vartheta \partial \vartheta')|_{\vartheta=\widehat{\vartheta}}$ where $E_{t-1}(\partial^2 \bar{l}_t(\vartheta) / \partial \vartheta \partial \vartheta')|_{\vartheta=\widehat{\vartheta}}$ can again be replaced by $\partial^2 \bar{l}_t(\widehat{\vartheta}) / \partial \vartheta \partial \vartheta'$. Another possibility, used in the empirical application of the paper, is given by the outer product form $T^{-1} \sum_{t=1}^T (\partial \bar{l}_t(\widehat{\vartheta}) \partial \vartheta)(\partial \bar{l}_t(\widehat{\vartheta}) / \partial \vartheta')$.

It is also of interest to consider the limiting distribution of $\widehat{W} = \widehat{B}'^{-1}$ where we have used the same notation as in the Gaussian case. It is straightforward to check that $\widehat{W} - W = -W(\widehat{B}' - B')W + o_p(T^{-1/2})$ and, by arguments similar to those used to obtain (18),

$$T^{1/2}(\text{vec}(\widehat{W}) - \text{vec}(W)) = -(W' \otimes W) K_{mn} T^{1/2}(\text{vec}(\widehat{B}) - \text{vec}(B)) + o_p(1).$$

Combining this with (19) it is straightforward to obtain the desired limiting distribution and a consistent estimator of its covariance matrix.

The above results can be used to construct conventional likelihood based tests for smooth constraints on the involved parameters. Because all the parameters are assumed to be identifiable no problems with potentially singular limiting distributions arise. However, this issue becomes relevant if at least two of the last $n - r$ diagonal elements of the matrix Ψ are identical. Without going into details we note that then the likelihood function should be modified in a way similar to that used in the Gaussian case.

Finally, note that above we assumed that the employed mixture distribution is correct. If that is not assumed one can proceed in the same way as in the case of Gaussian likelihood and modify the limiting distribution in (19). Because now the

information matrix is nonsingular this can be done in the conventional way described, for example, in Davidson (2000, p. 219-220).

4 Testing the Order of Conditional Heteroskedasticity

In practice the integer r , which we refer to as the order of conditional heteroskedasticity, is usually unknown. Therefore it is of interest to have test procedures which can help to specify an appropriate value for this parameter.

Suppose that a chosen value r represents the truth. From equations (4) and (5) it then follows that the r linear combinations of y_t given by $B_1'y_t$ are conditionally heteroskedastic whereas the $n - r$ linear combinations $B_2'y_t$ are homoskedastic or, in fact, $B_2'y_t \sim \text{i.i.d.}(0, I_{n-r})$. Therefore, a natural way to test the correctness of the specified order of conditional heteroskedasticity is to test whether the linear combinations $B_2'y_t$ really are free of conditional heteroskedasticity. The test procedures to be developed below are based on this idea.

A difficulty with the testing problem discussed above is that some parameters of the model may be identified under the alternative only. To see this, consider the Gaussian likelihood and the null hypothesis which states that $\alpha_r = \beta_r = 0$ in (6). Under this null hypothesis, the order of conditional heteroskedasticity is $r - 1$ and $v_{rt} = 1$ for all t . From equation (10) it can be seen that the last column of the matrix R_1 is then not identified because it can take any value without any effect on the conditional covariance matrix of the observed series. This nonidentifiability makes the likelihood ratio test intractable. Instead of the likelihood ratio test we shall therefore consider Lagrange multiplier type tests which are convenient because unrestricted estimation of the model is not required.

We shall explicitly only discuss the cases where no column of the parameter matrix B_2 is identifiable and where the parameter matrix B_2 is fully identifiable. The other cases can be handled by a straightforward combination of the arguments used in these two cases. Unless otherwise stated, the subsequent discussion assumes the null

hypothesis that the order of conditional heteroskedasticity equals r .

Our first test procedure is based on the approach of Ling and Li (1997). The idea is to test whether a sample analog of the univariate process $y_t' B_2 B_2' y_t$ is serially uncorrelated. This sample analog is formed from an estimator of B_2 denoted by \tilde{B}_2 . When no column of the parameter matrix B_2 is identifiable the estimator \tilde{B}_2 is supposed to be of the form $\tilde{B}_2 = \tilde{C}^{-1} \tilde{R}_2$ where \tilde{C} is any estimator of C consistent of order $O_p(T^{-1/2})$ and $\tilde{R}_2 = \tilde{R}_{1\perp} (\tilde{R}'_{1\perp} \tilde{R}_{1\perp})^{-1/2}$ can similarly be based on any estimator of R_1 consistent of order $O_p(T^{-1/2})$. Note that the use of efficient estimators is not necessary. If the parameter matrix B_2 is fully identifiable it can be estimated directly in which case we simply assume that the estimator \tilde{B}_2 is available and consistent of order $O_p(T^{-1/2})$.

Following Ling and Li (1997) we form the series $y_t' \tilde{B}_2 \tilde{B}_2' y_t$ and its centered version $\tilde{\xi}_{2t} = y_t' \tilde{B}_2 \tilde{B}_2' y_t - T^{-1} \sum_{t=1}^T y_t' \tilde{B}_2 \tilde{B}_2' y_t$. With this notation, the sample serial covariances of $y_t' \tilde{B}_2 \tilde{B}_2' y_t$ can be defined as

$$\tilde{\gamma}_{22}(k) = T^{-1} \sum_{t=k}^T \tilde{\xi}_{2t} \tilde{\xi}_{2,t-k}, \quad 0 \leq k < T,$$

and $\tilde{\gamma}_{22}(k) = \tilde{\gamma}_{22}(-k)$ for $T < k < 0$. Now we can introduce the test statistic

$$\mathcal{T}_1(K) = T \sum_{k=1}^K [\tilde{\gamma}_{22}(k) / \tilde{\gamma}_{22}(0)]^2 \xrightarrow{d} \chi^2(K),$$

where the limiting distribution is derived in the appendix. In addition to the assumptions already stated, its derivation assumes that the innovation process ε_t , and hence y_t , has finite fourth moments. Otherwise the distribution of ε_t may be arbitrary. In practice one may also investigate the validity of the null hypothesis by using the individual autocorrelation estimators $\tilde{\gamma}_{22}(k) / \tilde{\gamma}_{22}(0)$ which can be treated as approximately normally distributed with mean zero and variance $1/T$.

The above test is based on serial correlations of the univariate series $y_t' \tilde{B}_2 \tilde{B}_2' y_t$ where \tilde{B}_2 is as above. We shall now consider another test which is based on the analogous matrix valued series $\tilde{B}_2' y_t y_t' \tilde{B}_2$. By symmetry, some elements of this matrix are

redundant so that it is reasonable to use its half vectorized form $\text{vech}(\tilde{B}'_2 y_t y'_t \tilde{B}_2)$ and, upon centering, consider the $(n-r)(n-r+1)/2$ dimensional series $\tilde{\zeta}_{2t} = \text{vech}(\tilde{B}'_2 y_t y'_t \tilde{B}_2) - T^{-1} \sum_{t=1}^T \text{vech}(\tilde{B}'_2 y_t y'_t \tilde{B}_2)$. The related matrices of sample serial covariances are

$$\tilde{\Gamma}(k) = T^{-1} \sum_{t=k}^T \tilde{\zeta}_{2t} \tilde{\zeta}'_{2,t-k}, \quad 0 \leq k < T,$$

and $\tilde{\Gamma}(k) = \tilde{\Gamma}(-k)'$ for $T < k < 0$. These are used to define the test statistic

$$\mathcal{T}_2(K) = T \sum_{k=1}^K \text{tr} \tilde{\Gamma}(k)' \tilde{\Gamma}(0)^{-1} \tilde{\Gamma}(k) \tilde{\Gamma}(0)^{-1} \xrightarrow{d} \chi^2 (K(n-r)^2(n-r+1)^2/4),$$

where tr signifies the trace of a square matrix. The stated limiting distribution is again derived in the appendix under the same conditions as assumed for $\mathcal{T}_1(K)$.

It is interesting and useful that the limiting distributions of the above test statistics apply without any particular additional assumptions about the distribution of the observations. In particular, the obtained limiting distributions apply even when the errors ε_t are not normally distributed. This is in contrast with some previous tests on conditional heteroskedasticity in which such assumptions are required or modifications of test statistics are needed to guarantee the applicability of the conventional chi-square criterion (see e.g. Ling and Li (1996) and the references therein).

If no prior information about an appropriate value of r is available one may wish to apply the above tests sequentially by starting with $r = 1$ (assuming that $r = 0$ can be ruled a priori). If $r = 1$ is rejected one can next try $r = 2$ and so on until the first nonrejection is obtained. Note that it is reasonable to start testing from small values of r because, if the value of r is chosen too large, an unidentified model is estimated.

5 Empirical Application

As an empirical example we consider an application to a four-dimensional system consisting of four weekly foreign exchange rate return series. The data set comprises the exchange rates of the French Franc (FRF), Dutch Guilder (NLG), German Mark (DEM) and Swiss Franc (CHF) against the U.S. Dollar (USD) from the beginning of

1984 until the end of 1997 (782 observations). As Figure 1 shows, the series exhibit conditional heteroskedasticity, and against the dollar the rates appear to be relatively stable. However, the currencies belonging to the European Monetary System (the French franc, German mark and Dutch guilder) underwent some major realignments and the foreign exchange markets experienced a number of spells of turmoil during the sample period. These are clearly visible as spikes and excessive volatility in the implied cross rate returns in Figure 2, whereas the implied cross rates against the Swiss franc remained quite stable as exemplified by the implied CHF/DEM return series. It is also obvious from Figure 2 that the NLG/DEM rate fluctuates in a much narrower range than the rest.

Let us first consider models under normality. We start the analysis by selecting the order of conditional heteroskedasticity, r . A sequential application of test statistics \mathcal{T}_1 and \mathcal{T}_2 as described at the end of the previous section alongside plots of the allegedly conditionally homoskedastic and heteroskedastic linear combinations of the returns are employed in the selection process. According to the test results in the upper panel of Table 1 we cannot reject the null hypothesis of two conditionally heteroskedastic factors driving the system at the 5% level using either test. In contrast, the hypothesis of only one factor is easily rejected.

Estimation results for the system with two conditionally heteroskedastic factors under normality are presented in Table 2. The standard errors and test results to be reported below are based on the robust covariance matrix estimator, so they are asymptotically valid even if the assumed normal distribution is not correct. Interestingly, both of the conditionally heteroskedastic factors turn out to be implied exchange rate returns between European currencies. In the first column of the estimated B_1 matrix only the second and third elements are significant at the 5% level. They are approximately equal in absolute value but of opposite sign, suggesting that the first conditionally heteroskedastic linear combination could be the (scaled) difference between the NLG/USD and DEM/USD returns, i.e., the implied NLG/DEM return. The Wald test does not reject this hypothesis (p-value equals 0.13). In the second

column of the estimate of B_1 only the first and third elements are significant, and the null hypothesis of the second conditionally heteroskedastic linear combination being the implied (scaled) FRF/DEM return cannot be rejected by a Wald test (p-value equals 0.79). However, the coefficients of the first factor in the W matrix are very inaccurately estimated while those of the second factor are clearly significant, suggesting that the one-factor specification might be adequate.

