

VOLATILITY IMPULSE RESPONSE FUNCTIONS FOR MULTIVARIATE GARCH MODELS

Christian M. Hafner¹ and Helmut Herwartz²

September 2001

Abstract

In the empirical analysis of financial time series, multivariate GARCH models have been used in various forms. As it is typical for nonlinear models there is yet no unique framework available to uncover dynamic covariance relationships for vector return processes. We introduce a new concept of impulse response functions tracing the effects of independent shocks on volatility through time. The advocated methodology avoids typical orthogonalization and ordering problems. Theoretical properties of volatility impulse response functions are derived and compared with conditional moment profiles introduced by Gallant, Rossi and Tauchen (1993) for semi-nonparametric models. In an empirical study of a bivariate foreign exchange rate series we use volatility impulse response functions to compare alternative parametric volatility specifications. It is shown that for shocks affecting foreign exchange rates in an asymmetric way, the difference between our methodology and conditional volatility profiles can be substantial.

Keywords: Multivariate GARCH, impulse response, exchange rate, volatility

JEL Classification: C22

¹Electrabel, R&D Energy Markets, Place de l'Université 16, B-1348 Louvain-la-Neuve, Belgium. e-mail christian.hafner@electrabel.com.

²Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität zu Berlin, Spandauer Str.1, D-10178 Berlin, Germany. e-mail helmut@wiwi.hu-berlin.de.

This paper is a revised version of CORE discussion paper 9847. Both authors would like to thank Luc Bauwens, Jeroen Rombouts and Helmut Lütkepohl for helpful comments. Financial support by the Belgian Government and the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

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 authors.

1 Introduction

Multivariate GARCH models have been designed to model the conditional covariance matrix of multiple time series. The knowledge of this volatility matrix may give valuable information on risk measures associated with a given set of financial assets. It has been frequently observed both in univariate and multivariate frameworks that volatility changes over time. Given a multivariate parametric volatility model or an estimate thereof, there is no unified methodology available to uncover volatility dynamics operating between the involved variables. Since there are similarities between GARCH and VARMA-type models, it is appealing to generalize impulse response analysis as introduced by Sims (1980) to the analysis of shocks in volatility. A number of proposals have been made to trace the impact of some kind of shocks through time (Koop, Pesaran and Potter 1996, Engle and Ng 1993, Gallant, Rossi and Tauchen 1993, Lin 1997). In this paper we provide a new look at this issue. Volatility impulse response functions (VIRF) as defined in our paper will provide information on the impact of independent shocks on volatility.

Koop et al. (1996) define generalized impulse response functions for the conditional mean using the mean of the response vector conditional on history and a present shock, compared with a baseline that only conditions on history. Since in nonlinear models these conditional expectations are often not available analytically, they suggest to resort to Monte Carlo techniques. In their general framework, impulse responses depend on history, shock, intermediate innovations and model parameters, all of which can be regarded as random variables. Although our concept of VIRF is in the spirit of Koop et al. (1996), there is an important difference additional to the fact that we look at the conditional variance rather than the conditional mean: GARCH models can be viewed as being linear in squares, and multivariate GARCH models are known to have a VARMA representation with non-Gaussian errors. This particular structure allows us to calculate conditional expectations of volatility analytically. Moreover, our results are compared with traditional impulse response functions of VARMA systems: Some results such as shock persistence are analogous, others are different. For example, the distribution of impulse responses is no longer symmetric and the VIRF depend on the volatility state at the time when the shock occurs.

We provide an extensive comparison of VIRF with the conditional moment profiles proposed by Gallant et al. (1993, GRT hereafter). These conditional moment profiles have been introduced originally as a means to illustrate dynamic relationships between variables within a semi-nonparametric model. Adopted to parametric GARCH models, volatility profiles can easily be evaluated analytically. The definition of shock or news by GRT and Engle and Ng (1993) is a perturbation in the conditionally heteroskedastic error vector, ε_t say, which can either be observed or estimated. Apart from being dependent on the past, the components of ε_t are in general contemporaneously correlated.

Therefore, impulse response analysis is mostly undertaken for orthogonal transformations of ε_t . These transformations are derived from a causality scheme that the researcher has to assume *a priori*. Prespecifying causality patterns is in widespread use when analyzing macroeconomic systems. For financial time series, however, it appears hardly feasible to impose realistic causality structures *a priori*, because they are typically highly interrelated and observed at high frequencies.

News is inherently independent over time. If there were any predictable components of news, one could construct hedge portfolios that would at least partially eliminate the risks associated with them. This contradicts our intuition of news being unsystematic risk and unhedgeable. Therefore, we are defining news to be risk sources that are independent over time and thus unpredictable.

Having specified and estimated the multivariate distribution of the independent innovations, the fact that news is considered as independent and identically distributed allows us to generate realistic shock scenarios by drawing from this distribution. We propose a means to calculate VIRF for the vec representation of a multivariate GARCH model, which nests most popular multivariate GARCH specifications. We show that our new concept improves the understanding of the highly complex dynamic behavior of financial time series. In an empirical study, we compare two parsimonious representations of multivariate GARCH models, namely the BEKK-model (Engle and Kroner 1995) and its diagonal version, applied to a bivariate foreign exchange rate (FX-rate) series. For the impulse response functions, we first consider two historically observed shocks, one having a similar effect on both FX-rate volatilities, the other having an asymmetric effect. We show that GRT and our concept may imply quite different impulse responses in the case of asymmetric shocks. Furthermore, we consider shocks that are drawn randomly from the estimated innovation distribution. To describe the dispersion of impulse responses we use standard deviations and selected quantiles. One result is that the distribution of conditional volatility profiles as suggested by GRT is quite symmetric, whereas the distribution of VIRF is typically highly skewed.

The paper is organized as follows. In Section 2 we discuss identification of independent news. Section 3 briefly sketches representation and estimation issues of multivariate GARCH models. VIRF are defined in Section 4. We discuss thoroughly the implementation of the new method and compare it with alternative versions of conditional volatility profiles as advocated by GRT. Section 5 presents an empirical analysis of a bivariate FX-rate series, and Section 6 concludes.

2 Identification of news in multiple time series

In this section, we will address the issue of identifying news in multiple time series. The general problem is that the error vector shows contemporaneous correlation and, therefore,

the error components cannot be treated as news coming from independent sources. Let ε_t denote an N -dimensional random vector such that

$$\varepsilon_t = P_t \xi_t, \quad (1)$$

where $P_t P_t' = \Sigma_t$ and ξ_t denotes an *i.i.d.* random vector of dimension N with mean zero and covariance matrix the identity matrix I_N . Assuming that Σ_t is measurable with respect to the information set available at time $t-1$, $\mathcal{F}_{t-1}$, (1) implies that $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ and $\text{Var}[\varepsilon_t | \mathcal{F}_{t-1}] = \Sigma_t$. For example, ε_t could be the error of a VARMA process. If ε_t is a multivariate GARCH process as discussed in Section 3, then (1) may be called a strong GARCH model according to the terminology of Drost and Nijman (1993). This strong GARCH form is employed because we define news to appear in the *i.i.d.* innovation ξ_t .

In empirical practice, ε_t and Σ_t are either observed or can be estimated conveniently. Given ε_t and Σ_t , the identification of underlying innovations ξ_t has become a crucial issue in the analysis of vector autoregressive processes (Sims 1980, Lütkepohl 1993). A prominent solution to identify ξ_t is to assume that P_t is a lower triangular matrix, i.e., to use a Choleski decomposition of Σ_t . In this case, the elements of $\xi_t = P_t^{-1} \varepsilon_t$ depend recursively on the elements of the observation vector ε_t . With this type of orthogonalization, ξ_t will depend on the ordering of variables in ε_t which is often difficult to justify economically. Alternatively, one may consult economic theory to specify convenient zero restrictions in P_t . The latter approach is in widespread use for the structural analysis of dynamic macroeconomic systems. With respect to time series which are observed at high frequencies, daily say, structural assumptions may hardly be available. Note that instead of imposing only contemporaneous restrictions, one might also use long run restrictions to identify the structural innovations. Although identified by means of zero restrictions in P_t , the elements of ξ_t are, in general, not independently distributed and therefore fail the definition of news motivated above.

Independent news can often be identified via a Jordan decomposition of Σ_t . Let λ_{ti} , $i = 1, \dots, N$, denote the eigenvalues of Σ_t with corresponding eigenvectors γ_{ti} . The symmetric matrix $\Sigma_t^{1/2}$ is defined as

$$\Sigma_t^{1/2} = \Gamma_t \Lambda_t^{1/2} \Gamma_t',$$

with $\Gamma_t = (\gamma_{t1}, \dots, \gamma_{tN})$ and $\Lambda_t = \text{diag}(\lambda_{t1}, \dots, \lambda_{tN})$. Obviously, one has $\Sigma_t^{1/2} \Sigma_t^{1/2} = \Sigma_t$. Since setting $P_t = \Sigma_t^{1/2}$ avoids to impose any zero restrictions in identifying $\xi_t = \Sigma_t^{-1/2} \varepsilon_t$, the latter approach is purely a-theoretic. Under conditional normality of ε_t , however, ξ_t is not unique. To see this, consider *quasi innovations* ξ_t^* obtained after an orthogonal transformation,

$$\xi_t^* = R \xi_t, \quad R R' = I_N. \quad (2)$$

Due to the orthonormality of R , ξ_t^* is also a random vector with mean zero and covariance equal to the identity matrix. Indeed, if ξ_t is a vector of Gaussian random variables the

distributional properties of ξ_t^* and ξ_t cannot be distinguished. If, on the other hand, ξ_t is not normally distributed, then the components of quasi innovations ξ_t^* are dependent and thus fail our definition of news. This is in fact a characterization of the multivariate normal distribution and an important argument for our approach: If ξ_t is a vector of independent standardized variates with finite cumulants, and the non-trivial transformation (2) gives a vector of independent standardized variates, then each component of ξ_t^* is normal (and hence so is each component of ξ_t). This result due to Lancaster (1954) can be found, for example, in Kendall, Stuart and Ord (1987, pp. 499). It tells us that under the requirement of independence, the only situation where non-identifiability occurs is the case of a normal distribution. In other words, news can be considered as identified if the innovation vector is not normally distributed.

