CORE DISCUSSION PAPER
2001/46

FOURTH MOMENTS OF MULTIVARIATE GARCH PROCESSES

Christian M. Hafner[1]

June 2001

Abstract

This paper derives conditions for the existence of fourth moments of multivariate GARCH processes in the general vector specification and gives explicit results for the fourth moments and autocovariances of the squares and cross-products. Results are provided for the kurtosis and co-kurtosis between components. Applications of the results include the definition of impulse response functions for kurtosis and co-kurtosis, the derivation of the spectral density matrix of the squares and cross-products, and a measure for causality in volatility. A bivariate exchange rate example illustrates the applications.

Keywords: multivariate GARCH, fourth moments, kurtosis, co-kurtosis
JEL Classification: C22

[1]Electrabel, R&D Energy Markets, Traverse d'Esope 6, B–1348 Louvain-la-Neuve, Belgium. e-mail `christian.hafner@electrabel.com`.
The author would like to thank Luc Bauwens, Robert Engle, and Oliver Linton for helpful discussions. The paper will be presented at the 56th European Meeting of the Econometric Society in Lausanne, August 2001.
This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the author.

1 Introduction

Multivariate GARCH models have been frequently used in modelling volatility of multivariate time series. To give one example of the numerous applications in finance, Bollerslev, Engle and Wooldridge (1988) used a multivariate GARCH-M model to estimate time-varying betas within the capital asset pricing model. GARCH-type models capture conditional heteroskedasticity and leptokurtosis, two common features of financial data. Moreover, the statistical theory is developping. For example, Jeantheau (1998) proves strong consistency of quasi maximum likelihood (QML) estimators of multivariate GARCH models, and Comte and Lieberman (2000a) prove asymptotic normality.

An often used measure for the thickness of the tails is the kurtosis, or the standardized fourth moment. When fitting a model, one may therefore want to reproduce not only the first two moments, but also the fourth moment. For example, having estimated the model based on QML, the question arises which innovation distribution one should use in simulation studies such that the empirical kurtosis is approximated. In finance, this is particularly important for option pricing, where the degree of excess kurtosis determines the shape of the so-called smile. A second example is the estimation by generalized method of moments (GMM), where the efficiency may often be improved using correctly specified skewness and kurtosis, see e.g. Meddahi and Renault (1997) and Broze, Francq and Zakoian (2001). A final example is to use the kurtosis in hedging strategies, as shown by Mahieu (1995, ch.6) for foreign exchange risks.

Fourth moments of univariate GARCH models were investigated by He and Teräsvirta (1999) and Karanasos (1999). The present paper provides results for the general multivariate GARCH(p, q) process in vector specification. This specification nests many popular linear specifications such as the BEKK specification of Engle and Kroner (1995) and the factor GARCH models of Diebold and Nerlove (1989) and Engle, Ng and Rothschild (1990). In Section 2, we provide explicit results for the case of multinormal or multivariate t-distributed innovations. In Section 3, we discuss the relationship between the (co-)kurtosis and measures for conditional heteroskedasticity. Furthermore, impulse response functions are defined for the kurtosis and the co-kurtosis. Section 4 derives results for the spectral density matrix of the squares and cross-products of a multivariate GARCH process. The analogy to VARMA processes is quite obvious here, the only difficulty being the derivation of the error covariance matrix. A potentially useful result is that the frequently used multivariate GARCH process with diagnoality restriction implies coherences that do not depend on the frequency. Section 5 suggests a measure for causality in volatility, based on well-known causality measures in VARMA models. In Section 6, an empirical example of a bivariate exchange rate series illustrates the ideas. The applications in this paper are not meant to be exhaustive from an economic or financial point of view, but rather to illustrate the potential use of fourth moment characteristics of multiple economic time series.

Proofs of the theorems are provided in the appendix.

2 Main results

Consider the general multivariate GARCH(p, q) model for the vector of a random error term ε_t with K components,

$$\begin{aligned} \varepsilon_t &= H_t^{1/2} \xi_t && (1) \\ h_t &= \omega + \sum_{i=1}^{q} A_i \eta_{t-i} + \sum_{j=1}^{p} B_j h_{t-j} && (2) \end{aligned}$$

where $h_t = \text{vech}(H_t)$, $\omega = \text{vech}(\Omega)$, $\eta_t = \text{vech}(\varepsilon_t \varepsilon_t')$ and $N \times N, N = K(K+1)/2$ parameter matrices Ω, A_i, B_j. Throughout the paper, vec denotes the operator that stacks all columns of a matrix into a vector, and vech denotes the operator that stacks only the lower triangular part including the diagonal of a symmetric matrix into a vector. The innovation vector ξ_t is assumed to be i.i.d. with mean zero and identity covariance matrix. We define the square root of the matrix H_t as $H_t^{1/2} = \Xi_t \Lambda_t^{1/2} \Xi_t'$ with the matrix Ξ_t containing the eigenvectors and the diagonal matrix Λ_t containing the eigenvalues of H_t on its diagonal.

Denoting the information set available at time t by $\mathcal{F}_t$, we have the conditional moments $\mathrm{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ and $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = H_t$. A sufficient condition for the conditional covariance matrix H_t to be positive definite is that each of the parameter matrices is symmetric positive definite. The symmetry of the parameter matrices, however, is a rather restrictive assumption. There are yet no general results on necessary and sufficient conditions for positivity of H_t. This is why in practice one often considers restrictions, such as the BEKK specification, that ensure positivity.

Assumption 1 *The matrix H_t is positive definite almost surely.*

Note that (1) is a strong GARCH model by the definition of Drost and Nijman (1993).

By rearranging terms, the multivariate GARCH(p, q) model can be represented as a VARMA $(\max(p, q), p)$ model,

$$\eta_t = \omega + \sum_{i=1}^{\max(p,q)} (A_i + B_i) \eta_{t-i} - \sum_{j=1}^{p} B_j u_{t-j} + u_t, \quad (3)$$

where $u_t = \eta_t - h_t$ is a white noise vector, i.e., $\mathrm{E}[u_t] = 0$, $\mathrm{E}[u_t u_t'] = \Sigma_u$ and $\mathrm{E}[u_t u_s'] = 0$ for $s \neq t$. In (3), we set $A_{q+1} = \ldots = A_p = 0$ if $p > q$ and $B_{p+1} = \ldots = B_q = 0$ if $q > p$.

Assumption 2 *All eigenvalues of the matrix $\sum_{i=1}^{\max(p,q)}(A_i + B_i)$ have modulus smaller than one.*

The multivariate GARCH(p, q) process ε_t is covariance stationary if and only if Assumptions 1 and 2 hold, see e.g. Engle and Kroner (1995). In that case, the unconditional covariance matrix $\Sigma = \mathrm{Var}(\varepsilon_t)$ is given by

$$\sigma = \mathrm{vech}(\Sigma) = \left(I_N - \sum_{i=1}^{\max(p,q)} (A_i + B_i)\right)^{-1} \omega.$$

From the VARMA representation (3) it is possible to obtain the VMA(∞) representation

$$\eta_t = \sigma + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \tag{4}$$

where the $N \times N$ matrices Φ_i can be determined recursively by $\Phi_0 = I_N$, $\Phi_i = -B_i + \sum_{j=1}^{i}(A_j + B_j)\Phi_{i-j}, i = 1, 2, \ldots$, see Lütkepohl (1993, pp. 220). Now, the autocovariance of η_t is given by

$$\begin{aligned} \Gamma(\tau) &= \mathrm{E}\left[(\eta_t - \sigma)(\eta_{t-\tau} - \sigma)'\right] \\ &= \sum_{i=0}^{\infty} \Phi_{\tau+i} \Sigma_u \Phi_i'. \end{aligned} \tag{5}$$

Using the notation $\Sigma_\eta = \mathrm{E}[\eta_t \eta_t']$ and $\Sigma_h = \mathrm{E}[h_t h_t']$, we can write $\Sigma_u = \Sigma_\eta - \Sigma_h$, and $\Gamma(0) = \Sigma_\eta - \sigma\sigma'$.

Autocorrelations are obtained by

$$R(\tau) = W^{-1}\Gamma(\tau)W^{-1}, \tag{6}$$

where W is a diagonal matrix with the square roots of the diagonal elements of $\Gamma(0)$ on its diagonal.

In order to give explicit results, one has to assume a specific distribution for the innovations ξ_t. In the following, we will assume the most popular ones, a multinormal or, alternatively, a multivariate t-distribution. Extensions to other distributions are left to future research.

Assumption 3 *The innovations ξ_t are either*

1. *multinormally distributed,* $\xi_t \sim N(0, I_K)$*, or*

2. *multivariate t-distributed* (t_ν) *with density*

$$f(\xi_t) = \frac{\gamma\left(\frac{\nu+K}{2}\right)}{\{\pi(\nu-2)\}^{K/2}\gamma(\nu/2)}\left(1+\frac{\xi_t'\xi_t}{\nu-2}\right)^{-(\nu+K)/2} \tag{7}$$

   and $\nu > 4$*,*

where $\gamma(p) = \int_0^\infty x^{p-1}e^{-x}dx$ *is the gamma function.*

Note that the multivariate t-distribution is the distribution of the scaled ξ_t, so that $\mathrm{Var}(\xi_t) = I_K$. The general multivariate t-distribution is given by

$$f(\xi_t) = \frac{\gamma\left(\frac{\nu+K}{2}\right)}{(\pi\nu)^{K/2}\gamma(\nu/2)|S|^{1/2}}\left(1+\frac{\xi_t'S^{-1}\xi_t}{\nu}\right)^{-(\nu+K)/2}$$

with $|S| > 0$ and $\mathrm{Var}(\xi_t) = \frac{\nu}{\nu-2}S$, $\nu > 2$, see e.g. Johnson and Kotz (1972). Setting $S = \frac{\nu-2}{\nu}I_K$, we obtain the special case (7) with $\mathrm{Var}(\xi_t) = I_K$.