The first conditionally heteroskedastic factor exhibits rather low persistence while the second factor is highly persistent (the estimates of the α and β parameters sum to 0.639 and 0.997, respectively). The heteroskedastic and homoskedastic linear combinations of the observed series are depicted in Figure 3. According to these plots the normality assumption seems to be inadequate for this data set. In particular, the numerous spikes in these plots indicate excessive kurtosis. This together with the fact that the coefficients in the estimated W matrix corresponding to the first factor have very large standard errors suggests that a model with one factor and nonnormal error distribution might be a more appropriate specification. It is also noteworthy that the estimated β parameter of the first factor is insignificant and the estimated α parameter of the second factor is clearly greater than the estimate of the β parameter. These findings suggest that the factors may primarily be driven by "overreaction" to exceptionally big news and, in turn, lend additional support to the conjecture that the model may not be appropriate, but a separate regime for these events may be called for.

Instead of normality a leptokurtic error distribution seems to be needed. One commonly employed alternative for return series is the t distribution but it turned out to be inappropriate here. In the model under normality the estimated second factor primarily seems to capture the rather few outlying observations as attested by its graph in Figure 3 and the fact that it is dominated by the ARCH term. These observations lend support to a model in which the errors are generated by two different distributions of which one generates rather infrequent but large errors. The model based on the mixture of normal distributions introduced in Section 3.1 is designed

with this idea in mind. Test results for the order of conditional heteroskedasticity based on this error distribution are presented in the lower panel of Table 1. In this case the test results are ambiguous at the 5% significance level. Neither test rejects the null hypothesis of two factors (p-values of the \mathcal{T}_1 and \mathcal{T}_2 tests equal 0.12 and 0.09, respectively), while the \mathcal{T}_1 test rejects and the \mathcal{T}_2 test does not reject the hypothesis of one factor (p-values equal 0.03 and 0.08, respectively). However, based on the results of the normal specification above, we are inclined to favor the model with one factor.⁴

The estimation results of the one-factor model are presented in Table 3, including the conditionally homoskedastic linear combinations (B_2) and corresponding elements in the W matrix (W_2) that are identifiable in this specification. The standard errors are based on the outer product of the (analytical) score estimator. Because these standard errors can be badly behaved in finite samples, as recently pointed out by Mencia and Sentana (2004), we use their method of computing the gradient from a simulated realization of 100,000 observations. The estimates of the elements of Ψ are rather small (in particular less than unity) and the probability or mixing proportion p is estimated as small as 0.129. This is in accordance with the prior expectation that rather a small error variance prevails most of the time. The conditionally heteroskedastic factor exhibits rather moderate persistence (the sum of the estimated α and β parameters equals 0.935) and the GARCH term clearly dominates, suggesting that the introduction of the mixture distribution has successfully captured the aberrant observations. The factor in the first column of the estimated B matrix turns out to be the (scaled) difference between the German mark and Dutch guilder returns, i.e., the implied DEM/NLG return. This hypothesis cannot be rejected by a Wald test

⁴The \mathcal{T}_1 and \mathcal{T}_2 tests are essentially tests for conditional homoskedasticity and similar tests have been shown to overreject in the presence of aberrant observations. As pointed out by Franses et al. (2004), the problem is especially severe when there are patches of outliers as seems to be the case here. Moreover, we also attempted to estimate a two-factor model. The estimation algorithm converged very slowly and, according to Wald tests, only the first column of the B_1 matrix turned out to be statistically significant. These findings provide additional support for a model with one factor.

(p-value equals 0.10). In other words, changes in the DEM/NLG rate tend to increase conditional volatility of the factor. As pointed out above, the range of variation of the NLG/DEM rate is much narrower than that of the other rates, and it experienced no realignments during the sample period. Thus, a possible interpretation of the outcome might be that this particular exchange rate was the most sensitive such that even its small movements were seen as indicative of changes in market conditions, giving rise to higher volatility.⁵ According to Wald tests the equation for the French franc return is the only one where the conditionally heteroskedastic factor does not enter (at the 5% significance level). This would imply the rather unexpected finding that there is no dynamics in the conditional variance of the French franc return. As an additional check for this, we applied a likelihood ratio test. According to this test, the null hypothesis that the French franc return is not affected by the conditionally heteroskedastic factor was rejected at any reasonable significance level.

Because all the exchange rates in the system are expressed in terms of the U.S. dollar, a shock of the same size to all the returns (i.e., to all the elements of ε_t) can be interpreted as a shocks to the U.S. economy. The coefficients in the first column of W give the effects of the shock on the returns, and as all the estimates except the one for the DEM/USD return are negative, following a positive shock, the French franc, Dutch guilder and Swiss franc tend to appreciate vis-à-vis the dollar, while the German mark tends to depreciate. Also, the coefficients of the NLG/USD and CHF/USD returns are almost equal while that of the FRF/USD return is clearly smaller, suggesting substantial appreciation of the Dutch guilder and Swiss franc and moderate appreciation of the French franc vis-à-vis the German mark. If the shock is negative the effects are reversed. Moreover, the higher the conditional volatility in the system, the greater is the impact of such a shock on all the returns.

⁵We also tested for asymmetry by testing whether positive and negative values of the factor have equal effect on the conditional variance. A likelihood ratio test cannot reject the null hypothesis of symmetry (p-value equals 0.32).

6 Conclusion

In this paper we extend previous multivariate generalized orthogonal GARCH models to allow for a reduced number of conditionally heteroskedastic factors. Unlike in previous similar models we also develop test procedures which can be used to specify an appropriate number of factors needed to adequately describe the conditional heteroskedasticity in the data. In addition to Gaussian likelihood, estimation based on a mixture of normal distributions is also considered. The latter, motivated by the empirical application of the paper, appears useful when one needs to allow for aberrant observations, which in our case are due to realignments of the considered exchange rates.

It is shown that the Gaussian likelihood can be expressed in a convenient form and its numerical maximization can be facilitated by using simple preliminary estimates. In large systems such preliminary estimates may also be useful in their own right. Because in high-dimensional GARCH models full maximum likelihood estimation is, in general, difficult, our model may thus offer a feasible alternative.

Appendix 1

In this appendix the limiting distributions of test statistics $\mathcal{T}_1(K)$ and $\mathcal{T}_2(K)$ are derived. Unless otherwise stated, all the assumptions stated in the paper will be assumed, including the correctness of the null hypothesis that the order of conditional heteroskedasticity equals r . We shall explicitly only consider the case where no column of the parameter matrix B_2 is identifiable. The employed arguments can be straightforwardly modified to other relevant cases.

Derivation of the limiting distribution of $\mathcal{T}_1(K)$. Let $\gamma_{22}^*(k)$ be the counterpart of $\tilde{\gamma}_{22}(k)$ defined by using $y_t' B_2 B_2' y_t = \varepsilon_{2t}' \varepsilon_{2t}$ in place of $y_t' \tilde{B}_2 \tilde{B}_2' y_t$. We start by showing that, for any fixed k ,

$$\tilde{\gamma}_{22}(k) - \gamma_{22}^*(k) = \begin{cases} o_p(T^{-1/2}), & k > 0 \\ o_p(1), & k = 0. \end{cases} \quad (20)$$

Suppose first that $k > 0$ and define ξ_{2t} by replacing the estimator \tilde{B}_2 in the definition of $\tilde{\xi}_{2t}$ by B_2 . Then we can write

$$\begin{aligned} \tilde{\gamma}_{22}(k) - \gamma_{22}^*(k) &= T^{-1} \sum_{t=k}^T \xi_{2,t-k} (\tilde{\xi}_{2t} - \xi_{2t}) + T^{-1} \sum_{t=k}^T \xi_{2t} (\tilde{\xi}_{2,t-k} - \xi_{2,t-k}) \\ &\quad + T^{-1} \sum_{t=k}^T (\tilde{\xi}_{2t} - \xi_{2t}) (\tilde{\xi}_{2,t-k} - \xi_{2,t-k}). \end{aligned} \quad (21)$$

Using the definitions and assumptions it is straightforward to check that the first term on the right hand side can be expressed as

$$T^{-1} \sum_{t=k}^T \xi_{2,t-k} (\tilde{\xi}_{2t} - \xi_{2t}) = T^{-1} \sum_{t=k}^T \xi_{2,t-k} y_t' (\tilde{B}_2 \tilde{B}_2' - B_2 B_2') y_t + o_p(T^{-1/2}).$$

Now define the $n \times n$ matrices $M = B_2 R_2'$ and $\tilde{M} = \tilde{B}_2 \tilde{R}_2'$. Because $\tilde{M} = \tilde{C}^{-1'} \tilde{R}_2 \tilde{R}_2' = \tilde{C}^{-1'} (I_n - \tilde{R}_1 \tilde{R}_1')$ and because the estimators \tilde{C} and \tilde{R}_1 are assumed to be consistent of order $O_p(T^{-1/2})$, we have $\tilde{M} = M + O_p(T^{-1/2})$. Thus, $y_t' (\tilde{B}_2 \tilde{B}_2' - B_2 B_2') y_t = y_t' (\tilde{M} \tilde{M}' - M M') y_t = 2y_t' M (\tilde{M} - M)' y_t + y_t' (\tilde{M} - M) (\tilde{M} - M)' y_t$ and

$$\begin{aligned} T^{-1} \sum_{t=k}^T \xi_{2,t-k} (\tilde{\xi}_{2t} - \xi_{2t}) &= 2\text{tr}(\tilde{M} - M)' T^{-1} \sum_{t=k}^T \xi_{2,t-k} y_t y_t' M \\ &\quad + \text{tr}(\tilde{M} - M) (\tilde{M} - M)' T^{-1} \sum_{t=k}^T \xi_{2,t-k} y_t y_t' + o_p(T^{-1/2}). \end{aligned} \quad (22)$$

In the second term on the right hand side we can use the definition of $\xi_{2,t-k}$ to obtain

$$T^{-1} \sum_{t=k}^T \xi_{2,t-k} y_t y_t' = T^{-1} \sum_{t=k}^T \|\varepsilon_{2,t-k}\|^2 y_t y_t' - T^{-1} \sum_{t=1}^T \|\varepsilon_{2t}\|^2 T^{-1} \sum_{t=k}^T y_t y_t' = O_p(1).$$

Here the latter equality follows from a law of large numbers because the process $(\|\varepsilon_{2t}\|^2, y_t y_t')$ is stationary and ergodic with finite second moments. This, in conjunction with the fact $\tilde{M} - M = O_p(T^{-1/2})$, shows that the second term on the right hand side of (22) is of order $o_p(T^{-1/2})$. To see that the same is true for the first term, conclude from the definitions that $y_t' M = \varepsilon_t' H_t^{1/2} W' B_2 R_2' = \varepsilon_{2t}' R_2'$. Thus, from the

definition of $\xi_{2,t-k}$ and equation (9) it follows that

$$\begin{aligned} T^{-1} \sum_{t=k}^T \xi_{2,t-k} y_t y_t' M &= T^{-1} \sum_{t=k}^T \left(\|\varepsilon_{2,t-k}\|^2 - T^{-1} \sum_{t=1}^T \|\varepsilon_{2t}\|^2 \right) W_1 V_t^{1/2} \varepsilon_{1t} \varepsilon_{2t}' R_2' \\ &+ T^{-1} \sum_{t=k}^T \left(\|\varepsilon_{2,t-k}\|^2 - T^{-1} \sum_{t=1}^T \|\varepsilon_{2t}\|^2 \right) W_2 \varepsilon_{2t} \varepsilon_{2t}' R_2'. \end{aligned} \quad (23)$$

It is easy to check that on the right hand side the replacement of the sample mean $T^{-1} \sum_{t=1}^T \|\varepsilon_{2t}\|^2$ by its expected value $n - r$ causes an error of order $o_p(1)$ and, after this replacement, the resulting summands are in both cases realizations from zero mean stationary and ergodic processes. Thus, both terms on the right hand side of (23) are of order $o_p(1)$ and, since $\widetilde{M} - M = O_p(T^{-1/2})$, the first term on the right hand side of (22) is of order $o_p(T^{-1/2})$.