A well documented empirical result for financial time series is that residuals are fat tailed even after standardizing with a time-varying standard deviation (Bollerslev 1987), i.e. $E[\xi_t^4] > 3$. Therefore, we regard innovations ξ_t to be identified in empirical practice. Note that estimated innovations $\hat{\xi}_t$ are often used for diagnostic checking of univariate or multivariate GARCH models. Unlike under the often used Choleski decomposition, the approach advocated here provides innovation estimates and thus diagnostic results which are invariant with respect to the ordering of variables in the system.

3 Multivariate GARCH Models – Specification and Estimation

From a theoretical point of view the generalization of the univariate GARCH(p, q)–model (Bollerslev 1986) to the multivariate case is straightforward. As in the univariate GARCH(p, q) case one may allow Σ_t to depend on past observations ε_{t-i} , $i = 1, \dots, q$, and covariance matrices Σ_{t-i} , $i = 1, \dots, p$. In the multivariate framework, however, possible dependencies easily become intractable for empirical work. Let $\text{vech}(\cdot)$ denote the operator that stacks the lower fraction of an $N \times N$ –matrix into a $N^* = N(N + 1)/2$ dimensional vector. The general multivariate GARCH(p, q) model is given as:

$$\text{vech}(\Sigma_t) = c + \sum_{i=1}^q A_i \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}') + \sum_{i=1}^p B_i \text{vech}(\Sigma_{t-i}). \quad (3)$$

In the so–called vec representation (3), A_i and B_i are parameter matrices each containing $(N^*)^2$ parameters. The vector c accounts for time invariant variance components and contains N^* coefficients. Note that even in the case $N = 2$ and $p = q = 1$, 21 parameters characterize the dynamic relationship between ε_t and its history. To obtain tractable models for empirical work one usually has to impose parameter restrictions on the vec model (3). We will assume that all eigenvalues of the matrix $\sum_{i=1}^{\max(p,q)} (A_i + B_i)$ have modulus smaller than one, in which case the vector process ε_t is covariance stationary

(Engle and Kroner 1995). Under this assumption, the unconditional covariance matrix $\Sigma = \text{Var}(\varepsilon_t)$ is given by

$$\text{vech}(\Sigma) = \left(I_{N^*} - \sum_{i=1}^{\max(p,q)} (A_i + B_i) \right)^{-1} c,$$

where we set $A_{q+1} = \dots = A_p = 0$ if $p > q$ and $B_{p+1} = \dots = B_q = 0$ if $q > p$.

Engle and Kroner (1995) discuss in detail a dynamic specification of the following form:

$$\Sigma_t = C_0 C_0' + \sum_{k=1}^K \sum_{i=1}^q A'_{ki} \varepsilon_{t-i} \varepsilon_{t-i}' A_{ki} + \sum_{k=1}^K \sum_{i=1}^p G'_{ki} \Sigma_{t-i} G_{ki}. \quad (4)$$

In (4), C_0 is a lower triangular matrix and A_{ki} and G_{ki} are $N \times N$ parameter matrices. The so-called BEKK representation in (4) copes with two issues evolving in multivariate GARCH modelling. First, given positive definite initial covariances $\Sigma_0, \dots, \Sigma_{1-p}$, sample covariances $\Sigma_t, t = 1, \dots, T$, are positive definite under the weak (sufficient) condition that at least one of the matrices C_0 or G_{ki} has full rank (Engle and Kroner 1995). Second, the BEKK-representation allows for direct dependence of the conditional variance of one variable on the history of other variables within the system. A diagonal parameterization of the matrices A_{ki} and G_{ki} (diagonal BEKK) yields a restricted version of the so-called diagonal model proposed by Bollerslev, Engle and Wooldridge (1988). For the present analysis we take $K = 1$ and $p = q = 1$. In this case the assumption that the upper left elements of A_{11} and G_{11} are greater than zero is sufficient for the model parameters to be identified.

The estimation of the parameters in A_{11} and G_{11} is analogous to the univariate case. For the empirical part of this study we used the BHHH-algorithm introduced by Berndt et al. (1974) to maximize the Gaussian log-likelihood function. The contribution of $\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \Sigma_t)$ to the joint log-likelihood of a sample with T observations ($\log L = \sum_{t=1}^T l_t$) is:

$$l_t = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_t| - \frac{1}{2} \varepsilon_t' \Sigma_t^{-1} \varepsilon_t. \quad (5)$$

If the conditional distribution of ε_t is not normal, (5) is the quasi-likelihood function the maximization of which has become popular as quasi maximum likelihood (QML) estimation.

In empirical applications of univariate GARCH processes it has often been found that standardized residuals of estimated processes still have excess kurtosis. To take conditional leptokurtosis into account, Bollerslev (1987) advocates to evaluate (and maximize) the sample log-likelihood under the assumption of underlying t-distributed innovations. For this reason we utilize alternatively to the Gaussian distribution the multivariate t-distribution to specify the (quasi) log-likelihood function. The latter distribution is parameterized with mean zero, ν degrees of freedom, scaling matrix Ω_t , and has the following

density function $f(\cdot)$ (Johnson and Kotz 1972, pp. 134):

$$f(\varepsilon_t|0, \Omega_t, \nu) = c_t^{-1} |\Omega_t|^{-1/2} [\nu + \varepsilon_t' \Omega_t^{-1} \varepsilon_t]^{-(\nu+N)/2}, \quad (6)$$

with

$$c_t = \frac{\pi^{N/2} \Gamma(\nu/2)}{\nu^{\nu/2} \Gamma((\nu+N)/2)}. \quad (7)$$

In (7), $\Gamma(\cdot)$ is the gamma function. Under the model in (6), the following result relates the conditional variance of ε_t and Ω_t :

$$\Sigma_t = \text{Var}[\varepsilon_t | \mathcal{F}_{t-1}] = \frac{\nu}{\nu-2} \Omega_t, \quad \nu > 2.$$

The joint density of any subset of ε_t is again of the form (6) with appropriately modified moments. In particular, all marginal distributions share the same degrees of freedom parameter ν . Under the assumption of conditionally t-distributed error terms the contribution of ε_t to the log-likelihood is $l_t = \ln f(\varepsilon_t|0, \Omega_t, \nu)$.

In the multivariate framework, results for the asymptotic properties of the (Q)ML-estimator have been derived recently. Jeantheau (1998) proves the QML-estimator to be consistent under the main assumption that the considered multivariate process is strictly stationary and ergodic. Further assuming finiteness of moments of ε_t up to order eight, Comte and Lieberman (2000) derive asymptotic normality of the QML-estimator. The asymptotic distribution of the rescaled QML-estimator is analogous to the univariate case discussed in Bollerslev and Wooldridge (1992). Collecting the parameters of the empirical model in a vector θ , we apply the following result for inference on significance of parameter estimates:

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, D^{-1} S D^{-1}). \quad (8)$$

In (8), S is the expectation of the outer product of first order derivatives of $l_t(\theta)$ with respect to θ and D is the negative expectation of the matrix of second order derivatives.

4 Volatility Impulse Response Functions

This section provides VIRF as a new methodology to evaluate the dynamic impact of shocks on volatility. The GRT framework treats a shock as being added to the data. Conversely, our approach is similar to Koop et al. (1996) and regards a shock as being generated from the data generating process. In the latter case there is no difficulty in considering ‘realistic’ shocks, since they can be drawn from the estimated distribution of the innovations. We will see in Section 4.2 that in the former case it may be more difficult to construct ‘realistic’ shocks in combination with ‘realistic’ baselines in the framework of multivariate volatility models.

The formal representation of VIRF is given for the vec specification of multivariate GARCH models. Since for every BEKK model there exists a unique equivalent vec specification (Engle and Kroner 1995), corresponding results for BEKK models can be obtained by first transforming the model to its vec representation.

4.1 Independent shocks generated by the DGP

At time $t = 0$ some independent news is reflected by ξ_0 and it is not specified whether the news is ‘good’ or ‘bad’, as indicated by the sign of the individual components, and how important the news is, as indicated by the size. Because of our assumption that innovations ξ_t are i.i.d., one may consider a shock as being drawn from the distribution of ξ_t . In general, the impulse response is a function of the shock and, therefore, will have a certain distribution as well. The perspective of regarding the impulse response functions as random variables is similar to that in Koop et al. (1996). In traditional impulse response analysis, for instance, impulse responses are linear in the shock. If the shock is normally distributed, then so is the response at any horizon and the variance (or standard deviation) is sufficient to describe its distribution. In nonlinear models such as GARCH, however, the latter distribution in general will not be symmetric.

The conditional covariance matrix Σ_t is a function of the innovations $\xi_1, \dots, \xi_{t-1}$, the initial shock ξ_0 and Σ_0 . VIRF are now defined as the expectation of volatility conditional on an initial shock and history, subtracted by the baseline expectation that only conditions on history, i.e.

$$V_t(\xi_0) = E[\text{vech}(\Sigma_t) \mid \xi_0, \mathcal{F}_{-1}] - E[\text{vech}(\Sigma_t) \mid \mathcal{F}_{-1}]. \quad (9)$$

In (9), $V_t(\xi_0)$ is an N^* -dimensional vector. For example, if $N = 2$ ($N^* = 3$), the first and third element of $V_t(\xi_0)$ represent impulse responses of the conditional variances of the first and second variable, respectively, and the second element of $V_t(\xi_0)$ is the response of the conditional covariance.