Theorem 1 *Under Assumptions 1 to 3,*

$$vec(\Sigma_\eta) = G_K vec(\Sigma_h), \tag{8}$$

where

$$G_K = (L_K \otimes L_K)(I_{K^2} \otimes \tilde{N}_K)(I_K \otimes C_{KK} \otimes I_K)(D_K \otimes D_K) + I_{N^2}, \tag{9}$$

with L_m, D_m, C_{mn} *denoting the elimination, duplication and commutation matrices, respectively. The matrix* $\tilde{N}_n$ *is given by* $\tilde{N}_n = N_n = I_{n^2} + C_{nn}$ *in the Gaussian case and by*

$$\tilde{N}_n = I_{n^2} + C_{nn} + \frac{6}{\nu-4}\sum_{i=1}^{K} e_{l(i)}e_{l(i)}',$$

in the multivariate t-distribution case, where $l(i) = 1 + (i-1)(K+1)$ *and* e_j *is a vector with 1 at the* j*-th position and zeros elsewhere.* □

See Appendix A for a definition of L_m, D_m, C_{mn}, and N_n. The matrix G_K is square of order N^2. For example, in the bivariate case ($K = 2, N = 3$) it is given by

$$G_2 = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix} \quad \text{and} \quad G_2 = \begin{pmatrix} c_\nu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \tilde{c}_\nu & 0 & 0 & 0 & 0 \\ 0 & \tilde{c}_\nu & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & c_\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \tilde{c}_\nu & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_\nu \end{pmatrix}$$

for the Gaussian and t_ν case, respectively, where $c_\nu = 3+6/(\nu-4)$ and $\tilde{c}_\nu = 2+6/(\nu-4)$.

As a corollary, we obtain $\text{vec}(\Sigma_u) = (G_K - I_{N^2})\text{vec}(\Sigma_h)$. This can be used to calculate the autocovariances according to (5).

Assumption 4 *All eigenvalues of the matrix* $\sum_{i=1}^{\infty}(\Phi_i \otimes \Phi_i)(G_K - I_{N^2})$ *have modulus smaller than one.*

Theorem 2 *Under Assumptions 1 to 3, fourth moments of the multivariate GARCH(p, q) model are finite if and only if Assumption 4 holds. In that case, the fourth moment is given by*

$$vec(\Sigma_\eta) = G_K \left(I_{N^2} - \sum_{i=1}^{\infty}(\Phi_i \otimes \Phi_i)(G_K - I_{N^2}) \right)^{-1} vec(\sigma\sigma'). \tag{10}$$

□

Note that in the special case where all A_i and B_j are zero, i.e. the case of a Gaussian white noise vector process, all Φ_i's are zero and, hence, $\text{vec}(\Sigma_\eta) = G_K\text{vec}(\sigma\sigma')$.

If models of small order are considered, there are ways to find simpler expressions than (10). For the popular multivariate GARCH(1,1) model, we can derive the following results.

Theorem 3 *Under Assumption 3, fourth moments of the multivariate GARCH(1, 1) model are finite if and only if all eigenvalues of the matrix*

$$Z = (A \otimes A)G_K + A \otimes B + B \otimes A + B \otimes B \tag{11}$$

have modulus smaller than one. In that case, the fourth moment is given by

$$vec(\Sigma_\eta) = G_K(I_{N^2} - Z)^{-1} vec\left(\omega\omega' + \omega\sigma'(A+B)' + (A+B)\sigma\omega'\right) \tag{12}$$

and the τ-th order autocovariance matrix, $\tau \geq 1$, by

$$\Gamma(\tau) = (A+B)^{\tau-1}(A\Sigma_\eta + B\Sigma_h) - (A+B)^\tau \sigma\sigma'. \tag{13}$$

□

Covariance stationarity and existence of fourth moments implies that $\lim_{\tau\to\infty}\Gamma(\tau) = 0$. Note that, for $\tau > 1$, $\Gamma(\tau) = (A+B)\Gamma(\tau-1)$, so that a simple recursive calculation scheme can be used.

In the univariate Gaussian case, we have $G_1 = 3$ in the Gaussian case and $G_1 = c_\nu$ in the t_ν case and obtain as a special case of equation (12) with obvious notation

$$\mathrm{E}[\varepsilon_t^4] = 3\omega^2 \frac{1 + 2\frac{\alpha+\beta}{1-\alpha-\beta}}{1 - 3\alpha^2 - 2\alpha\beta - \beta^2},$$

for the Gaussian case and

$$\mathrm{E}[\varepsilon_t^4] = c_\nu\omega^2 \frac{1 + 2\frac{\alpha+\beta}{1-\alpha-\beta}}{1 - c_\nu\alpha^2 - 2\alpha\beta - \beta^2},$$

for the t_ν case. Finally, we obtain the well-known formulae for the kurtosis of a univariate GARCH(1,1) process,

$$\frac{\mathrm{E}[\varepsilon_t^4]}{\mathrm{E}[\varepsilon_t^2]^2} = 3 + \frac{6\alpha^2}{1 - 3\alpha^2 - 2\alpha\beta - \beta^2},$$

and

$$\frac{\mathrm{E}[\varepsilon_t^4]}{\mathrm{E}[\varepsilon_t^2]^2} = c_\nu + \frac{c_\nu(c_\nu - 1)\alpha^2}{1 - c_\nu\alpha^2 - 2\alpha\beta - \beta^2},$$

respectively.

3 Impulse response functions for kurtosis and co-kurtosis

In practice, one often considers fourth moments that are normalized with the square of the variance. In the multivariate case, Mardia (1970) has defined a measure of multivariate kurtosis for i.i.d. random variables. Suppose that X is a p-dimensional random vector with mean μ and covariance matrix Σ. Mardia's measure is given by

$$\beta_{2,p} = \mathrm{E}[\{(X-\mu)'\Sigma^{-1}(X-\mu)\}^2],$$

which is the mean of the squared Mahalanobis distance of X from its mean μ. The measure $\beta_{2,p}$ is invariant with respect to affine transformations and reduces to the standard kurtosis in the univariate case.

In our setting with $\mu = 0$, conditional and unconditional covariance matrix H_t and Σ, respectively, an analogous measure would be

$$\beta_{2,p} = \mathrm{E}[\{\varepsilon_t' \Sigma^{-1} \varepsilon_t\}^2]. \tag{14}$$

The reason for using Σ rather than H_t is that in the latter case the multivariate kurtosis would be trivial, i.e. only depending on the distribution of ξ_t, but not on the dynamics of H_t. For the notion (14), however, it is difficult to derive explicit results in the multivariate GARCH case.

In the following, we will therefore consider alternative measures based on the fourth moments of single components of the process. These measures are not invariant with respect to general affine transformations, but only with respect to diagonal transformations.

Definition 1 *Let the kurtosis of the $i-th$ component of ε_t be defined as*

$$k_{ii} = \frac{E[\varepsilon_{it}^4]}{E[\varepsilon_{it}^2]^2}.$$

and define the co-kurtosis of the i-th and j-th component of ε_t as

$$k_{ij} = \frac{E[\varepsilon_{it}^2 \varepsilon_{jt}^2]}{E[\varepsilon_{it}^2] E[\varepsilon_{jt}^2]}.$$

For a Gaussian i.i.d. process, it can be shown that the co-kurtosis takes the value $k_{ij} = 1 + 2\rho_{ij}^2$, where ρ_{ij} is the correlation between the i-th and j-th component. In the case of conditional heteroskedasticity, the following theorem gives useful expressions for the kurtosis and the co-kurtosis.

Theorem 4 *Under assumptions 1 to 4, the kurtosis and co-kurtosis in the Gaussian case can be expressed as*

$$k_{ii} = 3 + \frac{3\,Var(h_{ii,t})}{\sigma_{ii}^2},$$

$$k_{ij} = 1 + 2\rho_{ij}^2 + \frac{2\,Var(h_{ij,t}) + Cov(h_{ii,t}, h_{jj,t})}{\sigma_{ii}\sigma_{jj}},$$

and in the t_ν case as

$$k_{ii} = c_\nu + \frac{c_\nu\,Var(h_{ii,t})}{\sigma_{ii}^2},$$

$$k_{ij} = 1 + (c_\nu - 1)\rho_{ij}^2 + \frac{(c_\nu - 1)\,Var(h_{ij,t}) + Cov(h_{ii,t}, h_{jj,t})}{\sigma_{ii}\sigma_{jj}},$$

where $h_{ij,t}$ is the ij-th element of H_t and σ_{ii} the ii-th element of Σ. $\square$

The formula for k_{ii} (Gaussian case) can already be found in Gouriéroux (1997, pp.38). The kurtosis is linked to a natural measure of conditional heteroskedasticity, $\mathrm{Var}(h_{ii,t})/\sigma_{ii}^2$. Whenever the conditional variance is stochastic, the kurtosis is larger than three.

The co-kurtosis is linked to the covariance between the conditional variances and to the variance of the conditional covariance. The stronger the conditional variances are correlated and the higher the variance of the conditional covariance, the higher also the co-kurtosis.

In the following, we investigate the behavior of the expected kurtosis conditional on an initial shock. An impulse response function for volatility is given by

$$V_s(\xi_t) = \mathrm{E}[h_{t+s} \mid \xi_t, \mathcal{F}_{t-1}] - \mathrm{E}[h_{t+s} \mid \mathcal{F}_{t-1}]$$

as defined by Hafner and Herwartz (2001), who also provide an extensive comparison with the conditional moment profiles suggested by Gallant, Rossi, and Tauchen (1993), and the generalized impulse response functions of Koop, Pesaran, and Potter (1996).

We can adopt an analogous definition for the kurtosis and co-kurtosis.