Altogether we have shown that the first term on the right hand side of (21) is of order $o_p(T^{-1/2})$. By similar arguments it can be seen that the same is true for the second and third terms. For the second term this essentially amounts to considering (22) with $k < 0$ whereas the third term is clearly of a lower order of magnitude. Hence, we have established (20) when $k > 0$. The case $k = 0$ can be handled similarly.

Now recall that $\gamma_{22}^*(k)$ is the k th sample serial covariance obtained from $\|\varepsilon_{2t}\|^2$, ($t = 1, \dots, T$). Thus, well-known results about stationary time series show that the corresponding sample serial correlations $\gamma_{22}^*(k)/\gamma_{22}^*(0)$ are asymptotically independent and normally distributed with mean zero and variance $1/T$. From (20) it readily follows that the same is true for the corresponding observed quantities $\widetilde{\gamma}_{22}(k)/\widetilde{\gamma}_{22}(0)$. The limiting distribution of test statistic $\mathcal{T}_1(K)$ follows from this.

Derivation of the limiting distribution of $\mathcal{T}_2(K)$. We consider analogues of the sample serial covariance matrices $\widetilde{\Gamma}(k)$ defined by replacing the half vectorization operator vech by the ordinary vectorization operator vec . Thus, set $\widetilde{\zeta}_{2t} = \text{vec}(\widetilde{B}_2' y_t y_t' \widetilde{B}_2) - T^{-1} \sum_{t=1}^T \text{vec}(\widetilde{B}_2' y_t y_t' \widetilde{B}_2)$, and define

$$\widetilde{\Gamma}(k) = T^{-1} \sum_{t=k}^T \widetilde{\zeta}_{2t} \widetilde{\zeta}_{2,t-k}', \quad 0 < k < T,$$

and $\tilde{\Gamma}(k) = \tilde{\Gamma}(-k)'$ for $T < k < 0$. The unobserved counterparts of $\tilde{\zeta}_{2t}$ and $\tilde{\Gamma}(k)$ obtained by replacing the estimator \tilde{B}_2 by B_2 are denoted by ζ_{2t}^* and $\Gamma^*(k)$, respectively. In order to simplify notation we denote $\bar{w}_t = w_t - T^{-1} \sum_{t=1}^T w_t$ for any time series vector w_t ($t = 1, \dots, T$) and $A^{(2)} = A \otimes A$ for any matrix A . Various results of the Kronecker product and the vectorization operator to be used below can be found in Sections 2.4 and 7.2 of Lütkepohl (1996).

Analogously to (20) we first show that, for any fixed k ,

$$\tilde{R}_2^{(2)} \tilde{\Gamma}(k) \tilde{R}_2^{(2)'} - R_2^{(2)} \Gamma^*(k) R_2^{(2)'} = \begin{cases} o_p(T^{-1/2}), & k > 0 \\ o_p(1), & k = 0. \end{cases} \quad (24)$$

By properties of the vec operator, $\text{vec}(\tilde{B}_2' y_t y_t' \tilde{B}_2) = \tilde{B}_2^{(2)'} y_t^{(2)}$. Hence, $\tilde{\zeta}_{2t} = \tilde{B}_2^{(2)'} \overline{y_t^{(2)}}$ and $\tilde{\Gamma}(k) = \tilde{B}_2^{(2)'} \tilde{\Gamma}_y^{(2)}(k) \tilde{B}_2^{(2)}$ where $\tilde{\Gamma}_y^{(2)}(k) = T^{-1} \sum_{t=k}^T \overline{y_t^{(2)}} \overline{y_{t-k}^{(2)'}}$. In the same way it can be seen that $\Gamma^*(k) = B_2^{(2)'} \tilde{\Gamma}_y^{(2)}(k) B_2^{(2)}$.

Consider first the case $k > 0$. Using the matrices M and \tilde{M} defined in the previous proof we can write

$$\begin{aligned} \tilde{R}_2^{(2)} \tilde{\Gamma}(k) \tilde{R}_2^{(2)'} - R_2^{(2)} \Gamma^*(k) R_2^{(2)'} &= \tilde{M}^{(2)'} \tilde{\Gamma}_y^{(2)}(k) \tilde{M}^{(2)} - M^{(2)'} \tilde{\Gamma}_y^{(2)}(k) M^{(2)} \quad (25) \\ &= (\tilde{M}^{(2)} - M^{(2)})' \tilde{\Gamma}_y^{(2)}(k) M^{(2)} \\ &\quad + M^{(2)'} \tilde{\Gamma}_y^{(2)}(k) (\tilde{M}^{(2)} - M^{(2)}) \\ &\quad + (\tilde{M}^{(2)} - M^{(2)})' \tilde{\Gamma}_y^{(2)}(k) (\tilde{M}^{(2)} - M^{(2)}). \end{aligned}$$

Because the process y_t is stationary and ergodic with finite fourth moments, a law of large numbers implies that $\tilde{\Gamma}_y^{(2)}(k) = O_p(1)$. This, in conjunction with the result $\tilde{M} - M = O_p(T^{-1/2})$, shows that the last term in the last expression of (25) is of order $o_p(T^{-1/2})$. We show that the this is also the case for the first term.

From the identity

$$\tilde{M}^{(2)} - M^{(2)} = M \otimes (\tilde{M} - M) + (\tilde{M} - M) \otimes M + (\tilde{M} - M) \otimes (\tilde{M} - M)$$

it follows that

$$\begin{aligned}
(\widetilde{M}^{(2)} - M^{(2)})' \widetilde{\Gamma}_y^{(2)}(k) M^{(2)} &= (M' \otimes (\widetilde{M} - M)') \widetilde{\Gamma}_y^{(2)}(k) M^{(2)} \\
&\quad + ((\widetilde{M} - M)' \otimes M') \widetilde{\Gamma}_y^{(2)}(k) M^{(2)} \\
&\quad + (\widetilde{M} - M)^{(2)'} \widetilde{\Gamma}_y^{(2)}(k) M^{(2)}.
\end{aligned}$$

As noticed above, $\widetilde{\Gamma}_y^{(2)}(k) = O_p(1)$ which in conjunction with the result $\widetilde{M} - M = O_p(T^{-1/2})$ implies that the last term on the right hand side of the preceding equation is of order $o_p(T^{-1/2})$. Thus, we need to show that the same is true for the first and second terms. It suffices to consider the former which can be expressed as

$$\begin{aligned}
(M' \otimes (\widetilde{M} - M)') \widetilde{\Gamma}_y^{(2)} M^{(2)} &= (M' \otimes (\widetilde{M} - M)') T^{-1} \sum_{t=k}^T \overline{y_t^{(2)}} \overline{y_{t-k}^{(2)'} } M^{(2)} \\
&= (I_n \otimes (\widetilde{M} - M)') T^{-1} \sum_{t=k}^T \overline{(M' y_t \otimes y_t)} \overline{(M' y_{t-k} \otimes M' y_{t-k})}' \\
&= (I_n \otimes (\widetilde{M} - M)') T^{-1} \sum_{t=k}^T \overline{(R_2 \varepsilon_{2t} \otimes y_t)} \overline{(R_2 \varepsilon_{2,t-k} \otimes R_2 \varepsilon_{2,t-k})}'.
\end{aligned}$$

Here the last equality is based on the fact $M' y_t = R_2 \varepsilon_{2t}$ already used in the previous proof. Because the process $R_2 \varepsilon_{2t} \otimes y_t$ is stationary and ergodic with finite expectation it obeys a law of large numbers. Using this fact in conjunction with $\widetilde{M} - M = O_p(T^{-1/2})$ it is not difficult to check that the replacement of the sample mean in the last expression above by the corresponding expectation causes an error of order $o_p(T^{-1/2})$. Thus,

$$\begin{aligned}
&(M' \otimes (\widetilde{M} - M)') \widetilde{\Gamma}_y^{(2)}(k) M^{(2)} \\
&= (I_n \otimes (\widetilde{M} - M)') T^{-1} \sum_{t=k}^T (R_2 \varepsilon_{2t} \otimes y_t - \mathbf{E}(R_2 \varepsilon_{2t} \otimes y_t)) \\
&\quad \times (R_2 \varepsilon_{2,t-k} \otimes R_2 \varepsilon_{2,t-k} - \mathbf{E}(R_2 \varepsilon_{2,t-k} \otimes R_2 \varepsilon_{2,t-k}))' + o_p(T^{-1/2}).
\end{aligned}$$

Now recall from equation (9) the representation $y_t = W_1 V_t^{1/2} \varepsilon_{1t} + W_2 \varepsilon_{2t}$ and insert it in the first term of the last expression which becomes a sum of two terms. It is straightforward to check that, for $k > 0$, the summands in all these two terms

are stationary and ergodic martingale differences and, hence, of order $o_p(1)$. Thus, because $\widetilde{M} - M = O_p(T^{-1/2})$ it follows that $M' \otimes (\widetilde{M} - M)' \widetilde{\Gamma}_y^{(2)}(k) M^{(2)} = o_p(T^{-1/2})$.

We have thus shown that the first term in the last expression of (25) is of order $o_p(T^{-1/2})$. A similar proof shows that this is also the case for the second term. The proof is essentially based on arguments used above with $k < 0$. Altogether we have shown that, for $k > 0$, all the three terms in the last expression of (25) are of order $o_p(T^{-1/2})$. This proves (24) in the case $k > 0$. The case $k = 0$ can be obtained from (25) by using the facts $\widetilde{\Gamma}_y^{(2)}(k) = O_p(1)$ and $\widetilde{M} - M = O_p(T^{-1/2})$.