For illustration, let us consider the multivariate GARCH(1,1) case. Starting with $t = 1$, we obtain

$$\begin{aligned} V_1(\xi_0) &= A_1 \left\{ \text{vech}(\Sigma_0^{1/2} \xi_0 \xi_0' \Sigma_0^{1/2}) - \text{vech} \Sigma_0 \right\} \\ &= A_1 D_N^+(\Sigma_0^{1/2} \otimes \Sigma_0^{1/2}) D_N \text{vech}(\xi_0 \xi_0' - I_N), \end{aligned}$$

where D_N denotes the duplication matrix defined by the property $\text{vec}(Z) = D_N \text{vech}(Z)$ for any $(N \times N)$ matrix Z , and D_N^+ denotes its Moore-Penrose inverse. For any $t \geq 2$,

$$V_t(\xi_0) = (A_1 + B_1)^{t-1} A_1 D_N^+(\Sigma_0^{1/2} \otimes \Sigma_0^{1/2}) D_N \text{vech}(\xi_0 \xi_0' - I_N) \quad (10)$$

$$= (A_1 + B_1) V_{t-1}(\xi_0). \quad (11)$$

Since the innovation vector ξ_0 enters into the volatility equation only in the form $\xi_0 \xi_0'$,

we immediately get the following result:

$$V_t(\xi_0) = E \left[\frac{\partial \text{vech}(\Sigma_t)}{\partial \text{vech}(\xi_0 \xi_0')'} \middle| \mathcal{F}_{-1} \right] \text{vech}(\xi_0 \xi_0' - I_N). \quad (12)$$

That is, the VIRF are equal to the impact of an infinitesimal change in $\text{vec}(\xi_0 \xi_0')$ on volatility, scaled by the actual centered (squared) innovation vector $\text{vech}(\xi_0 \xi_0' - I_N)$. In the symmetric multivariate GARCH model, the conditional variance depends on past innovations only of the form $\text{vech}(\xi_{t-i} \xi_{t-i}')$. Therefore, equation (12) holds exactly. In asymmetric or nonlinear models, however, the right hand side of (12) could only be a first order approximation to VIRF.

To find an explicit expression of $V_t(\xi_0)$ for a multivariate GARCH(p, q) model, consider the VARMA representation. Let us define $\eta_t = \text{vech}(\varepsilon_t \varepsilon_t')$. By rearranging terms, the multivariate GARCH(p, q) model can be represented as a VARMA ($\max(p, q), p$) model,

$$\eta_t = \omega + \sum_{i=1}^{\max(p, q)} (A_i + B_i) \eta_{t-i} - \sum_{j=1}^p B_j u_{t-j} + u_t, \quad (13)$$

where $u_t = \eta_t - \text{vech}(\Sigma_t)$ is a white noise vector, i.e., $E[u_t] = 0$, $E[u_t u_t'] = \Sigma_u$ and $E[u_t u_s'] = 0$ for $s \neq t$. From the VARMA representation (13) it is possible to obtain the VMA(∞) representation

$$\eta_t = \text{vech}(\Sigma) + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \quad (14)$$

where the $N^* \times N^*$ matrices Φ_i can be determined recursively by $\Phi_0 = I_{N^*}$, $\Phi_i = -B_i + \sum_{j=1}^i (A_j + B_j) \Phi_{i-j}$, $i = 1, 2, \dots$, see Lütkepohl (1993, pp. 220). Obviously, we have

$$E \left[\frac{\partial u_t}{\partial \text{vech}(\xi_0 \xi_0')'} \middle| \mathcal{F}_{-1} \right] = 0$$

for all $t \geq 1$, and therefore

$$\begin{aligned} E \left[\frac{\partial \text{vech}(\Sigma_t)}{\partial \text{vech}(\xi_0 \xi_0')'} \middle| \mathcal{F}_{-1} \right] &= E \left[\frac{\partial \eta_t}{\partial \text{vech}(\xi_0 \xi_0')'} \middle| \mathcal{F}_{-1} \right] \\ &= \Phi_t E \left[\frac{\partial u_0}{\partial \text{vech}(\xi_0 \xi_0')'} \middle| \mathcal{F}_{-1} \right] \\ &= \Phi_t E \left[\frac{\partial \eta_0}{\partial \text{vech}(\xi_0 \xi_0')'} \middle| \mathcal{F}_{-1} \right] \\ &= \Phi_t D_N^+ (\Sigma_0^{1/2} \otimes \Sigma_0^{1/2}) D_N. \end{aligned}$$

By equation (12), we finally obtain

$$V_t(\xi_0) = \Phi_t D_N^+ (\Sigma_0^{1/2} \otimes \Sigma_0^{1/2}) D_N \text{vech}(\xi_0 \xi_0' - I_N). \quad (15)$$

In the case of a multivariate GARCH(1,1), we have $\Phi_t = (A_1 + B_1)^{t-1} A_1$ and we get equation (10) as a special case. Several points should be noted about (15), in particular when comparing the VIRF to the properties of traditional impulse response analysis of the conditional mean in linear systems, called traditional analysis hereafter.

1. In the traditional analysis, a shock δ has the opposite effect of $-\delta$. For the VIRF in (15), we have $V_t(\xi_0) = V_t(-\xi_0)$. In other words, the impulse response is an even function of the shock, as opposed to an odd function in the traditional analysis.
2. In the traditional analysis, shock linearity holds, i.e., a shock $k\delta$ has k times the effect of a shock δ . For $V_t(\xi_0)$ in (15), we do not have such a property. More generally, VIRF are not homogeneous functions of any degree.
3. In the traditional analysis, impulse response functions do not depend on the history of the process. The elements of $V_t(\xi_0)$ depend on the history, but only through the volatility state Σ_0 at the time when the shock occurs.
4. The decay or persistence of shocks is given by the moving average matrices Φ_t , analogously to the traditional analysis.

Considering the shock ξ_0 as being randomly drawn from its distribution, VIRF themselves become random functions. The first two moments are given by

$$E[V_t(\xi_0)] = 0, \quad (16)$$

$$\text{Var}[V_t(\xi_0)] = Z_1 \Sigma_{\xi\xi'} Z_1', \quad (17)$$

with $Z_1 = \Phi_t D_N^+ (\Sigma_0^{1/2} \otimes \Sigma_0^{1/2})$, and $\Sigma_{\xi\xi'} = \text{Var}[\text{vec}(\xi_0 \xi_0')]$. For example, in the case of $\xi_0 \sim N(0, I_N)$, the variance of $\text{vec}(\xi_0 \xi_0')$ is given by $I_{N^2} + C_{NN}$, where C_{NN} is the commutation matrix, see e.g. Magnus (1988). In the case of a multivariate t-distribution with $\nu > 4$, the following result is easily obtained:

$$\Sigma_{\xi\xi'} = I_{N^2} + C_{NN} + \frac{6}{\nu - 4} \sum_{i=1}^N e_{l(i)} e_{l(i)}',$$

where $l(i) = 1 + (i - 1)(N + 1)$ and e_j is a vector with 1 at the j -th position and zeros elsewhere. Despite the usefulness of the variance, it is not sufficient in describing the distribution of $V_t(\xi_0)$ which will typically be asymmetric and complicated to derive analytically. This is because a linear combination of the components of the vector $\text{vech}(\xi_0 \xi_0' - I_N)$ involves in the Gaussian case a linear combination of noncentral chi-square distributions and a mixture of normals. In the empirical part of the paper, we will therefore provide simulated quantiles to describe the distribution of $V_t(\xi_0)$.

4.2 External shocks in observables

In the following, we compare our approach with the methodology of GRT. The difference to the method outlined above is twofold. First, GRT assume a shock to occur in ε_t , which is either directly observable or is obtained after the correction for the conditional mean of the original process. Second, they compare the volatility profile relative to a baseline ε_0^0 . It is this latter fact where we will point out some inconvenient features of this definition in the context of volatility analysis.

4.2.1 Conditional volatility profiles

GRT define conditional moment profiles, which in the volatility case can be written as

$$v_t(\delta) = E[\text{vech}(\Sigma_t) \mid \varepsilon_0^*, \mathcal{F}_{-1}] - E[\text{vech}(\Sigma_t) \mid \varepsilon_0^0, \mathcal{F}_{-1}], \quad (18)$$

where $\varepsilon_0^* = \varepsilon_0^0 + \delta$, and $\delta = (\delta_1, \dots, \delta_N)'$ is the impulse vector.

Since $\text{vech}(\Sigma_t)$ is linear in lagged $\eta_t = \text{vech}(\varepsilon_t \varepsilon_t')$, using again the VMA(∞) representation in (14) we have

$$\begin{aligned} E \left[\frac{\partial \text{vech}(\Sigma_t)}{\partial \eta_0'} \middle| \mathcal{F}_{-1} \right] &= E \left[\frac{\partial \eta_t}{\partial \eta_0'} \middle| \mathcal{F}_{-1} \right] \\ &= \Phi_t \frac{\partial u_0}{\partial \eta_0'} \\ &= \Phi_t. \end{aligned}$$

Then, with $\eta_0^* = \text{vech}(\varepsilon_0^* \varepsilon_0^{*'})$ and $\eta_0^0 = \text{vech}(\varepsilon_0^0 \varepsilon_0^{0'})$, we have

$$\begin{aligned} v_t(\delta) &= E \left[\frac{\partial \text{vech}(\Sigma_t)}{\partial \eta_0'} \middle| \mathcal{F}_{-1} \right] (\eta_0^* - \eta_0^0) \\ &= \Phi_t \text{vech}(\delta \delta' + \varepsilon_0^0 \delta' + \delta \varepsilon_0^{0'}). \end{aligned} \quad (19)$$

Comparing (19) with (15), we note that both v_t and V_t have the same decay or degree of persistence over time, given by the moving average parameter matrices Φ_t . Moreover, if the baseline is not zero, then $v_t(\delta)$ is neither an odd nor an even function, nor is it homogenous of any degree. If the baseline is zero, then $v_t(\delta)$ is an even function and homogeneous of degree two. That is, $v_t(k\delta) = k^2 v_t(\delta)$ if $\varepsilon_0^0 = 0$. $v_t(\delta)$ does not explicitly depend on the history, $\mathcal{F}_{-1}$, but the choice of the baseline and/or the shock may be based on information in the past, e.g., a specific volatility state Σ_0 as in the VIRF analysis outlined above.