Definition 2 *We define the impulse response function of the kurtosis as*

$$\kappa_{s,ii}(\xi_t) = \frac{E[\varepsilon_{i,t+s}^4 \mid \xi_t, \mathcal{F}_{t-1}]}{E[\varepsilon_{i,t+s}^2 \mid \xi_t, \mathcal{F}_{t-1}]^2} - \frac{E[\varepsilon_{i,t+s}^4 \mid \mathcal{F}_{t-1}]}{E[\varepsilon_{i,t+s}^2 \mid \mathcal{F}_{t-1}]^2} \tag{15}$$

and of the co-kurtosis as

$$\kappa_{s,ij}(\xi_t) = \frac{E[\varepsilon_{i,t+s}^2 \varepsilon_{j,t+s}^2 \mid \xi_t, \mathcal{F}_{t-1}]}{E[\varepsilon_{i,t+s}^2 \mid \xi_t, \mathcal{F}_{t-1}]E[\varepsilon_{j,t+s}^2 \mid \xi_t, \mathcal{F}_{t-1}]} - \frac{E[\varepsilon_{i,t+s}^2 \varepsilon_{j,t+s}^2 \mid \mathcal{F}_{t-1}]}{E[\varepsilon_{i,t+s}^2 \mid \mathcal{F}_{t-1}]E[\varepsilon_{j,t+s}^2 \mid \mathcal{F}_{t-1}]}. \tag{16}$$

The second terms on the right hand sides of (15) and (16) are the baseline expectations that only condition on history. The first terms are the expectations of kurtosis and co-kurtosis, respectively, conditional on history and a given shock ξ_t. Note that the conditional one-step-ahead kurtosis is constant, so that $\kappa_{1,ii}(\xi_t)$ is flat. This is a consequence of the strong GARCH assumption (independent innovations).

In analogy to the definition of persistence in variance by Bollerslev and Engle (1993), we suggest the following definition of persistence in kurtosis.

Definition 3 *The stochastic process ε_t is defined to be persistent in kurtosis if*

$$\limsup_{s \longrightarrow \infty} \mid \kappa_{s,ij}(\xi_t) \mid > 0 \quad a.s.$$

for some $i, j = 1, \ldots, K$.

The multivariate GARCH(p, q) model is persistent in kurtosis if at least one of the eigenvalues of the matrix $\sum_{i=1}^{\infty}(\Phi_i \otimes \Phi_i)(G_K - I_{N^2})$ has a modulus equal or greater than one. In the case of the multivariate GARCH$(1, 1)$ model, the relevant matrix simplifies to Z, given in (11). If ε_t is not persistent in kurtosis, the closeness of the eigenvalue with maximum norm to unity may be considered as a measure for the 'degree of persistence'.

The denominator of the first term in (15) is the square of the conditional expectation of volatility, given shock and history. For example, for the multivariate GARCH(1,1) model it is given by the corresponding element of

$$\mathrm{E}[\eta_{t+s} \mid \xi_t, \mathcal{F}_{t-1}] = \sigma_s + (A+B)^{s-1}\left\{A\mathrm{vech}(\Sigma_t^{1/2}\xi_t\xi_t'\Sigma_t^{1/2}) + B\mathrm{vech}(\Sigma_t)\right\},$$

where $\sigma_s = (I_N - A - B)^{-1}(I_N - (A+B)^s)\omega$ and $\sigma_s \to \sigma$ as $s \to \infty$. The baseline expectation of volatility is given by

$$\mathrm{E}[\eta_{t+s} \mid \mathcal{F}_{t-1}] = \sigma_s + (A+B)^s\mathrm{vech}(\Sigma_t).$$

Note that by the law of iterated expectations, we have $\mathrm{E}[\eta_{t+s} \mid \xi_t, \mathcal{F}_{t-1}] = \mathrm{E}[h_{t+s} \mid \xi_t, \mathcal{F}_{t-1}]$ for $s \geq 1$ and $\mathrm{E}[\eta_{t+s} \mid \mathcal{F}_{t-1}] = \mathrm{E}[h_{t+s} \mid \mathcal{F}_{t-1}]$ for $s \geq 0$.

For the fourth moment, we denote

$$F_s(\xi_t) = \mathrm{E}[\mathrm{vec}(\eta_{t+s}\eta_{t+s}') \mid \xi_t, \mathcal{F}_{t-1}]$$

and obtain $F_s(\xi_t) = G_K F_s^*(\xi_t)$ with

$$F_1^*(\xi_t) = \mathrm{vec}(h_{t+1}h_{t+1}')$$

and, for $s > 1$,

$$\mathrm{vec}[F_s^*(\xi_t)] = \gamma_s + Z\mathrm{vec}[F_{s-1}^*(\xi_t)],$$

where Z is given in (11), and $\gamma_s = \mathrm{vec}[\omega\omega' + (A+B)\mathrm{E}[h_{t+s} \mid \xi_t, \mathcal{F}_{t-1}]\omega' + \omega\mathrm{E}[h_{t+s}' \mid \xi_t, \mathcal{F}_{t-1}](A+B)']$. The baseline expectation of the fourth moment is denoted by

$$F_s^b(\xi_t) = \mathrm{E}[\mathrm{vec}(\eta_{t+s}\eta_{t+s}') \mid \mathcal{F}_{t-1}].$$

We have $F_s^b(\xi_t) = G_K F_s^{b*}(\xi_t)$ with $\mathrm{vec}[F_1^{b*}(\xi_t)] = \gamma_1^b + Z\mathrm{vec}(h_t h_t')$, and, for $s > 1$,

$$\mathrm{vec}[F_s^{b*}(\xi_t)] = \gamma_s^b + Z\mathrm{vec}[F_{s-1}^{b*}(\xi_t)],$$

with $\gamma_s^b = \mathrm{vec}[\omega\omega' + (A+B)\mathrm{E}[h_{t+s} \mid \mathcal{F}_{t-1}]\omega' + \omega\mathrm{E}[h_{t+s}' \mid \mathcal{F}_{t-1}](A+B)']$.

4 The spectral density matrix of the squares and cross-products

Sometimes it might be useful to look at the frequency domain properties of the squares and cross-products of a multivariate GARCH process, rather than at the time domain. Even though autocovariances and spectral densities basically convey the same information, they both have their advantages and certain aspects may be more apparent in one or the other. To give an example of application, the model specification process might compare the empirical spectral density with the theoretical counterpart of a fitted model. Or, specification and inference may be performed directly in the frequency domain.

The spectral density matrix of a covariance stationary stochastic process is defined as the Fourier transform of the autocovariance matrix function, i.e.

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} e^{-i\tau\lambda} \Gamma(\tau).$$

The form of the spectral density matrix of a VARMA process is well-known, see, e.g., Brockwell and Davis (1991, pp.459). Therefore, using the VARMA representation of the multivariate GARCH model given in (3) and the result for the autocovariance matrix of the errors Σ_u, we obtain the spectral density matrix of the vector of squares and cross-products, η_t.

Theorem 5 *Under Assumptions 1 to 4, the vector process $\eta_t = vech(\varepsilon_t \varepsilon_t')$ of a multivariate GARCH(p, q) process has a spectral density matrix given by*

$$f(\lambda) = \frac{1}{2\pi} \Psi^{-1}(e^{-i\lambda}) \Theta(e^{-i\lambda}) \Sigma_u \Theta'(e^{i\lambda}) \Psi^{-1\prime}(e^{i\lambda}),$$

where $\Psi(z) = I_N - \sum_{i=1}^{\max(p,q)} (A_i + B_i) z^i$ is the autoregressive lag polynomial, $\Theta(z) = I_N - \sum_{j=1} B_j z^j$ the moving average lag polynomial, and the error covariance matrix Σ_u is given by

$$vec(\Sigma_u) = (G_K - I_{N^2}) \left(I_{N^2} - \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i)(G_K - I_{N^2}) \right)^{-1} vec(\sigma\sigma') \qquad (17)$$

with the parameter matrices Φ_i of the VMA(∞) representation (4) and G_K given by (9). □

Clearly, Assumption 3 could be relaxed, but no explicit results for Σ_u would yet be available. Note that in the multivariate GARCH(1,1) case, using Theorem 3 together with $\Sigma_u = \Sigma_\eta - \Sigma_h$, the expression for Σ_u simplifies to

$$\text{vec}(\Sigma_u) = (G_K - I_{N^2})(I_{N^2} - Z)^{-1} \text{vec}\left(\omega\omega' + \omega\sigma'(A+B)' + (A+B)\sigma\omega'\right),$$

where Z is given by (11).

Since the cross-covariance function $\Gamma_{ij}(\tau)$, $i \neq j$, is not symmetric, the cross-spectrum f_{ij} is complex-valued. Typically, the cross-spectrum is interpreted in terms of the coherence $COH_{ij}(\lambda)$, the gain $G_{ij}(\lambda)$, and the phase $PH_{ij}(\lambda)$. They are defined as

$$COH_{ij}(\lambda) = \frac{|f_{ij}(\lambda)|^2}{f_{ii}(\lambda) f_{jj}(\lambda)} \tag{18}$$

$$G_{ij}(\lambda) = \frac{|f_{ij}(\lambda)|}{f_{ii}(\lambda)} \tag{19}$$

$$PH_{ij}(\lambda) = \arctan \frac{-\mathrm{Im}(f_{ij}(\lambda))}{\mathrm{Re}(f_{ij}(\lambda))}. \tag{20}$$

The coherence can be viewed as the squared correlation between the two series at a particular frequency. The phase is giving the time shift between the series within a frequency domain regression. Finally, the gain is like the slope coefficient of a regression of one series against another at a particular frequency after having corrected for the time shift. Note that the coherence is symmetric, i.e., $COH_{ij}(\lambda) = COH_{ji}(\lambda)$ for all i, j, whereas $PH_{ij}(\lambda) = -PH_{ji}(\lambda)$.

Theorem 6 *Under Assumptions 1 to 4, if the multivariate GARCH process has diagonal parameter matrices A_i and B_j, then the coherence between the components i and j of η_t, $COH_{ij}(\lambda)$, does not depend on λ and is given by*

$$COH_{ij}(\lambda) = \frac{\Sigma_{u,ij}^2}{\Sigma_{u,ii}\Sigma_{u,jj}}, \tag{21}$$

where $\Sigma_{u,ij}$ is the ij-th element of Σ_u. $\square$

This theorem can be used as the basis for a test of the popular diagonal restriction of the parameter matrices. The null hypothesis of a constant coherence would be rejected if an estimate of the empirical coherence shows distinct structure.

