

Analytical quasi maximum likelihood inference in multivariate volatility models

Christian M. Hafner¹ Helmut Herwartz²

Econometric Institute Report EI 2003–21

July 2003

Abstract

Quasi maximum likelihood estimation and inference in multivariate volatility models remains a challenging computational task if, for example, the dimension is high. One of the reasons is that typically numerical procedures are used to compute the score and the Hessian, and often they are numerically unstable. We provide analytical formulae for the score and the Hessian and show in a simulation study that they clearly outperform numerical methods. As an example, we use the popular BEKK-GARCH model, for which we derive first and second order derivatives.

Keywords: Multivariate GARCH models, quasi maximum likelihood

JEL Classification: C14, C22.

¹Econometric Institute, Erasmus University Rotterdam, P.O.B. 1738, 3000 DR Rotterdam, Netherlands, chafner@few.eur.nl, corresponding author

²Institut für Statistik und Ökonometrie, Christian Albrechts Universität zu Kiel, Ohlshausenstr. 40-60, D-24118 Kiel, Germany, Herwartz@stat-econ.uni-kiel.de

1 Introduction

Over recent years, multivariate volatility models have become increasingly popular in research and practice. One of the reasons is certainly the improving computing power of modern computers, but also recent research on models that are possible to estimate even in high dimensions. A popular example of multivariate volatility models is the GARCH model class, see Bauwens et al. (2003) for a review. However, estimation and inference remain a difficult problem, in particular in high dimensions. One of the difficulties stems from the fact that most software packages available rely on numerical derivatives to compute the score and the Hessian of the likelihood function. This is often found to be numerically unstable, as noted by Lucchetti (2002), who provides analytical results for the scores of a particular GARCH model. It is known that numerical derivatives become even more unstable when they are used to compute the Hessian.

Maximum likelihood estimation relies on an assumption about the innovation distribution. Empirically one has often found that standardized residuals of estimated volatility models were still fat-tailed, so the assumption of Gaussian innovations is not innocuous and loses efficiency. Fiorentini et al. (2003) provide a general framework for maximum likelihood estimation using the Student-t distribution. The drawback of this approach is that, if the assumption is wrong, then in general the ML estimates are not even consistent. On the other hand, using a Gaussian likelihood, also known as quasi maximum likelihood (QML), retains consistency under misspecification. In practice, without prior information on the innovation distribution it may therefore be preferable to use QML.

The purpose of this paper is to provide an analytic framework to implement QML inference in multivariate volatility models. QML estimates are known to be consistent and asymptotically normal, under regularity conditions, in models with conditional heteroskedasticity, see Bollerslev and Wooldridge (1992). Comte and Lieberman (2003) provide a theoretical framework for multivariate

GARCH models in a general specification.¹ We provide analytic formulae for the score and the Hessian of a general multivariate volatility model. These depend in general on the first and second derivatives of the volatility matrix H_t with respect to its parameters, for which we provide results using the popular BEKK model class.

We present two methods to estimate the expectation of the Hessian; one method involves computation of second derivatives of the volatility matrix, the other does not. We compare their empirical performance in a simulation experiment, where we also include numerical derivatives as a third method. The striking result is that numerical derivatives are clearly outperformed by any of the analytic methods. The analytic method using second derivatives of H_t is outperformed by the one that does not in small and medium samples, but seems to be advantageous in large samples.

The following sections are organized as follows. Section 2 describes a general framework for QML estimation of multivariate volatility models, leaving H_t unspecified. Section 3 provides results for a popular example for H_t , the so-called BEKK-GARCH model. Section 4 presents the Monte Carlo experiment, and Section 5 concludes.

2 QML inference in multivariate volatility models

Let $H_t(\theta)$ be a positive definite ($N \times N$) conditional covariance matrix of some ($N \times 1$) error vector ε_t , parameterized by the vector θ . Denoting the information set available at time t by \mathcal{F}_t , and writing $E_t[\cdot] = E[\cdot | \mathcal{F}_t]$ for the conditional expectation operator, the model can be written as

$$\begin{aligned} E_{t-1}[\varepsilon_t] &= 0 \\ E_{t-1}[\varepsilon_t \varepsilon_t'] &= H_t(\theta) \end{aligned}$$

¹Consistency was already shown by Jeantheau (1998), but for a less general specification. For the same restricted specification, Ling and McAleer (2003) provide a theoretical framework for QML estimation and inference.

The conditional covariance matrix $H_t(\theta)$ can be explained by some multivariate GARCH model or any other multivariate volatility model as long as $H_t(\theta)$ is measurable with respect to \mathcal{F}_{t-1} , is continuous in θ and is twice continuously differentiable. We will give a popular example of $H_t(\theta)$ in the next section.