We shall next obtain an alternative expression for test statistic $\mathcal{T}_2(K)$. To this end, let A^+ signify the Moore-Penrose inverse of the matrix A . Then, we can write

$$\begin{aligned} \mathcal{T}_2(K) &= T \sum_{k=1}^K \text{tr} \left(\widetilde{R}_2^{(2)} \widetilde{\Gamma}(k) \widetilde{R}_2^{(2)'} \right)' \left(\widetilde{R}_2^{(2)} \widetilde{\Gamma}(0) \widetilde{R}_2^{(2)'} \right)^+ \\ &\quad \times \left(\widetilde{R}_2^{(2)} \widetilde{\Gamma}(k) \widetilde{R}_2^{(2)'} \right) \left(\widetilde{R}_2^{(2)} \widetilde{\Gamma}(0) \widetilde{R}_2^{(2)'} \right)^+. \end{aligned} \quad (26)$$

In order to justify this, conclude from the definitions of the vectors $\widetilde{\zeta}_{2t}$ and $\widetilde{\underline{\zeta}}_{2t}$ that $\widetilde{\underline{\zeta}}_{2t} = D_{n-r} \widetilde{\zeta}_{2t}$ and $\widetilde{\zeta}_{2t} = D_{n-r}^+ \widetilde{\underline{\zeta}}_{2t}$ where $D_{n-r}^+ = (D'_{n-r} D_{n-r})^{-1} D'_{n-r}$. Thus, $\widetilde{\Gamma}(k) = D_{n-r} \widetilde{\Gamma}(k) D'_{n-r}$ and $\widetilde{\Gamma}(k) = D_{n-r}^+ \widetilde{\Gamma}(k) D_{n-r}^+$. From these facts and the definition of the Moore-Penrose inverse it can be straightforwardly shown that $\left(\widetilde{R}_2^{(2)} \widetilde{\Gamma}(0) \widetilde{R}_2^{(2)'} \right)^+ = \widetilde{R}_2^{(2)} D_{n-r}^+ \widetilde{\Gamma}(0)^{-1} D_{n-r}^+ \widetilde{R}_2^{(2)'}$. Using this identity and the fact $\widetilde{R}_2^{(2)'} \widetilde{R}_2^{(2)} = I_{n-r}$ on the right hand side of (26) it can then be seen that the stated equation holds.

Next note that from the definitions it follows that $\underline{\Gamma}^*(k)$ is the sample serial covariance matrix of the random vector $\varepsilon_{2t} \otimes \varepsilon_{2t}$ at lag k . Thus, by standard results, $R_2^{(2)} \underline{\Gamma}^*(0) R_2^{(2)'}$ converges in probability to $R_2^{(2)} \text{Cov}(\varepsilon_{2t} \otimes \varepsilon_{2t}) R_2^{(2)'}$ and, by (24), the same is true for $\widetilde{R}_2^{(2)} \widetilde{\Gamma}(0) \widetilde{R}_2^{(2)'}$. The rank of the covariance matrix $\text{Cov}(\varepsilon_{2t} \otimes \varepsilon_{2t})$ is not full but $(n-r)(n-r+1)/2$ or the rank of the covariance matrix $\text{Cov}(D_{n-r}^+(\varepsilon_{2t} \otimes \varepsilon_{2t}))$ (cf. the relations $\widetilde{\zeta}_{2t} = D_{n-r}^+ \widetilde{\underline{\zeta}}_{2t}$ and $\widetilde{\Gamma}(k) = D_{n-r}^+ \widetilde{\Gamma}(k) D_{n-r}^+$ discussed above). Because the matrix $R_2^{(2)}$ satisfies $R_2^{(2)'} R_2^{(2)} = I_{n-r}$ the rank of $R_2^{(2)} \text{Cov}(\varepsilon_{2t} \otimes \varepsilon_{2t}) R_2^{(2)'}$ is also $(n-r)(n-r+1)/2$ and because $\widetilde{R}_2^{(2)} \widetilde{\Gamma}(0) \widetilde{R}_2^{(2)'} = R_2^{(2)} \text{Cov}(\varepsilon_{2t} \otimes \varepsilon_{2t}) R_2^{(2)'} + o_p(1)$ we must have $\text{rk} \left(\widetilde{R}_2^{(2)} \widetilde{\Gamma}(0) \widetilde{R}_2^{(2)'} \right) \geq (n-r)(n-r+1)/2$ with probability

approaching one (cf. the proof of Lemma 1 in Andrews (1987)). On the other hand, because the structure of the matrix $\tilde{R}_2^{(2)} \tilde{\Gamma}(0) \tilde{R}_2^{(2)'}$ is similar to that of the matrix $R_2^{(2)} \text{Cov}(\varepsilon_{2t} \otimes \varepsilon_{2t}) R_2^{(2)'}$ it follows that the inequality $\text{rk}\left(\tilde{R}_2^{(2)} \tilde{\Gamma}(0) \tilde{R}_2^{(2)'}\right) \leq (n-r)(n-r+1)/2$ must hold. Thus, with probability approaching one the rank of the matrix $\tilde{R}_2^{(2)} \tilde{\Gamma}(0) \tilde{R}_2^{(2)'}$ equals $(n-r)(n-r+1)/2$, the rank of its probability limit $R_2^{(2)} \text{Cov}(\varepsilon_{2t} \otimes \varepsilon_{2t}) R_2^{(2)'}$. From Theorem 2 of Andrews (1987) we therefore find that

$$\begin{aligned} \left(\tilde{R}_2^{(2)} \tilde{\Gamma}(0) \tilde{R}_2^{(2)'}\right)^+ &= \left(R_2^{(2)} \text{Cov}(\varepsilon_{2t} \otimes \varepsilon_{2t}) R_2^{(2)'}\right)^+ + o_p(1) \\ &= R_2^{(2)} D_{n-r}^{+'} \left(\text{Cov}(D_{n-r}^+ (\varepsilon_{2t} \otimes \varepsilon_{2t}))\right)^{-1} D_{n-r}^+ R_2^{(2)'} + o_p(1). \end{aligned} \quad (27)$$

Here the latter equality can be justified by using the definition of the Moore-Penrose inverse (cf. the expression obtained for $\left(\tilde{R}_2^{(2)} \tilde{\Gamma}(0) \tilde{R}_2^{(2)'}\right)^+$ after (26)). The first equality also holds with the left hand side replaced by $\left(R_2^{(2)} \Gamma^*(0) R_2^{(2)'}\right)^+$.

As will become clear below, $T^{1/2} R_2^{(2)} \Gamma^*(k) R_2^{(2)'} = O_p(1)$ for $k > 0$. This in conjunction with equations (24), (26) and (27) yields

$$\begin{aligned} \mathcal{T}_2(K) &= T \sum_{k=1}^K \text{tr} \left(R_2^{(2)} \Gamma^*(k) R_2^{(2)'} \right)' \left(R_2^{(2)} \Gamma^*(0) R_2^{(2)'} \right)^+ \\ &\quad \times \left(R_2^{(2)} \Gamma^*(k) R_2^{(2)'} \right) \left(R_2^{(2)} \Gamma^*(0) R_2^{(2)'} \right)^+ + o_p(1) \\ &= T \sum_{k=1}^K \text{tr} \Gamma^*(k)' \Gamma^*(0)^{-1} \Gamma^*(k) \Gamma^*(0)^{-1} + o_p(1), \end{aligned}$$

where $\Gamma^*(k)$ is the sample serial covariance matrix of $D_{n-r}^+(\varepsilon_{2t} \otimes \varepsilon_{2t})$ at lag k and the latter equality follows from (27) with arguments similar to those used for (26). The limiting distribution of test statistic $\mathcal{T}_2(K)$ can be derived from this.

To complete the proof, denote $e_t = D_{n-r}^+(\varepsilon_{2t} \otimes \varepsilon_{2t})$ and $F^*(k) = \Gamma^*(0)^{-1/2} \Gamma^*(k) \Gamma^*(0)^{-1/2}$. By standard arguments it can first be seen that

$$\Gamma^*(k) = T^{-1} \sum_{t=k}^T (e_t - \mathbb{E}e_t) (e_{t-k} - \mathbb{E}e_{t-k})' + o_p(T^{-1/2}) \quad (k > 0).$$

This and a standard application of a martingale central limit theorem show that the rows of the matrix $T^{1/2} F^*(k)$ converge in distribution to independent standard normal

variates and, moreover, these limits are independent for $k = 1, \dots, K$. Thus, it follows that $\mathcal{T}_2(K) = T \sum_{k=1}^K \text{tr} F^*(k) F^*(k)' + o_p(1) \xrightarrow{d} \chi^2(K(n-r)^2(n-r+1)^2/4)$.

Appendix 2

Derivatives of the Gaussian log-likelihood function. For numerical optimization and estimation of the matrices Q_0 and Q_1 in the limiting distributions (16), (17) and (18) it is useful to have analytical expressions of the first and second derivatives of the Gaussian log-likelihood function. It will be convenient to consider the log-likelihood of a single observation which, by the expression of the conditional density function (12) and the subsequent discussion, can be written as

$$\begin{aligned} \bar{l}_t(\Phi, R_1, \delta) &= \log \det(\Phi) - \frac{1}{2} y_t' \Phi' \Phi y_t \\ &\quad - \sum_{i=1}^r \left(\frac{1}{2} \log v_{it} - \frac{1}{2} (v_{it}^{-1} - 1) (\rho_{1i}' \Phi y_t)^2 \right) \\ &\stackrel{def}{=} \bar{l}_{0t}(\Phi) + \sum_{i=1}^r \bar{l}_{it}(\Phi, \rho_{1i}, \delta_i). \end{aligned}$$

With this parameterization, equation (6) can be written as

$$v_{it} = (1 - \alpha_i - \beta_i) + \beta_i v_{i,t-1} + \alpha_i (\rho_{1i}' \Phi y_{t-1})^2 \quad i = 1, \dots, r.$$

Note also that the formulation of some of the subsequent results assume that the parameter matrix Φ is symmetric although the transpose sign is not suppressed from the notation. The parameters of the model are collected in the vector $\theta = [\phi' \ \rho_{11}' \ \dots \ \rho_{1r}' \ \delta_1' \ \dots \ \delta_r']'$ where $\phi = \text{vech}(\Phi)$.