4.2.2 Choice of baseline and shock

We now turn to the choice of the baseline and the shock, for which four distinctions can be made as to whether the baseline and shock are fixed or random (GRT only consider fixed shocks):

A) Both shock and baseline are fixed: When fixing a shock in one variable, δ_i , the question is how to choose a realistic contemporaneous shock in the other variables, $\delta_j, j \neq i$. In the bivariate setting, GRT consider different scenarios corresponding to alternative combinations of large, medium, and small shocks in both variables. This procedure is likely to become intractable for higher dimensions.

If the baseline is set to zero, then the conditional volatility profile simplifies to

$$v_t(\delta) = \Phi_t \text{vech}(\delta \delta'). \quad (20)$$

In standard impulse response analysis for the conditional mean, it is natural to set the baseline ε_0^0 to its unconditional mean zero, because this is the steady state of the process. In volatility analysis, on the other hand, the choice of the baseline in the GRT framework turns out to be a non-trivial issue. The reason is that with a fixed baseline vector ε_0^0 one cannot represent the steady state of volatility given by Σ . This is because the matrix $\varepsilon_0^0 \varepsilon_0^{0'}$ is singular if $N > 1$. In other words, there is no natural baseline ε_0^0 for volatility analysis, because any given baseline deviates from the average volatility state. For example, a zero baseline will represent the lowest possible volatility state and, therefore, volatility forecasts will increase over time, no matter how big the initial shock was. The baseline itself can be viewed as a shock to volatility and it becomes difficult to disentangle the effects of the baseline and the shock vector.

Concerning the choice of the shock, consider a ‘zero shock’ $\delta = 0$ and compare $v_t(0) = 0$ with the effect of a ‘zero shock’ $\xi_0 = 0$ in (15). Obviously, $v_t(0) = 0$ for all t , but $V_t(0) = -\Phi_t \text{vech}(\Sigma_0)$, which in most empirical applications will be negative for the conditional variance components of V_t . In other words, a ‘zero shock’ means no impact using v_t but a negative one using VIRF for conditional variances.

To give another example, consider the choice $\delta = \Sigma_0^{1/2} \xi_0$. If the baseline is fixed to zero, then v_t becomes

$$\begin{aligned} v_t(\Sigma_0^{1/2} \xi_0) &= \Phi_t \text{vech}(\Sigma_0^{1/2} \xi_0 \xi_0' \Sigma_0^{1/2}) \\ &= V_t(\xi_0) + \Phi_t \text{vech}(\Sigma_0). \end{aligned} \quad (21)$$

If the baseline is fixed but different from zero, then

$$v_t(\Sigma_0^{1/2} \xi_0) = V_t(\xi_0) + \Phi_t \text{vech}(\Sigma_0 + \varepsilon_0^0 \xi_0' \Sigma_0^{1/2} + \Sigma_0^{1/2} \xi_0 \varepsilon_0^{0'}).$$

It is, in general, not possible to find a specific baseline such that $v_t = V_t$ for all t . As outlined above, this is due to the inherent mixture of shock and baseline that both affect volatility forecasts.

If unit shocks $\delta = e_j$ combined with $\varepsilon_0^0 = 0$ are considered, we obtain the definition of impulse response functions given by Lin (1997). In this case, $\delta \delta'$ is a matrix with 1 at the jj -th position and zeros elsewhere. It should be emphasized that Lin’s definition is a special case of the GRT definition.

B) Fixed shock and random baseline: Turning to the second possibility for the choice of $(\varepsilon_0^0, \delta)$, let δ be fixed and ε_0^0 be randomly drawn from the unconditional distribution of ε_t . As GRT note, this gives the same result for the average of the impulse responses as setting the baseline to zero. The first moments are given by

$$E[v_t(\delta)] = \Phi_t \text{vech}(\delta \delta'), \quad (22)$$

$$\text{Var}[v_t(\delta)] = Z_2 \Sigma Z_2', \quad (23)$$

with $Z_2 = \Phi_t D_N^+(I_{N^2} + C_{NN})(\delta \otimes I_N)$.

A minor difficulty of the random baseline perspective in the GRT framework is that the joint unconditional distribution of ε_t is not known for the class of multivariate GARCH models. Therefore, random baselines would have to be simulated using the estimated model or be drawn with replacement from the observations.

One could also condition on a specific volatility state Σ_0 and draw from the innovations ξ_0 , as in the VIRF analysis. In that case, the mean is the same as in (22) and the variance becomes $\text{Var}[v_t(\delta)] = Z_2 \Sigma_0 Z_2'$, which coincides with (23) if $\Sigma_0 = \Sigma$. However, the distribution has fatter tails if ε_0^0 is drawn from the unconditional distribution.

C) Random shock and fixed baseline: We will now address the possibility of a random δ . Again, one could use the unconditional distribution of ε_t or, if not available, simulated data from the model. If the baseline is fixed and the unconditional distribution is used to generate δ , then we obtain

$$\begin{aligned} \mathbb{E}[v_t(\delta)] &= \Phi_t \text{vech}(\Sigma) \\ \text{Var}[v_t(\delta)] &= \Phi_t \Sigma_\eta \Phi_t' + Z_3 \Sigma Z_3', \end{aligned}$$

with $\Sigma_\eta = \text{Var}[\eta_t]$ and $Z_3 = \Phi_t D_N^+(I_N + C_{NN})(\varepsilon_0^0 \otimes I_N)$. For the multinormal and multivariate t-distribution cases, an explicit expression for Σ_η can be found in Hafner (2001). If the baseline is set to zero, then $Z_3 = 0$ and the variance of v_t reduces to $\Phi_t \Sigma_\eta \Phi_t'$.

Instead, one could again condition on the volatility state Σ_0 and draw δ from the distribution of $\Sigma_0^{1/2} \xi_0$. As shown in (21), $v_t(\delta)$ is in this case just a shifted version of $V_t(\xi_0)$, with the shift given by $\Phi_t \text{vech}(\Sigma_0)$. Thus, the mean will take the value of the shift and the variance will be the same as for $V_t(\xi_0)$:

$$\begin{aligned} \mathbb{E}[v_t(\Sigma_0^{1/2} \xi_0)] &= \Phi_t \text{vech}(\Sigma_0), \\ \text{Var}[v_t(\Sigma_0^{1/2} \xi_0)] &= Z_1 \Sigma_{\xi\xi'} Z_1' = \text{Var}[V_t(\xi_0)]. \end{aligned}$$

The drawback is again that a fixed baseline cannot represent the steady state of volatility, as outlined above.

D) Random shock and random baseline: The final possibility is to let both δ and ε_0^0 be random, for example drawn from the unconditional distribution of ε_t . But then the perturbed scenario $\varepsilon_0^* = \varepsilon_0^0 + \delta$ is the sum of two independent random variables stemming both from the unconditional distribution of ε_t , so that a realization of ε_0^* may be very unrealistic for the distribution of ε_t .

Concluding our comparison between conditional volatility profiles and VIRF, the concepts have some similar properties but do not convey the same information. It is only

possible to reconcile both methodologies by skipping the baseline in the condition and considering

$$E[\text{vech}(\Sigma_t) \mid \varepsilon_0^*, \mathcal{F}_{-1}] - E[\text{vech}(\Sigma_t) \mid \mathcal{F}_{-1}], \quad (24)$$

which obviously abandons a major part of the GRT definition. Now, setting $\varepsilon_0^* = \Sigma_0^{1/2} \xi_0$, one obtains immediately the equivalence of (24) with the definition of VIRF given in (9).

5 Empirical Analysis

As in Bollerslev and Engle (1993) we analyze daily data for the Deutsche Mark (DEM) and British pound sterling (GBP) exchange rates vis-à-vis the US-Dollar (USD) to illustrate VIRF and compare them with conditional volatility profiles. Our extended estimation period is December 31, 1979 to April 1, 1994, providing 3720 observations. The corresponding log return processes are stationary and show typical patterns of volatility clustering.

5.1 Univariate Analysis

To obtain a convenient residual series for the analysis of the bivariate volatility system we first infer on the mean of univariate log FX-returns conditional on the history of both processes. To take conditional heteroskedasticity for inference on the conditional mean into account we assume that error variances of single equations can conveniently be approximated by means of a GARCH(1,1) model. Hafner and Herwartz (2000) discuss inference on linear dynamics under conditional heteroskedasticity. Following their conclusions from a comparison of QML-inference and OLS-based test statistics as, for instance, the t -ratio introduced in White (1980), we infer on linear dynamics by means of QML-methods advocated in Bollerslev and Wooldridge (1992). As a main characteristic t -statistics provided for univariate processes take nonnormality of underlying innovations into account.