Suppose that there is an underlying data generating process characterized by the unknown parameter vector θ_0 which one wants to estimate using a given sample of T observations. The quasi maximum likelihood (QML) approach estimates θ_0 by maximizing the Gaussian log likelihood function $L(\theta) = \sum_{t=1}^T l_t(\theta)$ with

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

Under conditions listed by Comte and Lieberman (2003), the QML estimates $\hat{\theta}$ are consistent and asymptotically normally distributed, even in the case of non-normally distributed innovations. The asymptotic distribution is given by

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \mathcal{J}^{-1} \mathcal{I} \mathcal{J}^{-1}),$$

where

$$\mathcal{I} = \text{E} \left[\frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta'} \Big|_{\theta_0} \right], \quad \mathcal{J} = -\text{E} \left[\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta_0} \right]$$

and where the expectation is taken with respect to the true process. The matrix \mathcal{I} is the expectation of the outer product of the score vector evaluated at the true parameter vector θ_0 and is often called the information matrix, whereas \mathcal{J} is the negative expectation of the Hessian evaluated at θ_0 . If the error process ε_t is conditionally Gaussian, then $\mathcal{I} = \mathcal{J}$ and the asymptotic covariance matrix reduces to \mathcal{I}^{-1} , the Cramer-Rao lower bound.

For inference on the estimates $\hat{\theta}$ one therefore needs to calculate the score vector and the Hessian. These are given by (see the proof of Lemma 1 of Comte and Lieberman, 2003)

$$\frac{\partial l_t(\theta)}{\partial \theta_i} = \text{Tr} \left[\dot{H}_{t,i} H_t^{-1} - \varepsilon_t \varepsilon_t' H_t^{-1} \dot{H}_{t,i} H_t^{-1} \right]$$

and

$$\begin{aligned} \frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} = & \text{Tr} \left[\ddot{H}_{t,i,j} H_t^{-1} - \varepsilon_t \varepsilon_t' H_t^{-1} \ddot{H}_{t,i,j} H_t^{-1} - \dot{H}_{t,i} H_t^{-1} \dot{H}_{t,j} H_t^{-1} \right. \\ & \left. + \varepsilon_t \varepsilon_t' H_t^{-1} \dot{H}_{t,j} H_t^{-1} \dot{H}_{t,i} H_t^{-1} + \varepsilon_t \varepsilon_t' H_t^{-1} \dot{H}_{t,i} H_t^{-1} \dot{H}_{t,j} H_t^{-1} \right] \end{aligned} \quad (2)$$

where we use the notation

$$\dot{H}_{t,i} = \frac{\partial H_t}{\partial \theta_i}, \quad \ddot{H}_{t,i,j} = \frac{\partial^2 H_t}{\partial \theta_i \partial \theta_j}$$

and where all H_t , $\dot{H}_{t,i}$ and $\ddot{H}_{t,i,j}$ are evaluated at θ .

Note that by the consistency of the QML estimate $\hat{\theta}$, the matrices \mathcal{I} and \mathcal{J} can be consistently estimated by

$$\hat{\mathcal{I}} = \frac{1}{T} \sum_{t=1}^T \frac{\partial l_t(\theta)}{\partial \theta} \frac{\partial l_t(\theta)}{\partial \theta'} \Bigg|_{\hat{\theta}} \quad (3)$$

and

$$\hat{\mathcal{J}} = -\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \Bigg|_{\hat{\theta}} \quad (4)$$

From the expression for the second derivatives of the likelihood in (2), the latter involves the second derivatives of H_t with respect to θ . However, by definition of H_t , $H_t(\theta_0) = \text{E}_{t-1}(\varepsilon_t \varepsilon_t')$, so that the first two terms of (2) just cancel under the conditional expectation operator, and

$$M_{t,ij}(\theta_0) := \text{E}_{t-1} \left[\frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} \Bigg|_{\theta_0} \right] = \text{Tr} \left[\dot{H}_{t,i} H_t^{-1} \dot{H}_{t,j} H_t^{-1} \right]$$

where H_t , $\dot{H}_{t,i}$ and $\dot{H}_{t,j}$ are evaluated at θ_0 . By the law of iterated expectations we have $\mathcal{J} = -\text{E}[M_t(\theta_0)]$ so that a computationally simpler estimate for \mathcal{J} is given by replacing the unknown true parameter vector θ_0 in $M_t(\theta_0)$ by the QML estimator, i.e.

$$\tilde{\mathcal{J}} = -\frac{1}{T} \sum_{t=1}^T M_t(\hat{\theta})$$

The estimator $\tilde{\mathcal{J}}$ avoids the computation of second derivatives of H_t and is therefore easier to implement than $\hat{\mathcal{J}}$. Both estimators are asymptotically

equivalent, so they are expected to perform equally well in large samples. In small and medium samples, on the other hand, it is not clear a priori if the computational feasibility of $\tilde{\mathcal{J}}$ goes without cost when compared with $\hat{\mathcal{J}}$. Note that in finite samples $\hat{\mathcal{J}}$ has a higher variance than $\tilde{\mathcal{J}}$ due to the additional noise terms in (2). Both estimators being unbiased, $\tilde{\mathcal{J}}$ should therefore be preferred in estimating the asymptotic distribution. However, there may be situations where $\hat{\mathcal{J}}$ is preferable in approximating the finite sample distribution. We will investigate the empirical performance of both estimators in our simulation study.