By straightforward differentiation,

$$\partial \bar{l}_{0t}(\theta) / \partial \phi = D_n' \text{vec}(\Phi^{-1}) - \frac{1}{2} D_n' \text{vec}(y_t y_t' \Phi') - \frac{1}{2} D_n' \text{vec}(\Phi y_t y_t') \quad (28)$$

and, for $1 \leq i \leq r$,

$$\begin{aligned} \partial \bar{l}_{it}(\theta) / \partial \phi &= -\frac{1}{2} v_{it}^{-1} (\partial v_{it} / \partial \phi) + \frac{1}{2} v_{it}^{-2} (\rho_{1i}' \Phi y_t)^2 (\partial v_{it} / \partial \phi) \\ &\quad - \frac{1}{2} (v_{it}^{-1} - 1) [D_n' \text{vec}(y_t y_t' \Phi' \rho_{1i} \rho_{1i}') + D_n' \text{vec}(\rho_{1i} \rho_{1i}' \Phi y_t y_t')], \end{aligned} \quad (29)$$

where

$$\partial v_{it}/\partial\phi = \beta_i \partial v_{i,t-1}/\partial\phi + \alpha_i [D'_n \text{vec}(y_{t-1} y'_{t-1} \Phi' \rho_{1i} \rho'_{1i}) + D'_n \text{vec}(\rho_{1i} \rho'_{1i} \Phi y_{t-1} y'_{t-1})].$$

From these one obtains

$$\partial \bar{l}_t(\theta)/\partial\phi = \sum_{i=0}^r \partial \bar{l}_{it}(\theta)/\partial\phi. \quad (30)$$

Next note that $\bar{l}_{0t}(\theta)$ does not depend on the parameters ρ_{1i} and δ_i . Hence,

$$\begin{aligned} \partial \bar{l}_t(\theta)/\partial\rho_{1i} &= -\frac{1}{2} v_{it}^{-1} (\partial v_{it}/\partial\rho_{1i}) + \frac{1}{2} v_{it}^{-2} (\rho'_{1i} \Phi y_t)^2 (\partial v_{it}/\partial\rho_{1i}) \\ &\quad - (v_{it}^{-1} - 1) \Phi y_t y'_t \Phi' \rho_{1i}, \quad i = 1, \dots, r, \end{aligned} \quad (31)$$

where

$$\partial v_{it}/\partial\rho_{1i} = \beta_i (\partial v_{i,t-1}/\partial\rho_{1i}) + 2\alpha_i \Phi y_{t-1} y'_{t-1} \Phi' \rho_{1i}.$$

Finally,

$$\partial \bar{l}_t(\theta)/\partial\alpha_i = -\frac{1}{2} v_{it}^{-1} (\partial v_{it}/\partial\alpha_i) + \frac{1}{2} v_{it}^{-2} (\rho'_{1i} \Phi y_t)^2 (\partial v_{it}/\partial\alpha_i), \quad i = 1, \dots, r, \quad (32)$$

and

$$\partial \bar{l}_t(\theta)/\partial\beta_i = -\frac{1}{2} v_{it}^{-1} (\partial v_{it}/\partial\beta_i) + \frac{1}{2} v_{it}^{-2} (\rho'_{1i} \Phi y_t)^2 (\partial v_{it}/\partial\beta_i), \quad i = 1, \dots, r, \quad (33)$$

where

$$\partial v_{it}/\partial\alpha_i = -1 + \beta_i (\partial v_{i,t-1}/\partial\alpha_i) + (\rho'_{1i} \Phi y_{t-1})^2$$

and

$$\partial v_{it}/\partial\beta_i = -1 + \beta_i (\partial v_{i,t-1}/\partial\beta_i) + v_{i,t-1}.$$

From equations (28) - (33) one can form $\partial \bar{l}_t(\theta)/\partial\theta$ which can further be used to obtain an estimator for the matrix Q_0 in (16), (17) and (18).

Instead of the second partial derivatives of the log-likelihood function we use their conditional expectations which are somewhat simpler than the second partial derivatives themselves. By straightforward differentiation we first obtain

$$\partial^2 \bar{l}_{0t}(\theta)/\partial\phi\partial\phi' = -D'_n(\Phi^{-1} \otimes \Phi^{-1})D_n - \frac{1}{2}D'_n(I_n \otimes y_t y'_t)D_n - \frac{1}{2}D'_n(y_t y'_t \otimes I_n)D_n$$

and, hence,

$$\begin{aligned} \mathbb{E}_{t-1}(\partial^2 \bar{l}_{0t}(\theta) / \partial \phi \partial \phi') &= -D'_n(\Phi^{-1} \otimes \Phi^{-1})D_n - \frac{1}{2}D'_n(I_n \otimes WH_tW')D_n \quad (34) \\ &\quad - \frac{1}{2}D'_n(WH_tW' \otimes I_n)D_n. \end{aligned}$$

For $1 \leq i \leq r$ we have,

$$\begin{aligned} \partial^2 \bar{l}_{it}(\theta) / \partial \phi \partial \phi' &= \frac{1}{2}v_{it}^{-2}(\partial v_{it} / \partial \phi)(\partial v_{it} / \partial \phi') - \frac{1}{2}v_{it}^{-1}(\partial^2 v_{it} / \partial \phi \partial \phi') \\ &\quad + \frac{1}{2}v_{it}^{-2}(\rho'_{1i}\Phi y_t)^2(\partial^2 v_{it} / \partial \phi \partial \phi') - v_{it}^{-3}(\rho'_{1i}\Phi y_t)^2(\partial v_{it} / \partial \phi)(\partial v_{it} / \partial \phi') \\ &\quad + \frac{1}{2}v_{it}^{-2}(\partial v_{it} / \partial \phi)[D'_n \text{vec}(y_t y'_t \Phi' \rho_{1i} \rho'_{1i}) + D'_n \text{vec}(\rho_{1i} \rho'_{1i} \Phi y_t y'_t)]' \\ &\quad + \frac{1}{2}v_{it}^{-2}[D'_n \text{vec}(y_t y'_t \Phi' \rho_{1i} \rho'_{1i}) + D'_n \text{vec}(\rho_{1i} \rho'_{1i} \Phi y_t y'_t)](\partial v_{it} / \partial \phi') \\ &\quad - \frac{1}{2}(v_{it}^{-1} - 1)[D'_n(\rho_{1i} \rho'_{1i} \otimes y_t y'_t)D_n + D'_n(y_t y'_t \otimes \rho_{1i} \rho'_{1i})D_n]. \end{aligned}$$

Since $\mathbb{E}_{t-1}((\rho'_{1i}\Phi y_t)^2) = v_{it}$ the second and third terms on the right hand side cancel and the first and fourth terms can be combined when conditional expectations are taken on both sides. Furthermore, $\mathbb{E}'_{t-1}(y_t y'_t \Phi' \rho_{1i} \rho'_{1i}) = v_{it} w_i \rho'_{1i} = v_{it} \Phi^{-1} \rho_{1i} \rho'_{1i}$. Hence, we find from the preceding equation that, for $1 \leq i \leq r$,

$$\begin{aligned} &\mathbb{E}_{t-1}(\partial^2 \bar{l}_{it}(\theta) / \partial \phi \partial \phi') \quad (35) \\ &= -\frac{1}{2}v_{it}^{-2}(\partial v_{it} / \partial \phi)(\partial v_{it} / \partial \phi') \\ &\quad + \frac{1}{2}v_{it}^{-1}(\partial v_{it} / \partial \phi) \left[D'_n \text{vec}(\Phi^{-1} \rho_{1i} \rho'_{1i}) + D'_n \text{vec}(\rho_{1i} \rho'_{1i} \Phi^{-1}) \right]' \\ &\quad + \frac{1}{2}v_{it}^{-1} \left[D'_n \text{vec}(\Phi^{-1} \rho_{1i} \rho'_{1i}) + D'_n \text{vec}(\rho_{1i} \rho'_{1i} \Phi^{-1}) \right] (\partial v_{it} / \partial \phi') \\ &\quad - \frac{1}{2}(v_{it}^{-1} - 1)[D'_n(\rho_{1i} \rho'_{1i} \otimes WH_tW')D_n + D'_n(WH_tW' \otimes \rho_{1i} \rho'_{1i})D_n]. \end{aligned}$$

From the above results one can form

$$\mathbb{E}_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial \phi \partial \phi') = \sum_{i=0}^r \mathbb{E}_{t-1}(\partial^2 \bar{l}_{it}(\theta) / \partial \phi \partial \phi'). \quad (36)$$

Recall that $\bar{l}_{0t}(\theta)$ does not depend on the parameters ρ_{1i} and δ_i . Hence,

$$\begin{aligned}\partial^2 \bar{l}_t(\theta) / \partial \rho_{1i} \partial \rho'_{1i} &= \frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial \rho_{1i}) (\partial v_{it} / \partial \rho'_{1i}) - \frac{1}{2} v_{it}^{-1} (\partial^2 v_{it} / \partial \rho_{1i} \partial \rho'_{1i}) \\ &\quad + \frac{1}{2} v_{it}^{-2} (\rho'_{1i} \Phi y_t)^2 (\partial^2 v_{it} / \partial \rho_{1i} \partial \rho'_{1i}) - v_{it}^{-3} (\rho'_{1i} \Phi y_t)^2 (\partial v_{it} / \partial \rho_{1i}) (\partial v_{it} / \partial \rho'_{1i}) \\ &\quad - (v_{it}^{-1} - 1) \Phi y_t y'_t \Phi' + v_{it}^{-2} \Phi y_t y'_t \Phi' \rho_{1i} (\partial v_{it} / \partial \rho'_{1i}).\end{aligned}$$

Here the second and third terms on the right hand side cancel and the first and fourth ones can be combined when conditional expectations are taken on both sides. Also, $E_{t-1}(\Phi y_t y'_t \Phi') = RH_t R'$ and $E_{t-1}(\Phi y_t y'_t \Phi' \rho_{1i}) = v_{it} \rho_{1i}$ and we can conclude that

$$\begin{aligned}E_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial \rho_{1i} \partial \rho'_{1i}) &= -\frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial \rho_{1i}) (\partial v_{it} / \partial \rho'_{1i}) - (v_{it}^{-1} - 1) RH_t R' \\ &\quad + v_{it}^{-1} \rho_{1i} (\partial v_{it} / \partial \rho'_{1i}).\end{aligned}\quad (37)$$

Let c_i stand for either α_i or β_i . Then,

$$\begin{aligned}\partial^2 \bar{l}_t(\theta) / \partial c_i^2 &= \frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial c_i)^2 - \frac{1}{2} v_{it}^{-1} (\partial^2 v_{it} / \partial c_i^2) \\ &\quad + \frac{1}{2} v_{it}^{-2} (\rho'_{1i} \Phi y_t)^2 (\partial^2 v_{it} / \partial c_i^2) - v_{it}^{-3} (\rho'_{1i} \Phi y_t)^2 (\partial v_{it} / \partial c_i)^2\end{aligned}$$

from which one obtains

$$E_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial c_i^2) = -\frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial c_i)^2, \quad c_i = \alpha_i, \beta_i, \quad i = 1, \dots, r, \quad (38)$$

and, similarly,

$$E_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial \alpha_i \beta_i) = -\frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial \alpha_i) (\partial v_{it} / \partial \beta_i), \quad i = 1, \dots, r. \quad (39)$$