Defining y_{1t} and y_{2t} as the log DEM/USD returns and log GBP/USD returns, respectively, the following time series representations are estimated:

- DEM/USD:

$$\begin{aligned} y_{1t} &= \ln(\text{DEM/USD})_t - \ln(\text{DEM/USD})_{t-1}, \\ &= \underset{(-0.28)}{-2.85\text{e-}05} + \underset{(1.96)}{0.035} y_{1t-1} + \hat{\varepsilon}_{1t}, \\ \varepsilon_{1t} | \mathcal{F}_{t-1} &\sim N(0, \sigma_{1t}^2), \quad \hat{\sigma}_{1t}^2 = \underset{(2.97)}{8.13\text{e-}05} + \underset{(7.68)}{0.082} \hat{\varepsilon}_{1t-1}^2 + \underset{(75.1)}{0.906} \hat{\sigma}_{1t-1}^2. \end{aligned} \quad (25)$$

- GBP/USD:

$$y_{2t} = \ln(\text{GBP/USD})_t - \ln(\text{GBP/USD})_{t-1},$$

	ARCH-LM(1)	JB	Q(10)	Q(20)
	DEM/USD			
$\hat{\xi}_{1t} = \hat{\varepsilon}_{1t}/\hat{\sigma}_{1t}$	0.486 (0.49)	353.28 (0.00)	19.69 (0.03)	34.22 (0.03)
	GBP/USD			
$\hat{\xi}_{2t} = \hat{\varepsilon}_{2t}/\hat{\sigma}_{2t}$	0.443 (0.51)	432.49 (0.00)	13.88 (0.18)	24.92 (0.20)

Table 1: *Diagnostic results for univariate processes, ARCH-LM Test on homoskedasticity, Jarque-Bera test on normality, $Q(k)$ is the Ljung-Box statistic on zero autocorrelation up to order k . p -values are given in parentheses underneath the test-statistics.*

$$\begin{aligned}
&= \begin{matrix} 1.13\text{e-}04 + 0.062 y_{2t-1} + \hat{\varepsilon}_{2t}, \\ (1.14) \quad (3.57) \end{matrix} \tag{26} \\
\varepsilon_{2t} | \mathcal{F}_{t-1} &\sim N(0, \sigma_{2t}^2), \quad \hat{\sigma}_{2t}^2 = \begin{matrix} 8.87\text{e-}07 + 0.070 \hat{\varepsilon}_{2t-1}^2 + 0.914 \hat{\sigma}_{2t-1}^2. \\ (3.39) \quad (6.16) \quad (64.7) \end{matrix}
\end{aligned}$$

QML t -ratios are given in parentheses underneath the coefficient estimates. The simple autoregressive representations of order one are obtained for both series although an initial specification allowed for lagged (cross) dynamics up to order 5. For both estimated equations all remaining 9 explanatory variables of this general model turned out to be insignificant at the 5% significance level.

The t -ratios are quite large for the GARCH parameters and thus indicate that conditional heteroskedasticity is present in both (univariate) error processes. Diagnostic statistics computed from standardized innovations $\hat{\xi}_{it} = \hat{\varepsilon}_{it}/\hat{\sigma}_{it}$, $i = 1, 2$, are provided in Table 1. It turns out that the univariate GARCH(1,1) captures sufficiently conditional heteroskedasticity of the investigated FX-rates. Applying ARCH-LM(1) tests (Engle 1982) to $\hat{\xi}_{it}$ provides no evidence in favor of remaining ARCH-type dynamics. Both residual processes show excess kurtosis such that the normality assumption is strongly violated. Moreover, inferring on remaining linear dependencies by means of Ljung-Box statistics, $Q(10)$ and $Q(20)$, the residual process of the GBP/USD-rate ($\hat{\xi}_{2t}$) is found to be uncorrelated. Applying the latter tests to the residual process of the DEM/USD-rate ($\hat{\xi}_{1t}$), however, the white noise hypothesis is rejected at the 5% level, but not at the 1% level. Looking somewhat deeper at the estimated autocorrelation pattern we find that only two autocorrelation estimates, namely those for lag 14 and 20, exceed $2/\sqrt{T}$ in absolute value. Moreover, augmenting the empirical model (25) with further (lagged) explanatory variables provides only negligible improvements of both residual statistics, $Q(10)$ and $Q(20)$. Summarizing our results we regard the empirical models (25) and (26) to be suitable for the provision of a bivariate residual series $\varepsilon_t = (\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t})'$. Joint volatility dynamics of this

series are now considered.

5.2 Bivariate volatility dynamics

Since ARCH-LM tests applied to the residuals of the univariate time series representations do not indicate the necessity of higher order models, we choose for the BEKK model in (4) order parameters $K = p = q = 1$. As initial values for QML-estimation we use the univariate GARCH estimates. Diagonal elements of the matrices A_{11} and G_{11} are taken to be the square root of the corresponding univariate estimates, off-diagonal elements of A_{11} and G_{11} are initialized with zero. The implied unconditional covariance matrix and the sample means of variances and covariances are used to determine initial values of C_0 . The estimation of the BEKK-model is performed alternatively under two distributional assumptions for ε_t , namely

1. $\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \Sigma_t)$, and
2. $\varepsilon_t | \mathcal{F}_{t-1} \sim t(0, \Omega_t, \nu)$, $\nu > 2$, $\Omega_t = \Sigma_t \frac{\nu-2}{\nu}$.

The estimated parameter matrices with t -ratios in parentheses are given in Table 2.

Enforcing the matrices A_{11} and G_{11} to be diagonal yields under conditional normality and under a conditional $t(0, \Omega_t, \nu)$ -distribution maximum values of the likelihood function of 28583.22 and 28975.28, respectively. The estimated degrees of freedom parameter of the $t(0, \Omega_t, \nu)$ -distribution is $\hat{\nu} = 4.7$. We regard both the estimated degrees of freedom and the log-likelihood improvement achieved by the $t(0, \Omega_t, \nu)$ -model as strong evidence against the Gaussian model and in favor of leptokurtic innovations ξ_t . Introducing four additional off-diagonal coefficients increases the log-likelihood by $\Delta \ln L = 12.14$ for the Gaussian and by $\Delta \ln L = 6.25$ under leptokurtic innovations ξ_t . Using analogously to common likelihood-ratio (LR) tests a $\chi^2(4)$ -distribution to indicate significance of $2\Delta \ln L$ we obtain p-values of 7.01e-05 and 0.015 for the Gaussian and the $t(0, \Omega_t, \nu)$ -model, respectively. Even if the LR-statistic for testing the restrictions implied by the diagonal BEKK model against its unrestricted competitor may not have a $\chi^2(4)$ distribution, we regard it as a meaningful descriptive tool indicating that the diagonal BEKK model might be too restrictive for this particular example.

The estimated off-diagonal elements are negative for the A_{11} -matrix and positive for the G_{11} -matrix. For both unrestricted BEKK-models we find two of four off-diagonal elements significant at the 5% level. Interestingly, estimates $\hat{a}_{21}$, $\hat{g}_{21}$ are significant (insignificant) in the Gaussian model ($t(0, \Omega_t, \nu)$ -model). Opposite results are obtained for $\hat{a}_{12}$, $\hat{g}_{12}$. The eigenvalues of the estimate of $A_{11} \otimes A_{11} + G_{11} \otimes G_{11}$ are also given in Table 2. They indicate that both parameterizations imply covariance stationarity and high persistence, since the largest eigenvalues are only slightly smaller than one.

Model	C_0		A_{11}		G_{11}		ρ_i	$\log L$
	$\varepsilon_t \mathcal{F}_{t-1} \sim N(0, \Sigma_t),$							
DIAG	1.30e-03		0.271		0.946		.980	28583.22
	(8.19)		(13.8)		(128.3)		.974	
	8.55e-04	6.70e-04	0.281		0.949		.974	
	(5.91)	(4.78)	(10.6)		(90.3)		.968	
BEKK	1.15e-03		0.284 -0.047		0.939 0.021		.984	28595.36
	(5.78)		(6.62) (-0.47)		(86.9) (0.42)		.966	
	4.40e-04	6.97e-04	-0.063 0.285		0.024 0.945		.965	
	(1.10)	(4.69)	(-2.38) (3.80)		(1.99) (75.0)		.958	
	$\varepsilon_t \mathcal{F}_{t-1} \sim t(0, \Omega_t, \nu)$							
DIAG	1.12e-03		0.272		0.953		.990	28975.28
	(10.1)		(17.3)		(177.8)		.986	
	8.43e-04	5.07e-04	0.280		0.955		.986	
	(8.18)	(6.42)	(13.8)		(142.0)		.982	
BEKK	1.04e-03		0.271 -0.055		0.954 0.016		.990	28981.53
	(10.1)		(9.58) (-2.82)		(100.9) (2.25)		.985	
	7.03e-04	5.29e-04	-0.047 0.290		0.012 0.950		.983	
	(5.76)	(6.03)	(-1.44) (9.37)		(1.08) (87.5)		.982	

Table 2: *Estimation results for the diagonal BEKK (DIAG) and BEKK models (t -statistics in parentheses). ρ_i , $i = 1, \dots, 4$, are the eigenvalues of the matrix $A_{11} \otimes A_{11} + G_{11} \otimes G_{11}$ and $\log L$ is the value of the log-likelihood function.*

As shown in Section 2, identifiability of underlying innovations ξ_t by means of $\Sigma_t^{1/2}$ requires that these variables are not Gaussian. To infer against multivariate normality of $\hat{\xi}_t = \hat{\Sigma}_t^{-1/2} \hat{\varepsilon}_t$ Table 3 provides the multivariate Jarque-Bera (JB) statistic for all estimated BEKK-specifications. Under the null hypothesis of normality this statistic is $\chi^2(4)$ distributed. Throughout, the obtained JB-statistics are highly significant, thereby rejecting the normality hypothesis. In addition, Table 3 displays a few empirical moments of $\hat{\xi}_{it}$. Although these innovations are significantly skewed it turns out that the strong evidence against the Gaussian model can essentially be attributed to residual leptokurtosis. All estimated fourth order moments are larger than 6. Allowing for conditional leptokurtosis via the $t(0, \Omega_t, \nu)$ -model estimated fourth order moments of $\hat{\xi}_{it}$ exceed substantially the corresponding estimates obtained under normality. For instance, the fourth order moment of innovations governing the DEM/USD rate is 7.998 under conditional leptokurtosis whereas it is 6.066 under the Gaussian assumption.