5 Measures for causality in volatility

Further extending well-known methods used in the analysis of VARMA models, we can introduce measures for the causality in volatility in a multivariate GARCH framework. In other words, we may be interested in a measure for how much one series contributes to explain the volatility of another series. In this paper, we use the term causality in the sense of 'Granger causality', which for volatility has been defined by Granger, Robins, and Engle (1984) and Comte and Lieberman (2000b).

Let ε_t be decomposed into ε_{1t} and ε_{2t}, being vectors of length K_1 and K_2, respectively, with $K = K_1 + K_2$. Denote the σ-algebra generated by innovations in ε_{1t} and ε_{2t} only by $\mathcal{F}_{1,t}$ and $\mathcal{F}_{2,t}$, respectively. Then ε_{2t} is said not to Granger cause ε_{1t} in variance, if $\forall t \in \mathbb{Z}$

$$\mathrm{Var}(\varepsilon_{1t} \mid \mathcal{F}_{t-1}) = \mathrm{Var}(\varepsilon_{1t} \mid \mathcal{F}_{1,t-1}). \tag{22}$$

Comte and Lieberman (2000b) point out that (22) is not the same definition as the one used by Granger, Robins, and Engle (1984), who define causality in second moments after having corrected for the conditional mean. In our case the conditional mean of ε_t is zero, so that the two definitions are equivalent.

In the following we consider measures for causality in variance for a given multivariate GARCH process (3). If the process were not known to be multivariate GARCH, then there would not be an equivalence relation between the causality measure being zero and no causality. To establish equivalence, one would have to resort to weaker notions of causality, based for example on orthogonal projections, allowing interpretations of *linear causality*, see e.g. Boudjellaba, Dufour, and Roy (1992).

In order to introduce a measure for the causality in variance from ε_{2t} to ε_{1t}, or to see whether ε_{2t} helps in explaining the volatility of ε_{1t}, let $x_t = \mathrm{vech}(\varepsilon_{1t}\varepsilon_{1t}')$ and $y_t = \mathrm{vech}(\varepsilon_{2t}\varepsilon_{2t}')$ and consider the restricted multivariate GARCH model for ε_{1t} which can be represented in VARMA form as

$$x_t = \omega^x + \sum_{i=1}^{\max(p,q)} (A_i^x + B_i^x) x_{t-i} - \sum_{j=1}^{p} B_j^x v_{t-j} + v_t, \tag{23}$$

where $v_t = x_t - E[x_t \mid \mathcal{F}_{1,t-1}]$, ω^x is an N_1 vector and A_i^x and B_j^x are $N_1 \times N_1$ matrices with $N_1 = K_1(K_1+1)/2$. Denote the variance of the white noise vector v_t by $\mathrm{Var}(v_t) = \Sigma_v$. Then, a measure for causality in volatility from y_t to x_t is given by

$$\mathcal{CV}_{x \leftarrow y} = \log \frac{|\Sigma_v|}{|\Sigma_{u,x}|}, \tag{24}$$

where $\Sigma_{u,x}$ is the part of the matrix Σ_u given in (17) that corresponds to x_t. This causality measure is analogous to the one introduced by Geweke (1982) for VARMA models. For a given multivariate GARCH process, there is no causality in variance from y to x if, and only if, $\mathcal{CV}_{x \leftarrow y} = 0$.

For the reverse causality direction, build a VARMA model for y_t with error term w_t and corresponding covariance matrix Σ_w,

$$y_t = \omega^y + \sum_{i=1}^{\max(p,q)} (A_i^y + B_i^y) y_{t-i} - \sum_{j=1}^{p} B_j^y w_{t-j} + w_t, \tag{25}$$

with appropriate parameter dimensions and where $w_t = y_t - E[y_t \mid \mathcal{F}_{2,t-1}]$. Then, a causality measure for x_t to y_t is given by

$$\mathcal{CV}_{x\to y} = \log \frac{|\Sigma_w|}{|\Sigma_{u,y}|},$$

where $\Sigma_{u,y}$ is the part of the matrix Σ_u that corresponds to y_t. For a given multivariate GARCH process, there is no causality in variance from x to y if, and only if, $\mathcal{CV}_{x\to y} = 0$.

A measure for instantaneous causality in volatility between x_t and y_t is given by

$$\mathcal{CV}_{x\leftrightarrow y} = \log \frac{|\Sigma_{u,x}||\Sigma_{u,y}|}{|\Sigma_{u,xy}|},$$

where $\Sigma_{u,xy}$ denotes the part of Σ_u that corresponds to x_t and y_t. For a given multivariate GARCH process, there is no instantaneous causality in volatility between x and y if, and only if, $\mathcal{CV}_{x\leftrightarrow y} = 0$.

Finally, a measure for linear dependence between x_t and y_t is defined by

$$\mathcal{CV}_{x,y} = \log \frac{|\Sigma_v||\Sigma_w|}{|\Sigma_{u,xy}|}.$$

It follows immediately that this measure can be decomposed into the three causality measures:

$$\mathcal{CV}_{x,y} = \mathcal{CV}_{x\to y} + \mathcal{CV}_{x\leftarrow y} + \mathcal{CV}_{x\leftrightarrow y}.$$

Furthermore, it is possible to decompose each unidirectional causality measure into the effects of different lags of the explanatory variable, see e.g. Gouriéroux and Monfort (1995, pp. 369).

For inference on causality, denote by $\widehat{\mathcal{CV}}$ an estimate of $\mathcal{CV}$ for a given sample of size n. If the errors v_t in (23) and w_t in (25) were Gaussian, then $n\widehat{\mathcal{CV}}_{x\to y}$ and $n\widehat{\mathcal{CV}}_{x\leftarrow y}$ would be the standard likelihood ratio statistics for testing the null hypothesis that $\mathcal{CV}_{x\to y} = 0$ and $\mathcal{CV}_{x\leftarrow y} = 0$, respectively. Similarly, Wald (used by Comte and Lieberman, 2000b) and Lagrange Multiplier statistics could be constructed. However, in all cases the presumed chi-square distribution under the null hypothesis of no causality has not been proven yet to match the asymptotic distribution of the statistics. In our case, proofs will be even more difficult since the errors v_t and w_t are not Gaussian and, hence, the scaled causality measures are not equal to the likelihood ratio statistics. At best, they could be called *quasi likelihood ratio statistics*, for which in the case of a univariate AR-GARCH model Hafner and Herwartz (2000) show that they have to be re-scaled to obtain a standard chi square distribution. An extension of their results to the multivariate case is however beyond the scope of this paper and left to future research. Therefore, inference and testing based on these measures should be treated with care. On the other hand, from a purely descriptive viewpoint the various possibilities to decompose the causality measures may give valuable information on the causality structure.

6 Empirical Example

For an empirical illustration, let us consider the bivariate exchange rate series DEM/USD and GBP/USD from december 31, 1979 to april 1, 1994, as discussed by Hafner and Herwartz (2001). A preliminary study eliminated first order autocorrelation by fitting univariate AR(1) models to the series. Then, a BEKK-GARCH(1,1) model was fitted to the bivariate residual series with two specifications of the conditional distribution: multinormal and multivariate t-distribution. For the latter case, the degrees of freedom parameter was estimated to be 4.7, implying finite fourth moments of the innovations ξ_t.

In the Gaussian case, the eigenvalues of the matrix Z all lie between 0.914 and 0.976, implying finite fourth moments of ε_t but a high persistence in kurtosis. The largest eigenvalue of Z in the t_ν case is 1.028, so that in this case ε_t does not have finite fourth moments.

The kurtosis as implied by the parameter estimates and assuming that the innovations are Gaussian are 3.50 for DEM/USD, 3.80 for GBP/USD and 2.52 for the co-kurtosis. This is considerably smaller than the empirical kurtosis of 4.94 for the residuals of DEM/USD, 5.47 for the residuals of GBP/USD, and 3.78 for the co-kurtosis of the residuals, which is a well-known empirical effect: Conditional heteroskedasticity is often not sufficient to explain the excess kurtosis of financial returns.

Table 1 reports the auto- and cross-correlations of η_t under $\xi_t \sim N(0, I_K)$, calculated using equations (6) in combination with (13). Clearly, the diagonal elements of $R(\tau)$ are higher than the off-diagonal elements. For example, the first order autocorrelation of ε_{1t}^2 (DEM/USD) is 0.122, of ε_{2t}^2 (GBP/USD) it is 0.144, and of the cross-product it is 0.123. For all autocorrelations, the decay of the autocorrelation function is slow, corresponding to the large eigenvalues of the matrix Z. Table 2 reports the corresponding empirical autocorrelations of the exchange rate series. Comparing the two tables, it appears that low (high) order correlations are somewhat higher (lower) for the model compared with the data. In modelling univariate financial time series, this is yet another well-known empirical fact: The empirical volatility often contains long memory that is approximated by a GARCH(1,1) model which is short memory. Furthermore, it is remarkable that all correlations in both tables are positive.