Next consider

$$\begin{aligned}\partial^2 \bar{l}_t(\theta) / \partial \rho_{1i} \partial \phi' &= \frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial \rho_{1i}) (\partial v_{it} / \partial \phi') - \frac{1}{2} v_{it}^{-1} (\partial^2 v_{it} / \partial \rho_{1i} \partial \phi') \\ &\quad + \frac{1}{2} v_{it}^{-2} (\rho'_{1i} \Phi y_t)^2 (\partial^2 v_{it} / \partial \rho_{1i} \partial \phi') - v_{it}^{-3} (\rho'_{1i} \Phi y_t)^2 (\partial v_{it} / \partial \rho_{1i}) (\partial v_{it} / \partial \phi') \\ &\quad + v_{it}^{-2} \Phi y_t y'_t \Phi' \rho_{1i} (\partial v_{it} / \partial \phi') \\ &\quad - (v_{it}^{-1} - 1) [D'_n(y_t y'_t \Phi' \rho_{1i} \otimes I_n) + D'_n(\rho_{1i} \otimes y_t y'_t \Phi')]' .\end{aligned}$$

The second and third terms on the right hand side cancel and the first and fourth ones can be combined when conditional expectations are taken on both sides. Since we also have $\mathbb{E}_{t-1}(\Phi y_t y_t' \Phi') = RH_t R'$, $\mathbb{E}_{t-1}(y_t y_t' \Phi' \rho_{1i}) = v_{it} \Phi^{-1} \rho_{1i}$ and $\mathbb{E}_{t-1}(y_t y_t' \Phi') = \Phi^{-1} RH_t R'$ we find that

$$\begin{aligned} \mathbb{E}_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial \rho_{1i} \partial \phi') &= -\frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial \rho_{1i}) (\partial v_{it} / \partial \phi') + v_{it}^{-1} \rho_{1i} (\partial v_{it} / \partial \phi') \\ &\quad - (v_{it}^{-1} - 1) [D'_n (v_{it} \Phi^{-1} \rho_{1i} \otimes I_n) + D'_n (\rho_{1i} \otimes \Phi^{-1} RH_t R')] \end{aligned} \quad (40)$$

for $i = 1, \dots, r$.

Further, for $c_i = \alpha_i, \beta_i$,

$$\begin{aligned} \partial^2 \bar{l}_t(\theta) / \partial c_i \partial \phi' &= \frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial c_i) (\partial v_{it} / \partial \phi') - \frac{1}{2} v_{it}^{-1} (\partial^2 v_{it} / \partial c_i \partial \phi') \\ &\quad + \frac{1}{2} v_{it}^{-2} (\rho'_{1i} \Phi y_t)^2 (\partial^2 v_{it} / \partial c_i \partial \phi') - v_{it}^{-3} (\rho'_{1i} \Phi y_t)^2 (\partial v_{it} / \partial c_i) (\partial v_{it} / \partial \phi') \\ &\quad + \frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial c_i) [D'_n \text{vec}(y_t y_t' \Phi' \rho_{1i} \rho'_{1i}) + D'_n \text{vec}(\rho_{1i} \rho'_{1i} \Phi y_t y_t')] \end{aligned}$$

This in conjunction with arguments similar to those already used give

$$\begin{aligned} \mathbb{E}_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial c_i \partial \phi') &= -\frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial c_i) (\partial v_{it} / \partial \phi') \\ &\quad + \frac{1}{2} v_{it}^{-1} (\partial v_{it} / \partial c_i) \left[D'_n \text{vec}(\Phi^{-1} \rho_{1i} \rho'_{1i}) + D'_n \text{vec}(\rho_{1i} \rho'_{1i} \Phi^{-1}) \right] \end{aligned} \quad (41)$$

for $c_i = \alpha_i, \beta_i$ and $i = 1, \dots, r$.

Next note that

$$\begin{aligned} \partial^2 \bar{l}_t(\theta) / \partial c_i \partial \rho'_{1i} &= \frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial c_i) (\partial v_{it} / \partial \rho'_{1i}) - \frac{1}{2} v_{it}^{-1} (\partial^2 v_{it} / \partial c_i \partial \rho'_{1i}) \\ &\quad + \frac{1}{2} v_{it}^{-2} (\rho'_{1i} \Phi y_t)^2 (\partial^2 v_{it} / \partial c_i \partial \rho'_{1i}) - v_{it}^{-3} (\rho'_{1i} \Phi y_t)^2 (\partial v_{it} / \partial c_i) (\partial v_{it} / \partial \rho'_{1i}) \\ &\quad + v_{it}^{-2} (\partial v_{it} / \partial c_i) \rho'_{1i} \Phi y_t y_t' \Phi', \end{aligned}$$

which yields

$$\mathbb{E}_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial c_i \partial \rho'_{1i}) = -\frac{1}{2} v_{it}^{-2} (\partial v_{it} / \partial c_i) (\partial v_{it} / \partial \rho'_{1i}) + v_{it}^{-1} (\partial v_{it} / \partial c_i) \rho'_{1i} \quad (42)$$

for $c_i = \alpha_i, \beta_i$, $i = 1, \dots, r$.

Finally,

$$\partial^2 \bar{l}_t(\theta) / \partial \rho_{1i} \partial \rho'_{1j} = 0, \quad i \neq j, \quad (43)$$

$$\partial^2 \bar{l}_t(\theta) / \partial \rho_{1i} \partial c_j = 0, \quad i \neq j, \quad (44)$$

and

$$\partial^2 \bar{l}_t(\theta) / \partial c_i \partial c_j = 0, \quad i \neq j, \quad (45)$$

where $c_i = \alpha_j, \beta_i$ as before.

Using the expressions given in equations (34) - (45) one can form the blocks on and below the diagonal of the matrix $E_{t-1}(\partial^2 \bar{l}_t(\theta) / \partial \theta \partial \theta')$ and furthermore obtain an estimator for the matrix Q_1 in (16), (17) and (18).

First derivatives of the log-likelihood function of the mixture Gaussian distribution. The log-likelihood of a single observation is

$$\bar{l}_t(\vartheta) = \log \det(B) - \frac{1}{2} \sum_{i=1}^r \log v_{it} + \log L_t(\vartheta).$$

We start with calculating

$$\begin{aligned} \partial \bar{l}_t(\vartheta) / \partial b &= \text{vec}(B^{-1'}) - \frac{1}{2} \sum_{i=1}^r v_{it}^{-1} (\partial v_{it} / \partial b) \\ &\quad + \frac{1}{L_t(\theta)} (p \partial L_{1t}(\vartheta_1) / \partial b + (1-p) \partial L_{2t}(\vartheta) / \partial b), \end{aligned}$$

where

$$\partial v_{it} / \partial b = [0' \ \dots \ 0' \ \partial v_{it} / \partial b'_{1i} \ 0' \ \dots \ 0']'$$

with

$$\partial v_{it} / \partial b_{1i} = \beta_i (\partial v_{i,t-1} / \partial b_{1i}) + 2\alpha_i y_{t-1} y'_{t-1} b_{1i}.$$

Note that all component vectors in $\partial v_{it} / \partial b$ are of order $n \times 1$ and the nonzero component is in the i th position. (Recall that b_{1i} signifies the i th column of the matrix B_1 or, equivalently, the i th column of the matrix B .) The vector $\partial v_{it} / \partial b$ will also appear in the following two partial derivatives needed to compute $\partial \bar{l}_t(\theta) / \partial b$.

From the definition of $L_{1t}(\vartheta)$ we find that

$$\begin{aligned}\partial L_{1t}(\vartheta)/\partial b &= -\frac{1}{2}L_{1t}(\vartheta)\partial\text{tr}(BH_t^{-1}\Psi_1(p)B'y_t y_t')/\partial b \\ &= -\frac{1}{2}L_{1t}(\vartheta)\left[2\text{vec}(y_t y_t' BH_t^{-1}\Psi_1(p)) - \sum_{i=1}^r (b'_{1i}y_t)^2 (p + (1-p)\psi_i) v_{it}^{-2} \frac{\partial v_{it}}{\partial b}\right],\end{aligned}$$

where the first term in the brackets is obtained by keeping B inside H_t in the previous expression fixed while the latter term results when B and B' in the previous expression are kept fixed. Similarly

$$\partial L_{2t}(\vartheta)/\partial b = -\frac{1}{2}L_{2t}(\vartheta)\left[2\text{vec}(y_t y_t' BH_t^{-1}\Psi_2(p)) - \sum_{i=1}^r (b'_{1i}y_t)^2 (p\psi_i^{-1} + (1-p)) v_{it}^{-2} \frac{\partial v_{it}}{\partial b}\right].$$

Now we have all the ingredients needed to compute $\partial \bar{l}_t(\theta)/\partial b$.

Next consider the partial derivative $\partial \bar{l}_t(\theta)/\partial \delta$ and recall that $\delta = [\delta'_1 \ \dots \ \delta'_r]'$ with $\delta_i = [\alpha_i \ \beta_i]'$ ($i = 1, \dots, r$). From the expression of $\bar{l}_t(\vartheta)$ we find that

$$\partial \bar{l}_t(\vartheta)/\partial \alpha_i = -\frac{1}{2}v_{it}^{-1}(\partial v_{it}/\partial \alpha_i) + \frac{1}{L_t(\vartheta)}(p\partial L_{1t}(\vartheta)/\partial \alpha_i + (1-p)\partial L_{2t}(\vartheta)/\partial \alpha_i).$$

Here $\partial v_{it}/\partial \alpha_i$ is as in (32) with $\rho'_{1i}\Phi$ replaced by b'_{1i} and

$$\begin{aligned}\partial L_{1t}(\vartheta)/\partial \alpha_i &= -\frac{1}{2}L_{1t}(\vartheta)(b'_{1i}y_t)^2 (p + (1-p)\psi_i) \frac{\partial}{\partial \alpha_i} v_{it}^{-1} \\ &= \frac{1}{2}L_{1t}(\vartheta)(b'_{1i}y_t)^2 (p + (1-p)\psi_i) v_{it}^{-2}(\partial v_{it}/\partial \alpha_i).\end{aligned}$$

Similarly,

$$\partial L_{2t}(\vartheta)/\partial \alpha_i = \frac{1}{2}L_{2t}(\vartheta)(b'_{1i}y_t)^2 (p\psi_i^{-1} + (1-p)) v_{it}^{-2}(\partial v_{it}/\partial \alpha_i).$$

Also,

$$\partial \bar{l}_t(\vartheta)/\partial \beta_i = -\frac{1}{2}v_{it}^{-1}(\partial v_{it}/\partial \beta_i) + \frac{1}{L_t(\vartheta)}(p\partial L_{1t}(\vartheta)/\partial \beta_i + (1-p)\partial L_{2t}(\vartheta)/\partial \beta_i),$$

where $\partial v_{it}/\partial \beta_i$ is as in (33) and $L_{1t}(\vartheta)/\partial \beta_i$ and $\partial L_{2t}(\vartheta)/\partial \beta_i$ are obtained as $L_{1t}(\vartheta)/\partial \alpha_i$ and $\partial L_{2t}(\vartheta)/\partial \alpha_i$, respectively, except that $\partial v_{it}/\partial \alpha_i$ is replaced by $\partial v_{it}/\partial \beta_i$. In other words,

$$\partial L_{1t}(\vartheta)/\partial \beta_i = \frac{1}{2}L_{1t}(\vartheta)(b'_{1i}y_t)^2 (p\psi_i^{-1} + (1-p)) v_{it}^{-2}(\partial v_{it}/\partial \beta_i)$$

and

$$\partial L_{2t}(\vartheta) / \partial \beta_i = \frac{1}{2} L_{2t}(\vartheta) (b'_{1i} y_t)^2 (p + (1-p)\psi_i) v_{it}^{-2} (\partial v_{it} / \partial \beta_i).$$

The above expressions can be used to form $\partial \bar{l}_t(\vartheta) / \partial \delta_i = [\partial \bar{l}_t(\vartheta) / \partial \alpha_i \quad \partial \bar{l}_t(\vartheta) / \partial \beta_i]'$ and furthermore $\partial \bar{l}_t(\vartheta) / \partial \delta = [\partial \bar{l}_t(\vartheta) / \partial \delta'_1 \quad \cdots \quad \partial \bar{l}_t(\vartheta) / \partial \delta'_r]'$.