Model	DEM/USD ($\bar{\xi}_1^T$)				GBP/USD ($\bar{\xi}_2^T$)				
r	1	2	3	4	1	2	3	4	JB
	$\varepsilon_t \mathcal{F}_{t-1} \sim N(0, \Sigma_t),$								
DIAG	1.4e-02	1.002	-0.251	6.144	-4.0e-03	1.004	0.315	6.232	3208.86
BEKK	1.5e-02	1.001	-0.227	6.066	-6.3e-03	1.008	0.262	6.127	2976.58
	$\varepsilon_t \mathcal{F}_{t-1} \sim t(0, \Omega_t, \nu)$								
DIAG	1.4e-02	.9912	-0.323	7.054	-3.7e-03	.9886	0.357	6.788	5315.86
BEKK	1.5e-02	1.009	-0.353	7.998	-6.2e-03	1.005	0.331	7.386	6759.85

Table 3: *Empirical moments of estimated innovations $\hat{\xi}_{it}$ and Jarque-Bera test on normality.*

5.3 VIRF for the estimated models

In the following, we illustrate VIRF for the estimated models. After considering given historical shocks in Section 5.3.1, we will look at random shocks in Sections 5.3.2 and 5.3.3.

5.3.1 Analysis of historical shocks

Before considering random shocks, let 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 European Community 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 appreciated against the USD whereas the GBP remained at about the same level. So this event had an asymmetric effect on volatility.

We will compare VIRF for these two events with conditional volatility profiles. For the former, we take the estimated residual ε_0 and the estimated volatility state Σ_0 on the day the shock occurred and construct standardized residuals ξ_0 , for which we can calculate $V_t(\xi_0)$. For the latter, we assume that the baseline is zero and the shock is given by the estimated residual, $\delta = \varepsilon_0$. Since the empirical analysis clearly preferred the unrestricted BEKK model with t-distributed error terms, we will only report the results for this specification. The impulse responses are scaled with respect to the estimated conditional volatilities at the time the shock occurred. 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 $\varepsilon_0 = (0.0221, 0.0321)'$ and the estimated volatility state (expressed in E-05) $\text{vech}(\Sigma_0) = (1.196, 0.962, 0.994)'$. The left panel of Figure 1 shows 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 impulse responses start from a high positive level, declining slowly, with the conditional volatility profiles slightly above the VIRF. The reason for the positive difference between the two impulse responses is due to equation (21). Apart from this shift, there is no major qualitative difference between the two notions in this case.

On August 2, 1993, the estimated residual vector was $\varepsilon_0 = (-0.014807, -0.003195)'$ and the estimated volatility state (expressed in E-05) $\text{vech}(\Sigma_0) = (0.3886, 0.3341, 0.4852)'$. 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. The GBP/USD variance, on the other hand, sees the corresponding element of $V_t(\xi_0)$ starting at a negative level, increasing with a peak at around $t = 150$ and then slowly decreasing to zero. The conditional volatility profile is again shifted upwards, but such that it takes positive values throughout and with an earlier peak at around $t = 70$. Note that the initial gap between the two functions is more than 6% of the current conditional GBP/USD variance. Investors using the conditional volatility profile may be misled in believing that the effect of the shock for GBP/USD volatility is positive and increasing over the first 70 days. As shown in Section 4.2, the positivity is an artefact of setting the baseline to zero.

Similarly, the VIRF of the covariance suggest that the shock is negative, slowly dampening to zero, whereas the conditional volatility profile shows a positive effect for the covariance. This might have implications for international portfolio management when the covariance between exchange rates is used as a component of risk measures.

5.3.2 Dispersion of (random) VIRF and conditional moment profiles

The impact of a random shock can be measured using some notion for dispersion of the distribution of VIRF. If the impact of a shock decays, then the dispersion of its distribution should decrease over time. A primary measure for dispersion is the variance, whose theoretical value is known and given in (17). Figure 2 plots the standard deviations of the individual components of VIRF, which are given by the square root of the diagonal elements of the matrix (17). We assume that at time zero when the shock occurs, the process is in the steady state of volatility, i.e., $\Sigma_0 = \Sigma$.

In order to compare VIRF with conditional volatility profiles, we selected of the many possibilities sketched in Section 4.2.2 the one that is typical for the original GRT approach, i.e., randomizing over the baseline and fixing the shock. On the other hand, we do not, as GRT suggest, draw the baseline from the unconditional distribution but rather from

the conditional distribution with $\Sigma_0 = \Sigma$. This is because VIRF are also conditioned on the volatility state and we want to make the underlying conditions of the two concepts as close as possible. As shown in Section 4.2.2, the variances of both procedures are equal and given by (23), but one would expect fatter tails using the unconditional distribution.

We will focus on the following three shock scenarios for GRT: One is $\delta_1 = (\sqrt{\Sigma_{11}}, 0)'$, i.e., a shock of one unconditional standard deviation in DEM/USD, zero in GBP/USD. The second is the reverse situation, $\delta_2 = (0, \sqrt{\Sigma_{22}})'$, and the third is $\delta_3 = (\sqrt{\Sigma_{11}}, \sqrt{\Sigma_{22}})'$.

It is obvious from the graphs that shocks persist very long in both GBP/USD and DEM/USD volatilities and in the covariance between GBP/USD and DEM/USD, which corresponds to the large eigenvalues of the matrix $A_{11} \otimes A_{11} + G_{11} \otimes G_{11}$ reported in Table 2. Even after 100 days the dispersion of the VIRF is still positive.

The conditional volatility profile of DEM/USD has a higher dispersion if the shock is in DEM/USD (δ_1) than when it is in GBP/USD (δ_2). The same holds for the reverse situation. Interestingly, a one-component shock has a higher impact on the same component than a shock in both components (δ_3). For a multivariate t-distribution this holds throughout, but in the Gaussian case only initially and the persistence of a joint shock is higher than that of a one-component shock. The reason for the higher initial impact of one-component shocks is that off-diagonal elements of the estimated parameter matrix A_{11} are negative.

The conditional covariance shows the strongest effect for δ_3 , which can be seen as a positive shock to the covariance. The one-component shocks δ_1 and δ_2 are negative shocks to the covariance and have a somewhat smaller impact.

Evaluating the dynamic relationships of the BEKK model under a conditional t-distribution, we see that all impulse response dispersions are somewhat higher than under conditional normality. The dispersion of VIRF in the Gaussian case is roughly speaking between the dispersions of the two one-component conditional volatility profiles (δ_1 and δ_2). On the contrary, under a multivariate t-distribution the dispersion of the first and third element of VIRF is higher than all conditional volatility profiles considered.

5.3.3 Simulated quantiles

As outlined in Section 4.1, the distribution of the VIRF will in general be asymmetric and far from being Gaussian. To shed light on the distribution, we suggest to look at quantiles of simulated distributions. To do this, we use 20,000 realizations of ξ_0 from the respective distribution (Gaussian vs. $t(0, \frac{\nu}{\nu-2}I_N, \nu)$ with $\nu = 4.7$). The VIRF can then be calculated according to (11), obtaining 20,000 realizations of VIRF. For a given time horizon, this yields an empirical distribution of impulse response functions, for which we report the 5, 50 and 95 percent quantiles. These are plotted in Figure 3 for the alternative models. Firstly we compare dynamic implications of the unrestricted vs the diagonal BEKK model under conditional normality (left hand side panels of Figure 3). Secondly,

for the unrestricted BEKK model we contrast the effect of alternative conditional distributions, Gaussian vs. $t(0, \frac{\nu}{\nu-2}I_N, \nu)$ (medium column of Figure 3). Thirdly, the preferred empirical model, the unrestricted BEKK model under conditional leptokurtosis, is used to compare VIRF vs. conditional volatility profiles with common shock, δ_3 (right hand side panels of Figure 3). We obtain that the quantiles of the diagonal BEKK specification are higher (in absolute value) initially but show a lower persistence than the BEKK specification. Furthermore, the 5 and 95 percent quantiles under a multivariate t-distribution are slightly higher (in absolute value) and have a slightly higher persistence than under a multinormal distribution. Finally, conditional volatility profiles have higher (in absolute value) 5 and 95 percent quantiles than VIRF. Moreover, the distribution of conditional volatility profiles is more symmetric, since the median is roughly between the 5 and 95 percent quantile, whereas the median is much closer to the 5 percent quantile for VIRF.

To compare our results with previous work, we also computed the VIRF for the estimation results reported in Bollerslev and Engle (1993) for a vec model and the shorter estimation period until 1985. Their results imply a relatively fast dissipation of shocks in the GBP/USD, as was visible in the corresponding plots (not shown). This may be due to the lack of events such as the October 1987 crash in their sample.

6 Conclusions and Outlook

The paper introduces volatility impulse response functions (VIRF) for multivariate time series exhibiting conditional heteroskedasticity. Unlike other recent approaches, we define news to appear in the independent innovation rather than in the conditionally heteroskedastic error. For the general vec representation of multivariate GARCH models we provide the analytic expressions for VIRF.