To illustrate the use of impulse response functions for the kurtosis and co-kurtosis, let

τ	1	2	3	4	5	10	50	100
$\mathrm{Corr}[\varepsilon_{1t}^2, \varepsilon_{1,t-\tau}^2]$	0.122	0.119	0.115	0.112	0.109	0.095	0.034	0.012
$\mathrm{Corr}[\varepsilon_{1t}\varepsilon_{2t}, \varepsilon_{1,t-\tau}^2]$	0.091	0.090	0.089	0.087	0.086	0.079	0.041	0.018
$\mathrm{Corr}[\varepsilon_{2t}^2, \varepsilon_{1,t-\tau}^2]$	0.066	0.066	0.065	0.065	0.065	0.062	0.041	0.021
$\mathrm{Corr}[\varepsilon_{1t}^2, \varepsilon_{1,t-\tau}\varepsilon_{2,t-\tau}]$	0.096	0.094	0.092	0.090	0.088	0.078	0.034	0.014
$\mathrm{Corr}[\varepsilon_{1t}\varepsilon_{2t}, \varepsilon_{1,t-\tau}\varepsilon_{2,t-\tau}]$	0.123	0.121	0.119	0.117	0.115	0.105	0.053	0.023
$\mathrm{Corr}[\varepsilon_{2t}^2, \varepsilon_{1,t-\tau}\varepsilon_{2,t-\tau}]$	0.105	0.104	0.103	0.102	0.101	0.095	0.058	0.028
$\mathrm{Corr}[\varepsilon_{1t}^2, \varepsilon_{2,t-\tau}^2]$	0.069	0.068	0.067	0.066	0.065	0.059	0.030	0.013
$\mathrm{Corr}[\varepsilon_{1t}\varepsilon_{2t}, \varepsilon_{2,t-\tau}^2]$	0.107	0.105	0.104	0.102	0.101	0.093	0.050	0.023
$\mathrm{Corr}[\varepsilon_{2t}^2, \varepsilon_{2,t-\tau}^2]$	0.144	0.142	0.140	0.137	0.135	0.125	0.067	0.031

Table 1: *Theoretical auto- and cross-correlations of η_t with lag τ, calculated using equation (6) in combination with (13) and parameter estimates for the bivariate exchange rate series.*

τ	1	2	3	4	5	10	50	100
$\mathrm{Corr}[\varepsilon_{1t}^2, \varepsilon_{1,t-\tau}^2]$	0.098	0.089	0.093	0.063	0.097	0.084	0.030	0.053
$\mathrm{Corr}[\varepsilon_{1t}\varepsilon_{2t}, \varepsilon_{1,t-\tau}^2]$	0.078	0.081	0.067	0.047	0.075	0.061	0.021	0.052
$\mathrm{Corr}[\varepsilon_{2t}^2, \varepsilon_{1,t-\tau}^2]$	0.078	0.097	0.071	0.070	0.074	0.055	0.006	0.043
$\mathrm{Corr}[\varepsilon_{1t}^2, \varepsilon_{1,t-\tau}\varepsilon_{2,t-\tau}]$	0.096	0.054	0.070	0.058	0.092	0.065	0.033	0.053
$\mathrm{Corr}[\varepsilon_{1t}\varepsilon_{2t}, \varepsilon_{1,t-\tau}\varepsilon_{2,t-\tau}]$	0.070	0.076	0.068	0.059	0.095	0.063	0.035	0.056
$\mathrm{Corr}[\varepsilon_{2t}^2, \varepsilon_{1,t-\tau}\varepsilon_{2,t-\tau}]$	0.076	0.081	0.083	0.080	0.091	0.045	0.021	0.045
$\mathrm{Corr}[\varepsilon_{1t}^2, \varepsilon_{2,t-\tau}^2]$	0.080	0.046	0.060	0.059	0.091	0.056	0.027	0.025
$\mathrm{Corr}[\varepsilon_{1t}\varepsilon_{2t}, \varepsilon_{2,t-\tau}^2]$	0.046	0.052	0.053	0.071	0.089	0.064	0.033	0.028
$\mathrm{Corr}[\varepsilon_{2t}^2, \varepsilon_{2,t-\tau}^2]$	0.090	0.082	0.082	0.111	0.088	0.067	0.030	0.027

Table 2: *Empirical auto- and cross-correlations of η_t with lag τ, where ε_{1t} and ε_{2t} are the residuals of the DEM/USD and GBP/USD series, respectively.*

us look at two observed historical shocks. The first is the so-called 'Black Wednesday', September 16, 1992, when the Italian lira and the pound sterling dropped out of the European Exchange Rate Mechanism (ERM). This event was followed by a very volatile period on the foreign exchange markets which is often referred to as currency crisis, see e.g. Rose and Svensson (1994) for a treatment of the reasons for the crisis. The second shock we consider is the announcement by the EC finance ministers to enlarge the ERM variability bands from 2 1/4 to 15 percent for six rates on August 2, 1993. As a consequence, the DEM/USD appreciated against the Dollar whereas the pound sterling remained at about the same level. So this event had an asymmetric effect on volatility. We will compare impulse response functions for the kurtosis with corresponding values for volatility as defined by Hafner and Herwartz (2001). We take the estimated residual ε_t and the estimated volatility state Σ_t on the day the shock occured and construct standardized residuals ξ_t, for which we can calculate the impulse response functions. The impulse responses are scaled with respect to the baseline expectation. This allows to interpret the scales as percentage deviations of the 'shock scenario' with respect to the 'baseline scenario'.

On September 16, 1992, the estimated residual vector was $\varepsilon_t = (0.0221, 0.0321)'$ and the estimated volatility state (expressed in E-05) $\mathrm{vech}(\Sigma_t) = (1.196, 0.962, 0.994)'$. The left panels of Figures 1 and 2 show the impulse responses. As the shock is strongly positive in both FX rates, the shock is positive not only for the variances but also for the covariance. The kurtosis impulse responses in the Gaussian case are quite flat with a small increase over the first 100 days and a decrease thereafter.

On August 2, 1993, the estimated residual vector was $\varepsilon_t = (-0.014807, -0.003195)'$ and the estimated volatility state (expressed in E-05) $\mathrm{vech}(\Sigma_t) = (0.3886, 0.3341, 0.4852)'$. Obviously, this was a much quieter period, so that the shock in DEM/USD was relatively strong, whereas GBP/USD did not change much. We see from the right hand panel in Figure 1 a familiar pattern for the DEM/USD variance. Conversely, the volatility impulse response of GBP/USD starts at a negative level. Again, the kurtosis impulse responses are very flat, but the co-kurtosis shows a big initial impact.

The picture changes substantially when looking at the corresponding functions for the case $\xi_t \sim t_\nu$ in Figure 2. All kurtosis and co-kurtosis impulse response functions clearly do not converge to zero, but rather to some positive value. By Definition 3, shocks are in this case persistent in kurtosis, which is due to the largest eigenvalue of Z exceeding unity.

Turning to the theoretical spectral densities as discussed in Section 4, we plot in Figure 3 the coherence between the squared residuals of DEM/USD and the squared residuals of GBP/USD. The function decreases sharply at very low frequencies, then increases and remains on the same level, about 0.34, at medium and high frequencies. Thus, a regression of the squares of one series on the squares of the other series gives an R^2 of about 34% at medium and high frequencies. Both gain functions have a similar shape. The phase between the two series is close to zero for medium to high frequencies.

Finally, let us investigate measures for the causality in volatility as introduced in Section 5. First, a univariate GARCH(1,1) model was estimated for DEM/USD, giving an empirical variance for the error term v_t. This is compared with the element in the first row, first column, of the estimated covariance matrix Σ_u, $\Sigma_{u,11}$ say, of the estimated bivariate GARCH(1,1) model to give $\mathcal{CV}_{DEM/USD \leftarrow GBP/USD}$. Second, we estimate a univariate GARCH(1,1) for GBP/USD, obtain Σ_w which is compared with $\Sigma_{u,33}$ to give $\mathcal{CV}_{DEM/USD \rightarrow GBP/USD}$. To obtain the instantaneous causality, we compare the product of $\Sigma_{u,11}$ and $\Sigma_{u,33}$ to the determinant of the matrix $\Sigma_{u,-2}$, which is obtained by eliminating the second row and column from Σ_u. The measure of linear dependence is then given by the sum of these causality measures. This gives us the values

$$\begin{aligned}
\mathcal{CV}_{DEM/USD \leftarrow GBP/USD} &= 0.0063 \\
\mathcal{CV}_{DEM/USD \rightarrow GBP/USD} &= 0.0324 \\
\mathcal{CV}_{DEM/USD \leftrightarrow GBP/USD} &= 0.5508 \\
\mathcal{CV}_{DEM/USD, GBP/USD} &= 0.5895.
\end{aligned}$$

Obviously, DEM/USD has a higher explanatory power for GBP/USD volatility than GBP/USD has for DEM/USD volatility. The instantaneous causality is much stronger than the unidirectional causalities. This is plausible since the foreign exchange market is the most efficient and transparent financial market, so that unpredicted shocks that increase volatility in one series will almost immediately spill over to the volatility of a related series. The impact after one day is still strong, but smaller than the instantaneous effect. Corresponding to the causality measures, we can calculate quasi likelihood ratio statistics, $n\widehat{\mathcal{CV}}$, which take the values 23.26, 120.5, 2047.9, and 2191.8, respectively. Under a chi square distribution with respectively four, four, two and ten degrees of freedom, all statistics would reject the corresponding null hypotheses $\mathcal{CV} = 0$ at a one percent level. As noted in Section 5, however, this result should be treated with care.

It should be emphasized that all above considerations take the estimated models as given and neglect estimation uncertainty. A theory for inference on impulse responses, spectral density and causality measures in volatility has yet to be developed and is beyond the scope of this paper.

7 Concluding remarks

This paper provides explicit results for the fourth moment structure of multivariate GARCH(p,q) processes in their general vector specification. Unlike second moments, fourth moments crucially depend on the innovation distribution, for which we assume alternatively a multinormal or multivariate t-distribution. The paper suggests some potential applications of the results, such as impulse response functions for the (co-)kurtosis, spectral density of the squares and cross-products, and measures for the causality in volatility. They are illustrated using a bivariate exchange rate example. Many further applications are possible. For example, the issue of temporal aggregation of multivariate GARCH models has not been addressed yet. Results as in Drost and Nijman (1993) and Drost and Werker (1996) for univariate GARCH models would certainly be useful to investigate the (co-)kurtosis as a function of the sampling frequency. The extension of their results to the multivariate case is straightforward using the results of this paper. The details however are tedious to work out so that they were not included in this paper. Last, but not least, the results of this paper will have to be extended to multivariate GARCH processes allowing for asymmetric impacts of positive and negative shocks to volatility, as in Hafner and Herwartz (1998) and Kroner and Ng (1998), who both use threshold variables.