For the partial derivative $\partial \bar{l}_t(\vartheta) / \partial \psi$ we have

$$\partial \bar{l}_t(\vartheta) / \partial \psi_i = \frac{p}{L_t(\theta)} \partial L_{1t}(\vartheta) / \partial \psi_i + \frac{1-p}{L_t(\theta)} \partial L_{2t}(\vartheta) / \partial \psi_i, \quad i = 1, \dots, n.$$

By the definitions,

$$\begin{aligned} \partial L_{1t}(\vartheta) / \partial \psi_i &= \exp \left\{ -\frac{1}{2} y'_t B H_t^{-1} \Psi_1(p) B' y_t \right\} \frac{\partial}{\partial \psi_i} \left(\prod_{j=1}^n (p + (1-p)\psi_j)^{1/2} \right) \\ &\quad - \frac{1}{2} L_{1t}(\vartheta) \partial (y'_t B H_t^{-1} \Psi_1(p) B' y_t) / \partial \psi_i. \end{aligned}$$

Let b_j and h_{jt} denote the j th column of the matrix B and the j th diagonal element of the matrix H_t , respectively. On the right hand side of the preceding equality we then have

$$\begin{aligned} \partial (y'_t B H_t^{-1} \Psi_1(p) B' y_t) / \partial \psi_i &= \frac{\partial}{\partial \psi_i} \sum_{j=1}^n (y'_t b_j)^2 h_{jt}^{-1} (p + (1-p)\psi_j) \\ &= (y'_t b_i)^2 h_{it}^{-1} (1-p) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \psi_i} \left(\prod_{j=1}^n (p + (1-p)\psi_j)^{1/2} \right) &= \frac{\partial}{\partial \psi_i} (p + (1-p)\psi_i)^{1/2} \prod_{\substack{j=1 \\ j \neq i}}^n (p + (1-p)\psi_j)^{1/2} \\ &= \frac{1-p}{2} (p + (1-p)\psi_i)^{-1/2} \prod_{\substack{j=1 \\ j \neq i}}^n (p + (1-p)\psi_j)^{1/2} \\ &= \frac{1-p}{2} (p + (1-p)\psi_i)^{-1} \prod_{j=1}^n (p + (1-p)\psi_j)^{1/2} \\ &= \det(\Psi_1(p))^{1/2} \frac{1-p}{2} (p + (1-p)\psi_i)^{-1}. \end{aligned}$$

Thus, we can conclude that

$$\partial L_{1t}(\vartheta) / \partial \psi_i = \frac{1-p}{2} L_{1t}(\vartheta) \left[(p + (1-p)\psi_i)^{-1} - (y'_t b_i)^2 h_{it}^{-1} \right], \quad i = 1, \dots, n.$$

As for $\partial L_{2t}(\vartheta) / \partial \psi_i$, one similarly obtains

$$\begin{aligned} \partial L_{2t}(\vartheta) / \partial \psi_i &= \exp \left\{ -\frac{1}{2} y_t' B H_t^{-1} \Psi_2(p) B' y_t \right\} \frac{\partial}{\partial \psi_i} \left(\prod_{j=1}^n (p \psi_j^{-1} + (1-p))^{1/2} \right) \\ &\quad - \frac{1}{2} L_{2t}(\vartheta) \partial (y_t' B H_t^{-1} \Psi_2(p) B' y_t) / \partial \psi_i, \end{aligned}$$

where

$$\begin{aligned} \partial (y_t' B H_t^{-1} \Psi_2(p) B' y_t) / \partial \psi_i &= \frac{\partial}{\partial \psi_i} \sum_{j=1}^n (y_t' b_j)^2 h_{jt}^{-1} (p \psi_j^{-1} + (1-p)) \\ &= -p \psi_i^{-2} (y_t' b_i)^2 h_{it}^{-1} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \psi_i} \left(\prod_{j=1}^n (p \psi_j^{-1} + (1-p))^{1/2} \right) &= \frac{\partial}{\partial \psi_i} (p \psi_i^{-1} + (1-p))^{1/2} \prod_{\substack{j=1 \\ j \neq i}}^n (p \psi_j^{-1} + (1-p))^{1/2} \\ &= -\frac{p}{2 \psi_i^2} (p \psi_i^{-1} + (1-p))^{-1/2} \prod_{\substack{j=1 \\ j \neq i}}^n (p \psi_j^{-1} + (1-p))^{1/2} \\ &= -\frac{p}{2 \psi_i^2} (p \psi_i^{-1} + (1-p))^{-1} \prod_{j=1}^n (p \psi_j^{-1} + (1-p))^{1/2} \\ &= -\det(\Psi_2(p))^{1/2} \frac{p}{2 \psi_i^2} (p \psi_i^{-1} + (1-p))^{-1}. \end{aligned}$$

Hence,

$$\partial L_{2t}(\vartheta) / \partial \psi_i = -\frac{p}{2 \psi_i^2} L_{2t}(\vartheta) \left[(p \psi_i^{-1} + (1-p))^{-1} - (y_t' b_i)^2 h_{it}^{-1} \right], \quad i = 1, \dots, n.$$

Now all the ingredients needed to obtain $\partial \bar{l}_t(\vartheta) / \partial \psi$ are available.

Finally, to obtain the partial derivative $\partial \bar{l}_t(\vartheta) / \partial p$ we first note that

$$\partial \bar{l}_t(\theta) / \partial p = \frac{1}{L_t(\theta)} (p \partial L_{1t}(\vartheta) / \partial p - p \partial L_{2t}(\vartheta) / \partial p).$$

Here

$$\begin{aligned} \partial L_{1t}(\vartheta) / \partial p &= \exp \left\{ -\frac{1}{2} y_t' B H_t^{-1} \Psi_1(p) B' y_t \right\} \frac{\partial}{\partial p} \left(\prod_{j=1}^n (p + (1-p) \psi_j)^{1/2} \right) \\ &\quad - \frac{1}{2} L_{1t}(\vartheta) \frac{\partial}{\partial p} y_t' B H_t^{-1} \Psi_1(p) B' y_t. \end{aligned}$$

For the first term on the right hand side,

$$\begin{aligned}
\frac{\partial}{\partial p} \left(\prod_{j=1}^n (p + (1-p)\psi_j)^{1/2} \right) &= \frac{\partial}{\partial p} \exp \left\{ \frac{1}{2} \sum_{j=1}^n \log (p + (1-p)\psi_j) \right\} \\
&= \exp \left\{ \frac{1}{2} \sum_{j=1}^n \log (p + (1-p)\psi_j) \right\} \\
&\quad \times \frac{\partial}{\partial p} \left(\frac{1}{2} \sum_{j=1}^n \log (p + (1-p)\psi_j) \right) \\
&= \prod_{j=1}^n (p + (1-p)\psi_j)^{1/2} \frac{1}{2} \sum_{j=1}^n \frac{1 - \psi_j}{p + (1-p)\psi_j} \\
&= \det(\Psi_1(p))^{1/2} \frac{1}{2} \sum_{j=1}^n \frac{1 - \psi_j}{p + (1-p)\psi_j}.
\end{aligned}$$

Further,

$$\begin{aligned}
\frac{\partial}{\partial p} y_t' B H_t^{-1} \Psi_1(p) B' y_t &= \frac{\partial}{\partial p} \sum_{j=1}^n (y_t' b_j)^2 h_{jt}^{-1} (p + (1-p)\psi_j) \\
&= \sum_{j=1}^n (y_t' b_j)^2 h_{jt}^{-1} (1 - \psi_j).
\end{aligned}$$

Combining the above derivations gives

$$\begin{aligned}
\partial L_{1t}(\vartheta) / \partial p &= \frac{1}{2} L_{1t}(\vartheta) \sum_{j=1}^n \frac{1 - \psi_j}{p + (1-p)\psi_j} \\
&\quad - \frac{1}{2} L_{1t}(\vartheta) \sum_{j=1}^n (y_t' b_j)^2 h_{jt}^{-1} (1 - \psi_j).
\end{aligned}$$

In a similar way,

$$\begin{aligned}
\partial L_{2t}(\vartheta) / \partial p &= \exp \left\{ -\frac{1}{2} y_t' B H_t^{-1} \Psi_2(p) B' y_t \right\} \frac{\partial}{\partial p} \left(\prod_{j=1}^n (p\psi_j^{-1} + (1-p))^{1/2} \right) \\
&\quad - \frac{1}{2} L_{2t}(\vartheta) \frac{\partial}{\partial p} y_t' B H_t^{-1} \Psi_2(p) B' y_t
\end{aligned}$$

and on the right hand side

$$\begin{aligned}
\frac{\partial}{\partial p} \left(\prod_{j=1}^n (p\psi_j^{-1} + (1-p))^{1/2} \right) &= \frac{\partial}{\partial p} \exp \left\{ \frac{1}{2} \sum_{j=1}^n \log (p\psi_j^{-1} + (1-p)) \right\} \\
&= \exp \left\{ \frac{1}{2} \sum_{j=1}^n \log (p\psi_j^{-1} + (1-p)) \right\} \\
&\quad \times \frac{\partial}{\partial p} \left(\frac{1}{2} \sum_{j=1}^n \log (p\psi_j^{-1} + (1-p)) \right) \\
&= \prod_{j=1}^n (p\psi_j^{-1} + (1-p))^{1/2} \frac{1}{2} \sum_{j=1}^n \frac{\psi_j^{-1} - 1}{p\psi_j^{-1} + (1-p)} \\
&= \frac{1}{2} \det(\Psi_2(p))^{1/2} \sum_{j=1}^n \frac{\psi_j^{-1} - 1}{p\psi_j^{-1} + (1-p)}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial p} y_t' B H_t^{-1} \Psi_2(p) B' y_t &= \frac{\partial}{\partial p} \sum_{j=1}^n (y_t' b_j)^2 h_{jt}^{-1} (p\psi_j^{-1} + (1-p)) \\
&= \sum_{j=1}^n (y_t' b_j)^2 h_{jt}^{-1} (\psi_j^{-1} - 1).
\end{aligned}$$

Thus,

$$\begin{aligned}
\partial L_{2t}(\vartheta) / \partial p &= \frac{1}{2} L_{2t}(\vartheta) \sum_{j=1}^n \frac{\psi_j^{-1} - 1}{p\psi_j^{-1} + (1-p)} \\
&\quad - \frac{1}{2} L_{2t}(\vartheta) \sum_{j=1}^n (y_t' b_j)^2 h_{jt}^{-1} (\psi_j^{-1} - 1).
\end{aligned}$$

References

Aitchison, J. and S.D. Silvey (1959). Maximum-likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics* 29, 813-828.