The advocated methodology is applied to a bivariate system consisting of the DEM/USD and the GBP/USD FX-rate. We find an unrestricted BEKK model implemented under the assumption of conditional t-distributed error terms to outperform both the diagonal BEKK model and its counterpart under conditional normality. The estimated VIRF illustrate that allowing for causality in variance between the DEM/USD and GBP/USD volatilities dampens the initial dispersion of the impulse responses but increases the shock persistence. Under the t-model, particular shocks are found to die out at a slightly slower rate compared to the Gaussian case. A comparison of VIRF and conditional volatility profiles is complicated due to the difficulty of selecting convenient shock scenarios. If shocks are assumed to hit both variables, we find the distribution of conditional volatility profiles to be quite symmetric whereas VIRF distributions are highly skewed. We discuss two examples of historical shocks and their impacts on volatility. We find considerably different interpretations using VIRF and conditional volatility profiles, respectively, for shocks affecting the volatilities in an asymmetric way, that is, a shock that perturbs one

series but not the other.

In the empirical analysis of stock markets, the leverage effect is well documented, see e.g. Engle and Ng (1993). That is, the impact of negative shocks on volatility may be different from the impact of positive shocks. The standard multivariate GARCH model as treated in this paper is not able to account for asymmetric responses to shocks. However, the methodology proposed in this paper is easily extended to multivariate threshold GARCH models. Extensions to multivariate exponential GARCH models will inevitably require assumptions about the innovation distribution, if the analytic form of the VIRF is of interest. This will give many directions for future research.

Throughout our discussion of empirical VIRF, we assumed the estimated models to be equal to the true data generating process. As another direction of future research, one may investigate the distributional properties of VIRF estimates. To this end, bootstrap methods may be considered.

Finally, it should be mentioned that time-varying volatility of multivariate financial time series may also be modeled by stochastic volatility, see e.g. Harvey, Ruiz and Shephard (1994) and Danielsson (1998). Stochastic volatility models may be more parsimonious and are sometimes found to provide better fits than GARCH models, but estimation is typically less straightforward than for GARCH-type models. An extension of VIRF to stochastic volatility models may enhance the comprehension of the involved dynamics in comparison with ARCH-type models. This is also a potential topic for future research.

References

- Berndt, E.K., Hall, B.H., Hall, R.E., Hausman, J.A. (1974), “Estimation and Inference in nonlinear structural models”, *Annals of Economic and Social Measurement*, 3/4, 653-665.
- Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics*, 31, 307–327.
- Bollerslev, T., (1987), “A Conditional Heteroscedastic Time Series Model for Speculative Prices and Rates of Return”, *Review of Economics and Statistics*, 69, 542-547.
- Bollerslev, T., Engle, R.F. (1993), “Common Persistence in Conditional Variances”, *Econometrica*, 61, 167–186.
- Bollerslev, T., Engle, R.F., Wooldridge, J.M. (1988), “A Capital Asset Pricing Model with Time Varying Covariances”, *Journal of Political Economy*, 96, 116–131.
- Bollerslev, T., Wooldridge, J.M. (1992), “Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances”, *Econometric Reviews*, 11, 143–172.

- Comte, F., Lieberman, O. (2000), “Asymptotic Theory for Multivariate GARCH Processes”, *Manuscript, Universities Paris 6 and Paris 7*.
- Danielsson, J., (1998), “Multivariate Stochastic Volatility Models: Estimation and a Comparison with VGARCH Models”, *Journal of Empirical Finance*, 5, 155–174.
- Drost, F., Nijman, T. (1993), “Temporal Aggregation of GARCH Processes”, *Econometrica*, 61, 909–927.
- Engle, R. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation”, *Econometrica*, 50, 987–1008.
- Engle, R.F., Kroner, K.F. (1995), “Multivariate Simultaneous Generalized ARCH”, *Econometric Theory*, 11, 122–150.
- Engle, R.F., Ng, V.K. (1993), “Measuring and Testing the Impact of News on Volatility”, *The Journal of Finance*, 48, 1749–1778.
- Gallant, A.R., Rossi, P.E., Tauchen, G. (1993), “Nonlinear Dynamic Structures”, *Econometrica*, 61, 871–907.
- Hafner, C.M. (2001), “Fourth Moments of Multivariate GARCH Processes”, CORE discussion paper, Université catholique de Louvain, Louvain-la-Neuve, Belgium.
- Hafner, C.M., Herwartz, H. (2000), “Testing for Linear Autoregressive Dynamics Under Heteroskedasticity”, *The Econometrics Journal*, 3, 177–197.
- Harvey, A.C., Ruiz, E., Shephard, N. (1994), “Multivariate Stochastic Variance Models”, *Review of Economic Studies*, 61, 247–264.
- Jeantheau, T. (1998) “Strong Consistency of Estimators for Multivariate ARCH Models”, *Econometric Theory*, 14, 70–86.
- Johnson, N.L., Kotz, S. (1972), “Distributions in Statistics: Continuous Multivariate Distributions”, Wiley, New York
- Kendall, M., Stuart, A., and Ord, J.K. (1987), “Kendall’s Advanced Theory of Statistics”, 5th ed., Charles Griffin and Co., London.
- Koop, G., Pesaran, M.H., Potter, S.M. (1996), “Impulse Response Analysis in Nonlinear Multivariate Models”, *Journal of Econometrics*, 74, 119–147.
- Lancaster, H. O. (1954), “Traces and cumulants of quadratic forms in normal variables”, *Journal of the Royal Statistical Society, Series B*, 16, 247–254.

- Lin, W.-L. (1997), “Impulse Response Function for Conditional Volatility in GARCH Models”, *Journal of Business & Economic Statistics*, 15, 15–25.
- Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin.
- Magnus, J.R. (1988), *Linear Structures*, Charles Griffin and Co., London.
- Rose, A.K., Svensson, L.E.O. (1994), “European Exchange Rate Credibility Before the Fall”, *European Economic Review*, 38, 1185–1216.
- Sims, C. (1980), “Macroeconomics and Reality”, *Econometrica*, 48, 1–48.
- White, H. (1980), “A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”, *Econometrica* 48, 817–838.

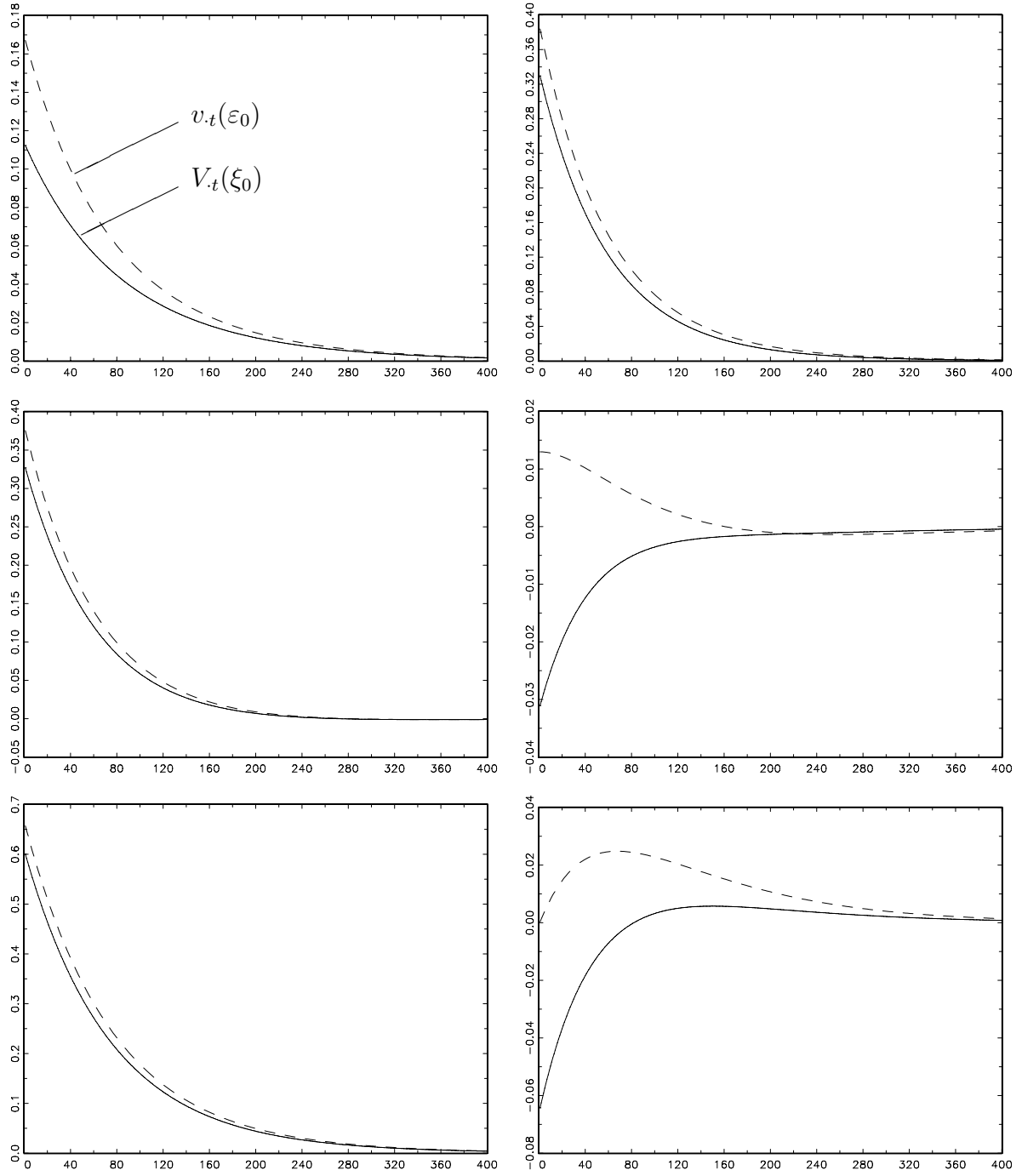


Figure 1: *Left panel: shock on Sept 16, 1992. Right panel: Shock on Aug 2nd, 1993. First row: DEM/USD variance, second row: covariance between DEM/USD and GBP/USD, third row: GBP/USD variance. $V_t(\xi_0)$ and $v_t(\varepsilon_0)$ are typical elements of VIRF and conditional volatility profiles, respectively.*