Appendix A: The elimination, duplication and commutation matrices

The $(mn \times mn)$ commutation matrix C_{mn} is defined by

$$C_{mn}\text{vec}(A) = \text{vec}(A')$$

for every $(m \times n$ matrix A. A useful property of C_{mn} is the following: If A is $(n \times p)$ and B is $(q \times m)$, then

$$\text{vec}(A \otimes B) = (I_p \otimes C_{mn} \otimes I_q)(\text{vec}(A) \otimes \text{vec}(B)).$$

Let E_{ij} be the $(m \times n)$ matrix with 1 in its ij-th position and zeros elsewhere. Then an explicit expression for C_{mn} is given by

$$C_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n}(E_{ij} \otimes E'_{ij}).$$

For example, C_{22} is given by

$$C_{22} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The matrix N_n is square of order n^2 with the property that

$$N_n \text{vec}(A) = \text{vec}(A + A')$$

for every square matrix A of order n. It is related to the commutation matrix by

$$N_n = I_{n^2} + C_{nn}.$$

For example, N_2 is given by

$$N_2 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

The $(n^2 \times n(n+1)/2)$ duplication matrix D_n is defined so that

$$D_n \text{vech}(A) = \text{vec}(A)$$

for every symmetric matrix A of order n. For example, D_2 is given by

$$D_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

An explicit expression is given by

$$D_n = \sum_{i>j} \text{vec}(E_{ij} + E_{ji}) \text{vech}(E_{ij})' + \sum_{i=j} \text{vec}(E_{ij}) \text{vech}(E_{ij})'.$$

Finally, the elimination matrix L_n is an $(n(n+1)/2 \times n)$ matrix, which is defined so that

$$\text{vech}(A) = L_n \text{vec}(A)$$

for every symmetric matrix A of order n. For example, L_2 is given by

$$L_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

An explicit expresion is given by

$$L_n = \sum_{i \geq j} \text{vech}(E_{ij})(\text{vec}(E_{ij}))'.$$

For details, see e.g. Magnus (1988).

Appendix B: Proofs

Proof of Theorem 1: Under model (1) and Assumption 3.1, we have $\varepsilon_t \mid \mathcal{F}_{t-1} \sim N(0, H_t)$. Applying Theorem 10.2 of Magnus (1988), one obtains

$$\text{Var}(\text{vec}(\varepsilon_t \varepsilon_t') \mid \mathcal{F}_{t-1}) = \tilde{N}_K (H_t \otimes H_t) \tag{26}$$

with $\tilde{N}_n = N_n = I_{n^2} + C_{nn}$. Alternatively, under Assumption 3.2 ($\xi_t \sim t_\nu$), (26) holds but with

$$\tilde{N}_n = I_{n^2} + C_{nn} + \frac{6}{\nu - 4} \sum_{i=1}^{K} e_{l(i)} e_{l(i)}',$$

where $l(i) = 1 + (i-1)(K+1)$ and e_j is a vector with 1 at the j-th position and zeros elsewhere. By definition, $\eta_t = L_K \text{vec}(\varepsilon_t \varepsilon_t')$, so that

$$\text{Var}(\eta_t \mid \mathcal{F}_{t-1}) = L_K \tilde{N}_K (H_t \otimes H_t) L_K'.$$

Applying the vec-operator, we get

$$\text{vec}(\text{Var}[\eta_t \mid \mathcal{F}_{t-1}]) = (L_K \otimes L_K)(I_{K^2} \otimes \tilde{N}_K) \text{vec}(H_t \otimes H_t).$$

Now, $\text{vec}(H_t \otimes H_t)$ can be expressed as the product of the vecs using the commutation matrix C to obtain

$$\text{vec}(\text{Var}[\eta_t \mid \mathcal{F}_{t-1}]) = (L_K \otimes L_K)(I_{K^2} \otimes \tilde{N}_K)(I_K \otimes C_{KK} \otimes I_K)(\text{vec}(H_t) \otimes \text{vec}(H_t)).$$

By definition, $\text{vec}(H_t) = D_K h_t$, so that $\text{vec}(H_t) \otimes \text{vec}(H_t) = (D_K \otimes D_K)\text{vec}(h_t h_t')$. Recalling the identity $\text{Var}(\eta_t \mid \mathcal{F}_{t-1}) = \text{E}[\eta_t \eta_t' \mid \mathcal{F}_{t-1}] - h_t h_t'$ and, hence,

$$\Sigma_\eta = \text{E}[\text{Var}(\eta_t \mid \mathcal{F}_{t-1})] + \Sigma_h$$

with $\Sigma_h = \text{E}[h_t h_t']$, and taking vecs, the result is obtained. □

Proof of Theorem 2: Since $\Gamma(0) = \Sigma_\eta - \sigma\sigma'$ and $\Sigma_u = \Sigma_\eta - \Sigma_h$, we can write

$$\Sigma_\eta = \sum_{i=0}^{\infty} \Phi_i (\Sigma_\eta - \Sigma_h) \Phi_i' + \sigma\sigma'.$$

Taking vecs and applying Theorem 1, we obtain

$$G_K \mathrm{vec}(\Sigma_h) = \sum_{i=0}^{\infty} (\Phi_i \otimes \Phi_i)(G_K - I_{N^2})\mathrm{vec}(\Sigma_h) + \mathrm{vec}(\sigma\sigma'),$$

and, since $\Phi_0 = I_N$,

$$\left(I_{N^2} - \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i)(G_K - I_{N^2}) \right) \mathrm{vec}(\Sigma_h) = \mathrm{vec}(\sigma\sigma').$$

The matrix $I_{N^2} - \sum_{i=1}^{\infty}(\Phi_i \otimes \Phi_i)(G_K - I_{N^2})$ is invertible if and only if all its eigenvalues have modulus smaller than one. □

Proof of Theorem 3: Taking expectations, we have

$$\begin{aligned} \mathrm{E}[\mathrm{vec}(h_t h_t')] = \gamma \quad &+ \quad (A \otimes A) E[\mathrm{vec}(\eta_{t-1}\eta_{t-1}')] \\ &+ \quad \{(A \otimes B) + (B \otimes A) + (B \otimes B)\}\mathrm{E}[\mathrm{vec}(h_{t-1}h_{t-1}')], \end{aligned}$$

where $\gamma = \mathrm{vec}\{\omega\omega' + (A+B)\sigma\omega' + \omega\sigma'(A+B)'\}$. By Theorem 1, $\mathrm{vec}(\Sigma_\eta) = G_K \mathrm{vec}(\Sigma_h)$ so that

$$\mathrm{E}[\mathrm{vec}(h_t h_t')] = \gamma + \{(A \otimes A)G_K + (A \otimes B) + (B \otimes A) + (B \otimes B)\}\mathrm{E}[\mathrm{vec}(h_{t-1}h_{t-1}')]. \quad (27)$$

Subsequent substitution leads to

$$\mathrm{E}[\mathrm{vec}(h_t h_t')] = (I_{N^2} + Z + \ldots + Z^{\tau-1})\gamma + Z^{\tau}\mathrm{E}[\mathrm{vec}(h_{t-\tau}h_{t-\tau}')] \quad (28)$$

with $Z = (A \otimes A)G_K + (A \otimes B) + (B \otimes A) + (B \otimes B)$. A necessary and sufficient condition for $Z^\tau \to 0$ as $\tau \to \infty$ is that all eigenvalues of Z are less than one in modulus. This condition is equivalent to $(I_{N^2} + Z + \ldots + Z^{\tau-1}) \to (I_{N^2} - Z)^{-1}$ as $\tau \to \infty$. Therefore, Σ_h converges in probability as $\tau \to \infty$ to $(I_{N^2} - Z)^{-1}\gamma$ if and only if all eigenvalues of Z are less than one in modulus. This proves the first part of the theorem.

To prove the second part, we have, with $\tau \geq 2$,

$$\begin{aligned} \mathrm{E}[h_t \mid \mathcal{F}_{t-\tau}] &= \mathrm{E}[\omega + A\eta_{t-1} + Bh_{t-1} \mid \mathcal{F}_{t-\tau}] \\ &= \omega + (A+B)\mathrm{E}[h_{t-1} \mid \mathcal{F}_{t-\tau}] \\ &= (I_N + A + B)\omega + (A+B)^2\mathrm{E}[h_{t-2} \mid \mathcal{F}_{t-\tau}] \\ &= \cdots \\ &= \{I_N + A + B + \ldots + (A+B)^{\tau-2}\}\omega + (A+B)^{\tau-1}h_{t-\tau+1} \\ &= \{I_N + A + B + \ldots + (A+B)^{\tau-1}\}\omega + (A+B)^{\tau-1}(A\eta_{t-\tau} + Bh_{t-\tau}) \\ &= \{I_N - (A+B)^{\tau}\}\sigma + (A+B)^{\tau-1}(A\eta_{t-\tau} + Bh_{t-\tau}) \end{aligned}$$

and therefore

$$\begin{aligned}
\mathrm{E}[\eta_t\eta_{t-\tau}'] &= \mathrm{E}[\mathrm{E}(\eta_t\eta_{t-\tau}' \mid \mathcal{F}_{t-\tau})] \\
&= \mathrm{E}[\mathrm{E}(h_t \mid \mathcal{F}_{t-\tau})\eta_{t-\tau}'] \\
&= \{I_N - (A+B)^\tau\}\sigma\sigma' + (A+B)^{\tau-1}(A\Sigma_\eta + B\Sigma_h).
\end{aligned}$$

Subtracting $\sigma\sigma'$, the result for $\Gamma(\tau)$ is obtained. □

Proof of Theorem 4: Only the equation for k_{ij} will be proved. We have

$$\begin{aligned}
\mathrm{E}[\varepsilon_{it}^2\varepsilon_{jt}^2] &= \mathrm{E}[\mathrm{E}(\varepsilon_{it}^2\varepsilon_{jt}^2 \mid \mathcal{F}_{t-1})] \\
&= \mathrm{E}[h_{ii,t}h_{jj,t} + 2h_{ij,t}^2] \\
&= \sigma_{ii}\sigma_{jj} + \mathrm{Cov}(h_{ii,t}, h_{jj,t}) + 2\left(\mathrm{Var}(h_{ij,t}) + \sigma_{ij}^2\right). \qquad (29)
\end{aligned}$$

The second equality follows from a general property of product-moments of multinormal variates: If $X \sim N(0,\Sigma)$, then $\mathrm{E}[X_iX_jX_kX_l] = \Sigma_{ij}\Sigma_{kl} + \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}$. In (29), dividing by $\sigma_{ii}\sigma_{jj}$, the result is obtained. □

Proof of Theorem 5: Follows immediately from the VARMA representation of η_t in equation (3) and the results for the spectral density matrix of VARMA processes given by e.g. Brockwell and Davis (1991, pp.459). Theorem 2 gives the explicit formula for Σ_η, which also yields Σ_u using $\Sigma_u = \Sigma_\eta - \Sigma_h$. □

Proof of Theorem 6: To simplify the proof, let us consider a multivariate ARCH(1) process. Generalizations to multivariate GARCH(p,q) are obvious. Under the assumption of diagonality, the matrix $\Psi(e^{-i\lambda})$ is diagonal with elements $1-\psi_{jj}e^{-i\lambda} = 1-\psi_{jj}(\cos\lambda - i\sin\lambda)$ on its diagonal.