Alexander, C. (2001). A primer on orthogonal GARCH model. Mimeo. ISMA Centre.

- Andrews, D.W.K. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory* 3, 348-358.
- Comte, F. and O. Lieberman (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 1, 61-84.
- Bauwens, L., S. Laurent, and J.V.K. Rombouts (2003). Multivariate GARCH models: A survey. *Journal of Applied Econometrics* (forthcoming).
- Davidson, J. (2000). *Econometric Theory*. Blackwell Publishers Ltd, Oxford.
- Engle, R.F. and K.F. Kroner (1995). Multivariate simultaneous generalized GARCH. *Econometric Theory* 11, 122-150.
- Engle, R.F., V.K. Ng, and M. Rotschild (1990). Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics* 45, 213-237.
- Franses, P.H., D. van Dijk, and A. Lucas (2004). Short patches of outliers, ARCH and volatility modelling. *Applied Financial Economics* 14, 221-231.
- Hafner, C.M. and H. Herwartz (2003). Analytical quasi maximum likelihood inference in multivariate volatility models, *Econometric Institute Report 2003/21*, Erasmus Universiteit Rotterdam.
- Horn, R.A. and C.H. Johnson (1985). *Matrix Analysis*. Cambridge University Press, Cambridge.
- Kawakatsu, H. (2003). Cholesky factor GARCH. Mimeo. Quantitative Micro Software.
- Klaassen, F. (2000). Have exchange rates become more closely tied? Evidence from a new multivariate GARCH model. Discussion Paper. University of Amsterdam.

Ling, S. and W.K. Li (1997). Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis* 18, 447-464.

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons. Chichester.

Mencía, F.J., and E. Sentana (2004). Estimation and testing of dynamic models with generalised hyperbolic innovations. CEMFI Working Paper No. 0411.

Silvey, S.D. (1959). The Lagrangian multiplier test. *Annals of Mathematical Statistics* 30, 389-407.

van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* 17, 549-564.

Vrontos, I.D., P. Dellaportas and D.N. Politis (2003). A full-factor multivariate GARCH model. *Econometrics Journal* 6, 312-334.

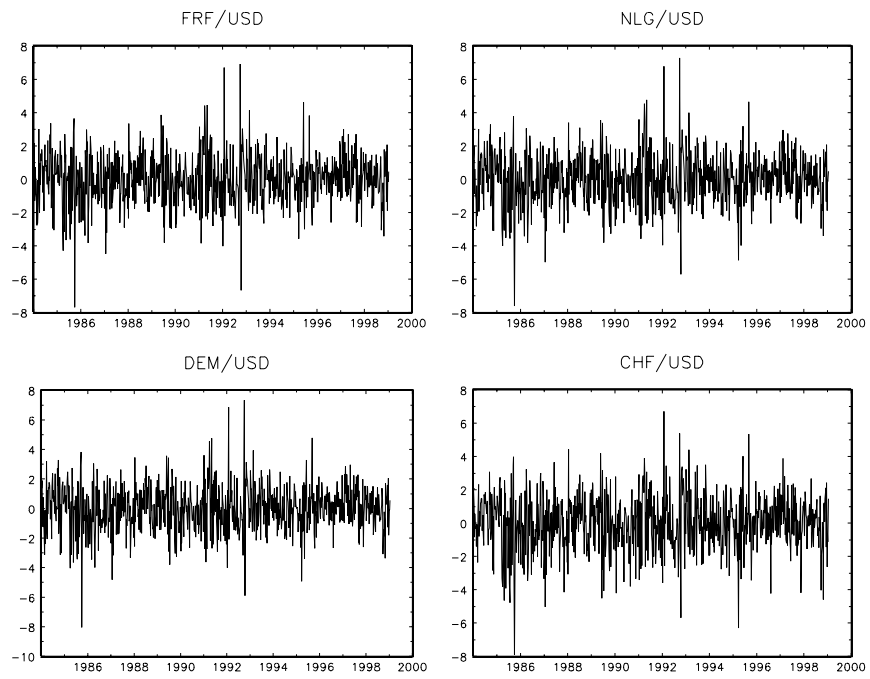


Figure 1: The exchange rate return series.

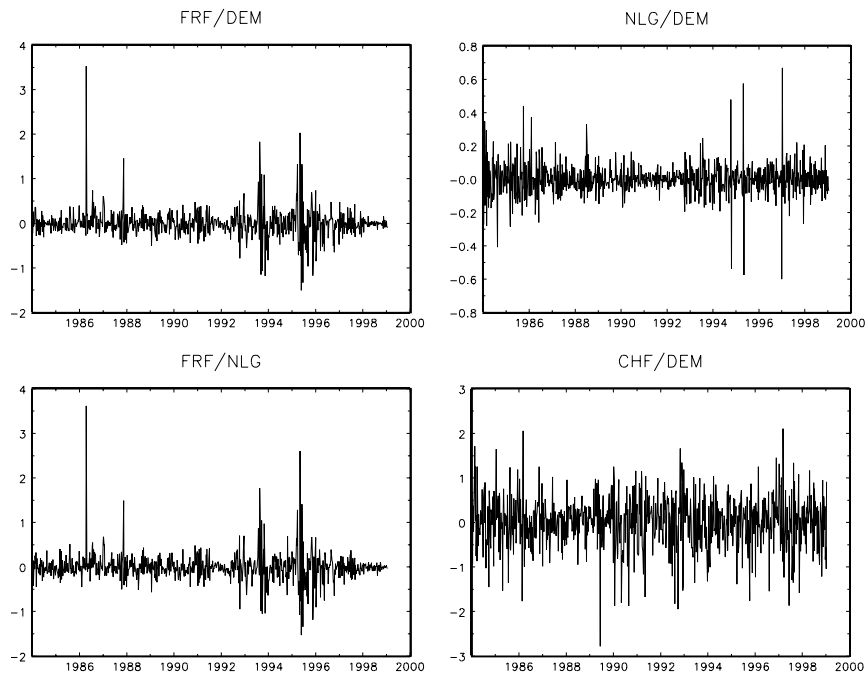


Figure 2: The implied cross rate returns computed from the exchange rates against the U.S: dollar.

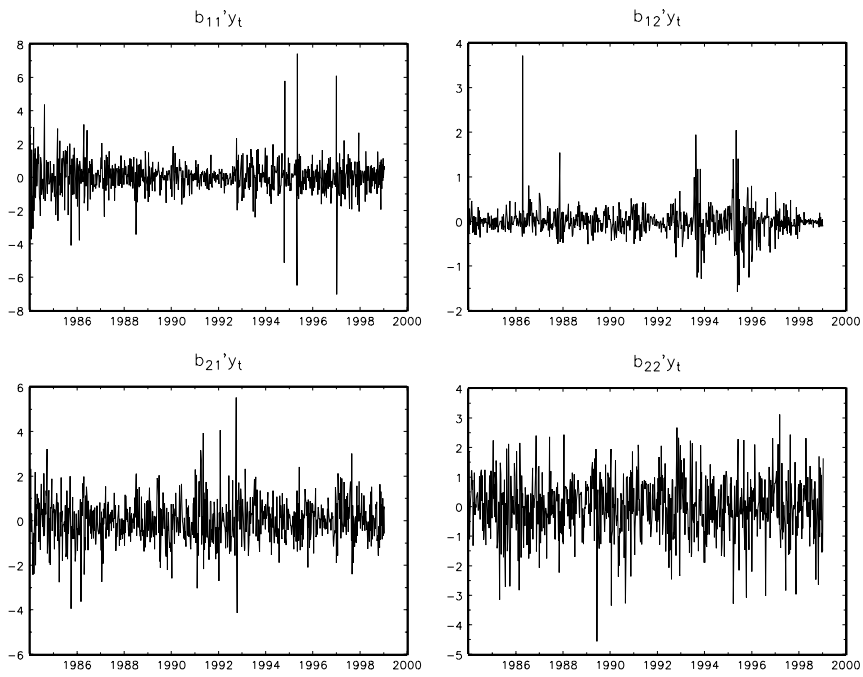


Figure 3: The two conditionally heteroskedastic factors (upper panel) and homoskedastic linear combinations (lower panel) under normality.

Table 1: Tests for the order of conditional heteroskedasticity.

Model	Hypothesis	\mathcal{T}_1	\mathcal{T}_2
Gaussian	$r = 1$	23.828 (1.05e-6)	140.269 (3.08e-14)
	$r = 2$	2.677 (0.102)	16.506 (0.057)
Mixture of Gaussians	$r = 1$	4.997 (0.025)	48.732 (0.076)
	$r = 2$	2.421 (0.120)	15.017 (0.090)

The figures in parentheses are marginal significance levels.

Table 2: Estimation results of the model with two conditionally heteroskedastic factors under normality.

Parameter	Estimate	Std.err.	Parameter	Estimate	Std.err.
α_1	0.195	0.076	α_2	0.669	0.129
β_1	0.444	0.650	β_2	0.314	0.137
b_{11}	0.604	0.348	b_{12}	1.060	0.007
	-10.511	0.818		0.187	0.668
	10.079	0.558		-1.262	0.592
	-0.216	0.221		0.015	0.063
w_{11}	0.085	0.163	w_{12}	-0.262	0.060
	-0.025	0.162		-1.139	0.090
	0.067	0.161		-1.193	0.068
	-0.065	0.162		-0.994	0.045

Table 3: Estimation results of the one-factor mixture model.

α_1	0.302		ψ_1	0.144
	(0.065)			(0.016)
β_1	0.633		ψ_2	0.048
	(0.060)			(0.005)
p	0.129		ψ_3	0.317
	(0.017)			(0.042)
			ψ_4	0.466
				(0.078)
B	-0.005	3.260	0.458	-0.119
	(0.067)	(0.044)	(0.047)	(0.035)
	-7.238	-1.494	0.898	-1.803
	(0.068)	(0.045)	(0.048)	(0.035)
	7.243	-1.708	-0.003	0.638
	(0.069)	(0.045)	(0.048)	(0.036)
	-0.066	0.021	-0.743	1.524
	(0.077)	(0.050)	(0.046)	(0.047)
W	-0.025	0.139	1.361	0.661
	(0.018)	(0.025)	(0.413)	(0.254)
	-0.099	-0.171	1.388	0.675
	(0.018)	(0.027)	(0.427)	(0.263)
	0.038	-0.171	1.398	0.686
	(0.018)	(0.027)	(0.429)	(0.264)
	-0.135	-0.119	1.164	1.220
	(0.020)	(0.029)	(0.456)	(0.284)

The figures in the parentheses are standard errors.