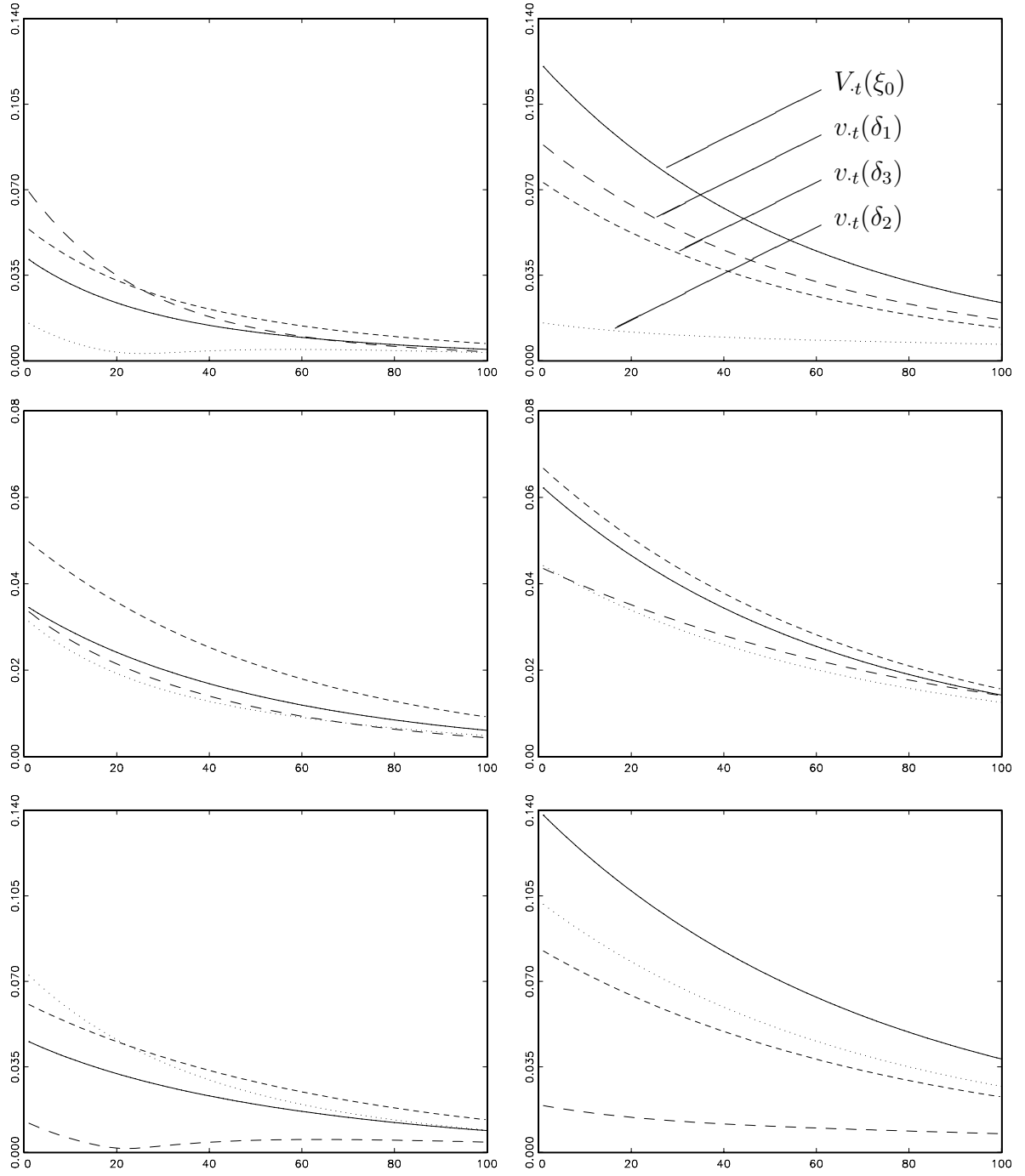


Figure 2: *Standard deviations of VIRF ($V_t(\xi_0)$) and conditional volatility profiles $v_t(\delta_i)$, $\delta_1 = (\sqrt{\Sigma_{11}}, 0)'$, $\delta_2 = (0, \sqrt{\Sigma_{22}})'$, $\delta_3 = (\sqrt{\Sigma_{11}}, \sqrt{\Sigma_{22}})'$, implied by the BEKK Model estimated under standard Gaussian (left hand side panels) and t -distributed (right hand side panels) innovations ξ_t . First row: DEM/USD variance, second row: covariance between DEM/USD and GBP/USD, third row: GBP/USD variance.*

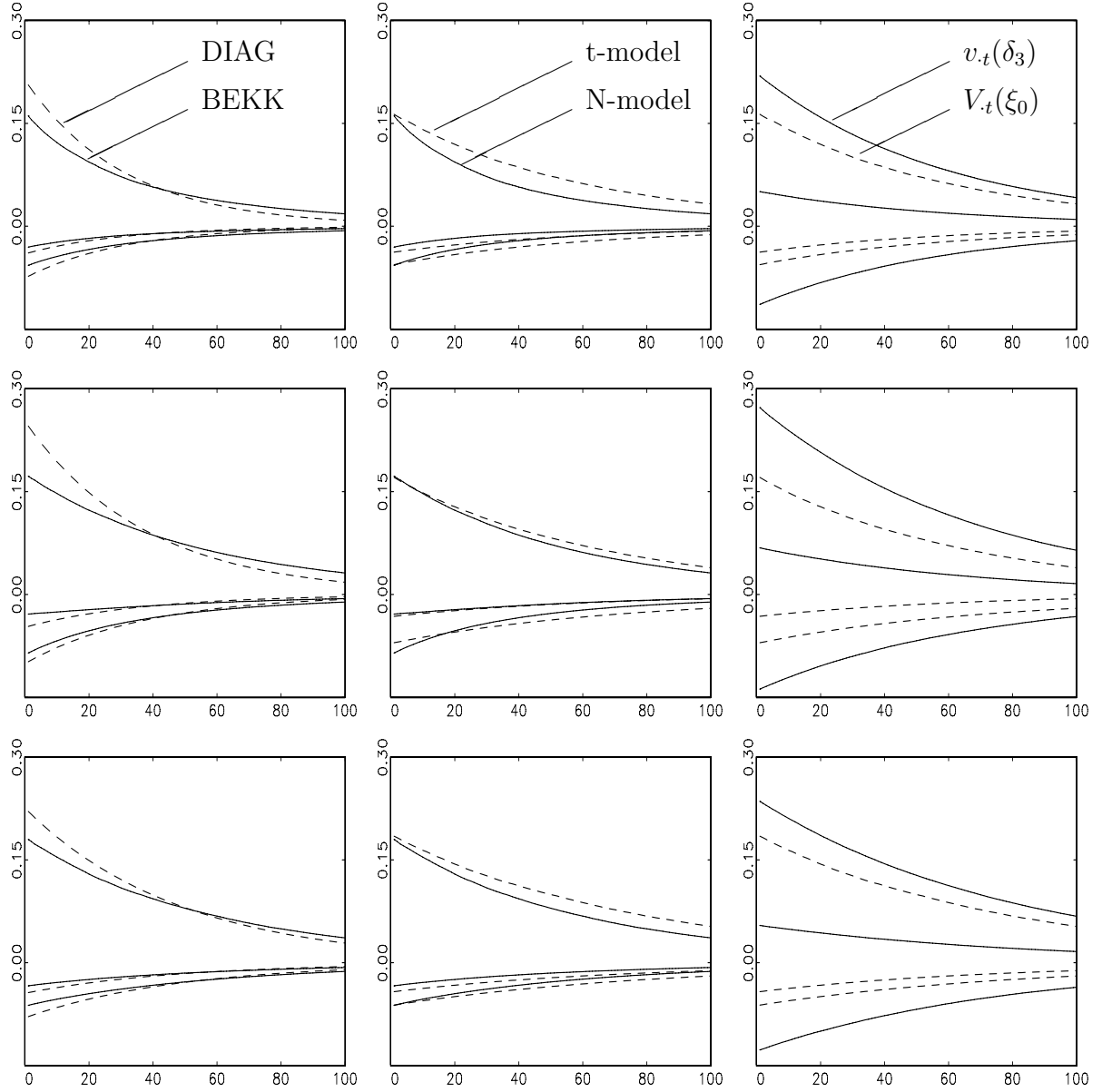


Figure 3: *Simulated 5, 50, and 95 percent quantiles of VIRF and conditional volatility profiles. Left hand side panels: Quantiles of $V_t(\xi_0)$ obtained from the unrestricted BEKK vs. the diagonal BEKK under normality. Medium panels: Quantiles of $V_t(\xi_0)$ obtained from the unrestricted BEKK under normality vs. t -distributed innovations. Right hand side panels: Quantiles of $V_t(\xi_0)$ vs. $v_t(\delta_3)$, $\delta_3 = (\sqrt{\Sigma_{11}}, \sqrt{\Sigma_{22}})'$ implied by the unrestricted BEKK with t -distributed innovations. First row: DEM/USD variance, second row: covariance between DEM/USD and GBP/USD, third row: GBP/USD variance.*