Since $f(\lambda) = \frac{1}{2\pi}\Psi^{-1}(e^{-i\lambda})\Sigma_u\Psi^{-1}(e^{i\lambda})'$, the jk-th element of the spectral density matrix can be written as

$$\begin{aligned}
f_{jk}(\lambda) &= \frac{1}{2\pi}\frac{\Sigma_{u,jk}}{(1-\psi_{jj}(\cos\lambda - i\sin\lambda))(1-\psi_{kk}(\cos\lambda + i\sin\lambda))} \qquad (30) \\
&= \frac{1}{2\pi}\frac{\Sigma_{u,jk}}{1+\psi_{jj}\psi_{kk} - (\psi_{jj}+\psi_{kk})\cos\lambda + i(\psi_{jj}-\psi_{kk})\sin\lambda} \qquad (31)
\end{aligned}$$

For the real valued marginal spectral densities ($j=k$), this expression simplifies to

$$f_{jj} = \frac{1}{2\pi}\frac{\Sigma_{u,jj}}{1-2\psi_{jj}\cos\lambda + \psi_{jj}^2}.$$

Now, straightforward calculations show that

$$|f_{ij}(\lambda)|^2 = \mathrm{Re}(f_{ij}(\lambda))^2 + \mathrm{Im}(f_{ij}(\lambda))^2$$

$$= \frac{1}{4\pi^2} \frac{\Sigma^2_{u,jk}}{1+\psi^2_{jj}\psi^2_{kk}+\psi^2_{jj}+\psi^2_{kk}+4\psi_{jj}\psi_{kk}\cos^2\lambda-2(1+\psi_{jj}\psi_{kk})(\psi_{jj}+\psi_{kk})\cos\lambda}$$

$$= \frac{1}{4\pi^2} \frac{\Sigma^2_{u,jk}}{(1-2\psi_{jj}\cos\lambda+\psi^2_{jj})(1-2\psi_{kk}\cos\lambda+\psi^2_{kk})}. \tag{32}$$

Thus, we obtain

$$\frac{|f_{jk}(\lambda)|^2}{f_{jj}(\lambda)f_{kk}(\lambda)} = \frac{\Sigma^2_{u,jk}}{\Sigma_{u,jj}\Sigma_{u,kk}},$$

as stated. □

References

Bollerslev, T. & R.F. Engle (1993) Common persistence in conditional variances. *Econometrica* 61, 167–186.

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

Boudjellaba, H., J.-M. Dufour & R. Roy (1992) Testing causality between two vectors in multivariate autoregressive moving average models, *Journal of the American Statistical Association* 87, 1082–1090.

Brockwell, P.J. and R.A. Davis (1991) *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag.

Broze, L., C. Francq & J.-M. Zakoian (2001) Non redundancy of high order moment conditions for efficient GMM estimation of weak AR processes. *Economics Letters* 71, 317–322.

Comte, F. and O. Lieberman (2000a) Asymptotic theory for multivariate GARCH processes, discussion paper, University of Paris 6 and 7.

Comte, F. and O. Lieberman (2000b) Second order noncausality in multivariate GARCH processes, *Journal of Time Series Analysis* 21, 535–557.

Diebold, F.X. & M. Nerlove (1989) The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics* 4, 1–21.

Drost, F. C. & T. E. Nijman (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909–927.

Drost, F. C. & B. J. M. Werker (1996) Closing the GARCH gap: continuous time GARCH modeling. *Journal of Econometrics* 74, 31–57.

Engle, R.F. & K.F. Kroner (1995) Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.

Engle, R.F., V.M. Ng & Rothschild (1990) Asset pricing with a factor-ARCH structure: empirical estimates for Treasury Bills. *Journal of Econometrics* 45, 213–237.

Gallant, A.R., P.E. Rossi & G. Tauchen (1993) Nonlinear dynamic structures. *Econometrica* 61, 871–907.

Geweke, J. (1982) Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association* 77, 304–313.

Gouriéroux, C. & A. Monfort (1995) *Séries Temporelles et Modèles Dynamiques*, Economica, Paris.

Gouriéroux, C. (1997) *ARCH Models and Financial Applications.* New York: Springer.

Granger, C.W.J., R.P. Robins, & R.F. Engle (1984). Wholesale and retail prices: Bivariate time series modelling with forecastable error variances, in: D. Belsley and E. Kuh, eds., *Model reliability*, MIT Press, Cambridge, MA.

Hafner, C.M. & H. Herwartz (1998) Structural analysis of portfolio risk using beta impulse response functions. *Statistica Neerlandica* 52, 336–355.

Hafner, C.M. & H. Herwartz (2000) Testing for linear autoregressive dynamics under heteroskedasticity, *The Econometrics Journal* 3, 177–197.

Hafner, C.M. & H. Herwartz (2001) Volatility impulse response functions for multivariate GARCH models. CORE discussion paper, Université catholique de Louvain, Louvain-la-Neuve, Belgium. Revised version of CORE discussion paper 9847.

He, C. & T. Teräsvirta (1999) Fourth moment structure of the GARCH(p, q) process. *Econometric Theory* 15, 824–846.

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models, *Econometric Theory* 14, 70–86.

Karanasos, M. (1999) The second moment and the autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics* 90, 63–76.

Koop, G., Pesaran, M.H., Potter, S.M. (1996), Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* 74, 119–147.

Kroner, K.F. & V.K. Ng (1998) Modeling asymmetric comovements of asset returns. *The Review of Financial Studies* 11, 817–844.

Lütkepohl, H. (1993) *Introduction to Multiple Time Series Analysis*, 2nd ed., Berlin, New York: Springer.

Magnus, J.R. (1988) *Linear Structures*. London: Griffin.

Mahieu, R. (1995) *Financial Market Volatility – Statistical Models and Empirical Results*, PhD thesis, Universitaire Pers Maastricht, The Netherlands.

Mardia, K.V. (1970) Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 519–530.

Meddahi, N. & E. Renault (1997) Linear statistical inference for ARCH-type processes, GREMAQ, Université de Toulouse, France.

Rose, A.K., Svensson, L.E.O. (1994) European exchange rate credibility before the fall. *European Economic Review* 38, 1185–1216.

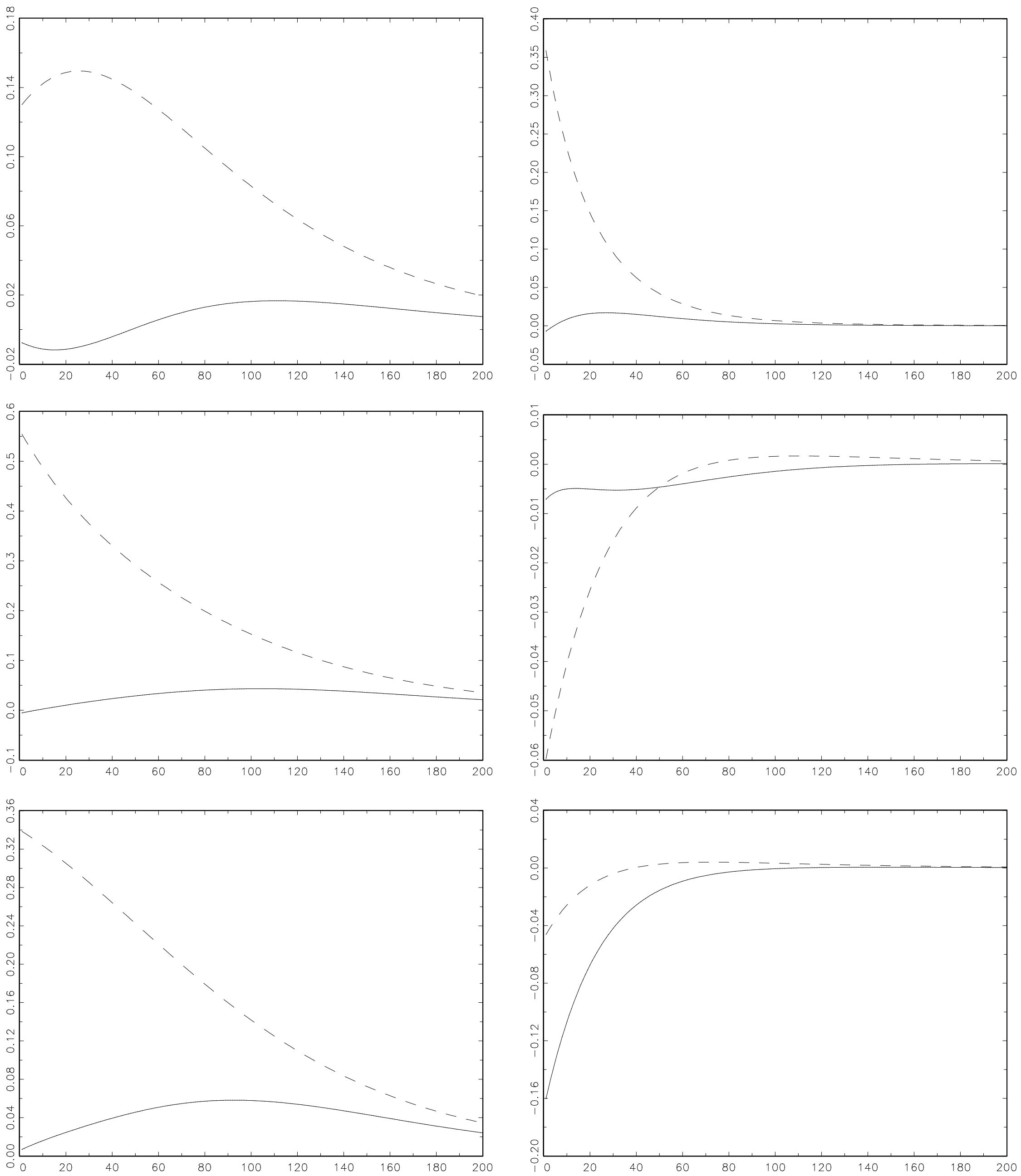

Figure 1: *Impulse response functions for* $\xi_t \sim N(0, I_K)$. *Left panel: shock on Sept 16, 1992 (Black Wednesday), right panel: shock on August 2, 1993. First row: Impulse response functions of DEM/USD kurtosis (solid) and volatility (dashed). Second row: Impulse response functions of GBP/USD kurtosis (solid) and volatility (dashed). Third row: Impulse response functions of the co-kurtosis (solid) and covariance (dashed).*

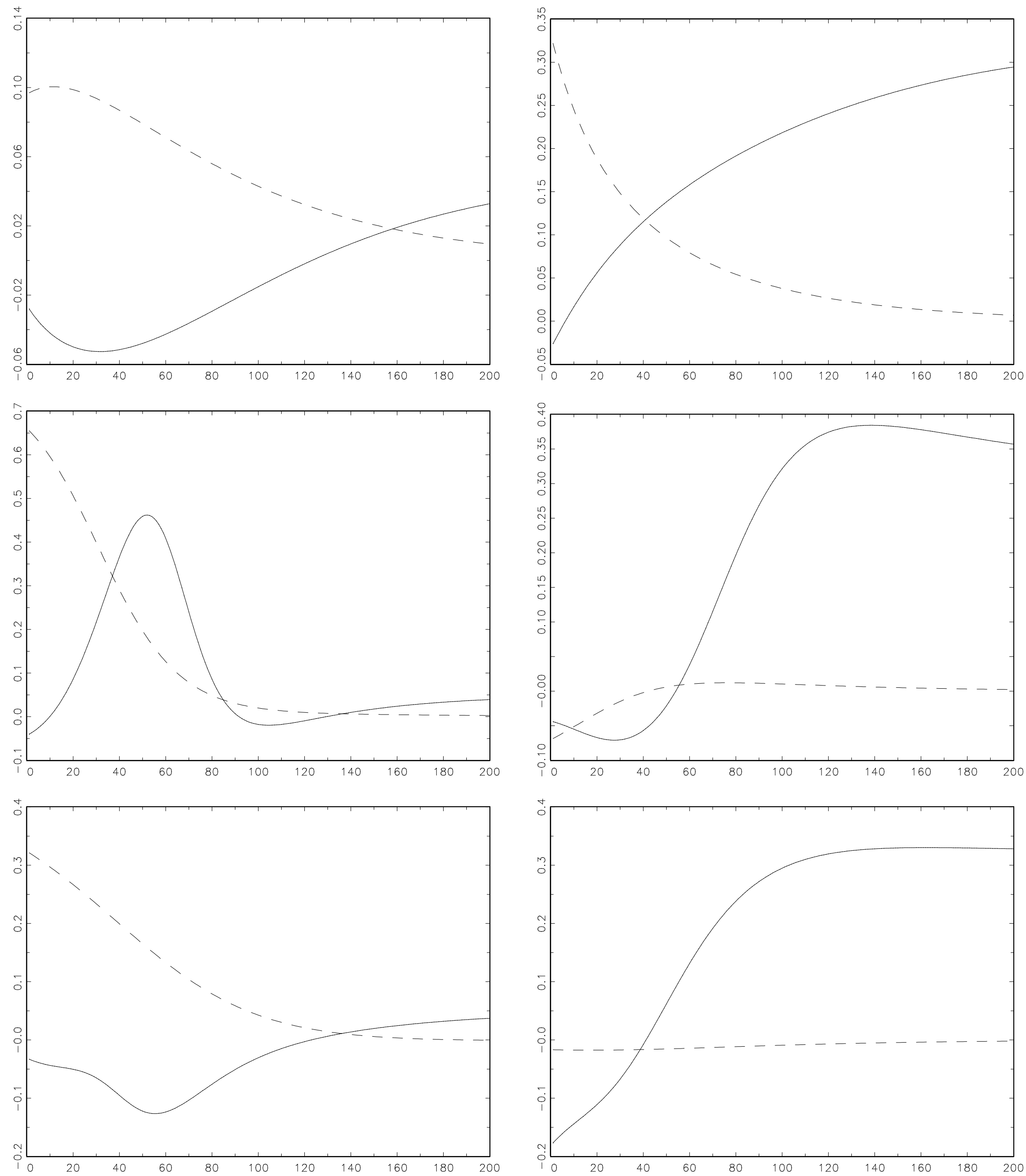

Figure 2: *Impulse response functions for* $\xi_t \sim t_\nu$. *Left panel: shock on Sept 16, 1992 (Black Wednesday), right panel: shock on August 2, 1993. First row: Impulse response functions of DEM/USD kurtosis (solid) and volatility (dashed). Second row: Impulse response functions of GBP/USD kurtosis (solid) and volatility (dashed). Third row: Impulse response functions of the co-kurtosis (solid) and covariance (dashed).*

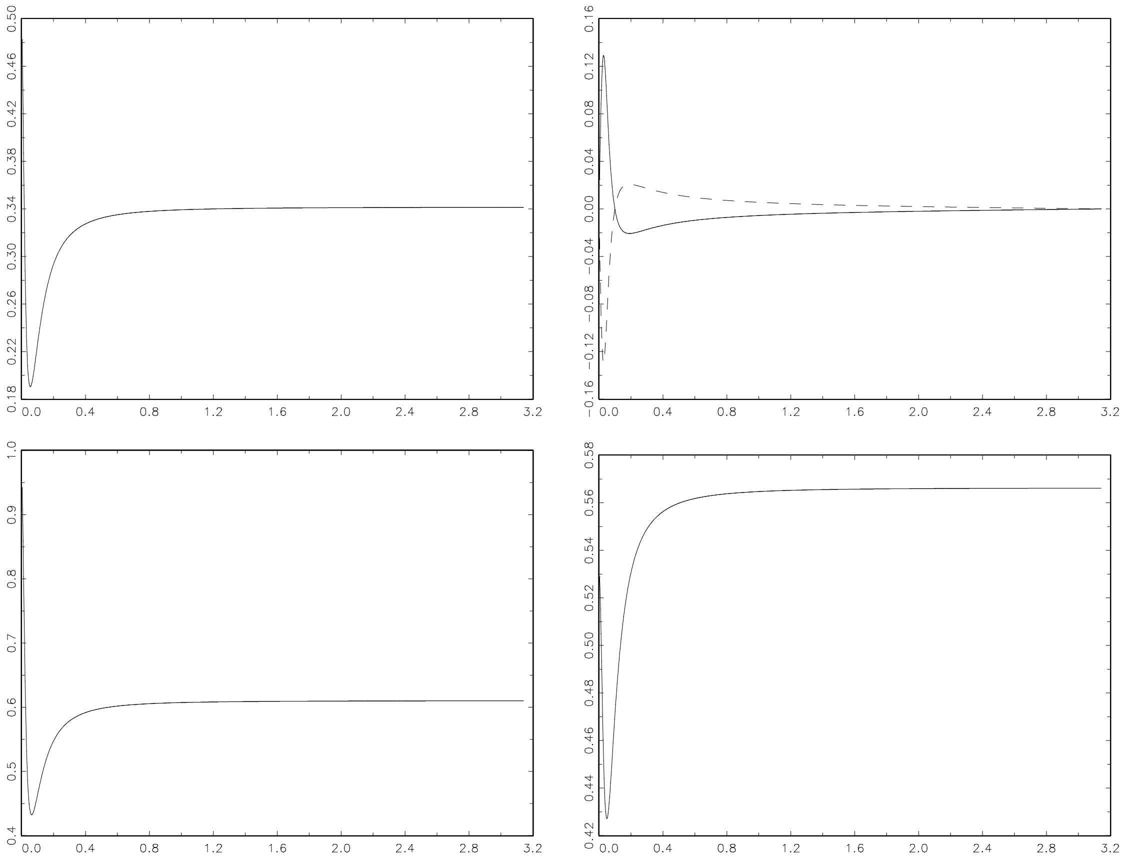

Figure 3: *Frequency domain characteristics of a multivariate GARCH(1,1) model fitted to the residuals of DEM/USD (ε_{1t}) and GBP/USD (ε_{2t}). Upper left: theoretical coherence between ε^2_{1t} and ε^2_{2t}. Upper right: theoretical phase of a frequency domain regression of ε^2_{2t} on ε^2_{1t} (solid) and of ε^2_{1t} on ε^2_{2t} (dashed). Lower left: Theoretical gain of a frequency domain regression of ε^2_{2t} on ε^2_{1t}. Lower right: Theoretical gain of a frequency domain regression of ε^2_{1t} on ε^2_{2t}.*