Of course, the derivatives of H_t depend on the particular volatility model used, and we will give an example in the next section. One of the objectives of our paper is to motivate the use of analytic first and second derivatives of H_t instead of numerical ones, which can be quite unstable. We will demonstrate this also in our simulation study.

3 QML inference in the BEKK-GARCH model

In this section we discuss a popular example of a model for H_t , the conditional covariance matrix, the so-called BEKK model of Engle and Kroner (1995). It has the attractive feature that H_t is positive definite by construction. Many other multivariate GARCH variants are special cases of the BEKK specification, for example the factor model of Engle, Ng and Rothschild (1990), the orthogonal GARCH model of Alexander (2001) and the GO-GARCH model of van der Weide (2002). For details on these models we refer to Bauwens et al. (2003).

In its general form, the BEKK(p, q, K) model can be written as

$$H_t = CC' + \sum_{k=1}^K \sum_{i=1}^q A'_{ki} \varepsilon_{t-i} \varepsilon'_{t-i} A_{ki} + \sum_{k=1}^K \sum_{i=1}^p B'_{ki} H_{t-i} B_{ki}, \quad (5)$$

where C is a lower triangular matrix and A_{ki} and B_{ki} are $N \times N$ parameter matrices. For illustrative purposes, we will only consider the case $p = q = K = 1$, which is also by far the most popular model order. Thus, the model

simplifies to

$$H_t = CC' + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'H_{t-1}B. \quad (6)$$

For this model, the parameter vector is given by $\theta = (\text{vech}(C)', \text{vec}(A)', \text{vec}(B)')$. In the bivariate case, this amounts to 11 parameters.

In the following we calculate the derivatives $\dot{H}_{t,i}$ and $\ddot{H}_{t,i,j}$ that were required in the previous section. Rather than deriving with respect to specific components of θ , it is more convenient to calculate the derivative of the vectorized H_t with respect to the vector θ . Of course, the former expression can easily be obtained by transforming the latter expression appropriately.

For the first and second derivatives of H_t with respect to θ we use the following notation:

$$H_{Xt} = \frac{\partial \text{vec}(H_t)}{\partial \text{vec}(X)'}, \quad (N^2 \times N^2)$$

$$H_{XYt} = \frac{\partial \text{vec}(H_{Xt})}{\partial \text{vec}(Y)'}, \quad (N^4 \times N^2)$$

If X is a lower triangular matrix, we write

$$H_{Xt}^\Delta = \frac{\partial \text{vec}(H_t)}{\partial \text{vech}(X)'}, \quad (N^2 \times N^*)$$

where $N^* = N(N+1)/2$. If X is lower triangular, but not Y ,

$$H_{XYt}^{\Delta\cdot} = \frac{\partial \text{vec}(H_{Xt}^\Delta)}{\partial \text{vec}(Y)'}, \quad (N^2 N^* \times N^2)$$

If Y is lower triangular, but not X ,

$$H_{XYt}^{\cdot\Delta} = \frac{\partial \text{vec}(H_{Xt})}{\partial \text{vech}(Y)'}, \quad (N^4 \times N^*)$$

and if both X and Y are lower triangular,

$$H_{XYt}^{\Delta\Delta} = \frac{\partial \text{vec}(H_{Xt}^\Delta)}{\partial \text{vech}(Y)'}, \quad (N^2 N^* \times N^*)$$

The dimensions are given in parentheses, where $N^* = N(N+1)/2$.

Let us begin by considering the first derivatives of H_t . All results follow by applying standard rules for matrix calculus, see e.g. Lütkepohl (1996). Deriving with respect to the parameters in A, B and C we get

$$H_{At} = 2D_N D_N^+ (I_N \otimes A' \varepsilon_{t-1} \varepsilon'_{t-1}) + (B \otimes B)' H_{At-1} \quad (7)$$

$$H_{Bt} = 2D_N D_N^+ (I_N \otimes B' H_{t-1}) + (B \otimes B)' H_{Bt-1} \quad (8)$$

$$H_{Ct}^\Delta = 2D_N D_N^+ (C \otimes I_N) L_N + (B \otimes B)' H_{Ct-1}^\Delta \quad (9)$$

where (7) and (8) are matrices of dimension $(N^2 \times N^2)$ and (9) is a $(N^2 \times N^*)$ matrix. The matrix D_N is the $(N^2 \times N^*)$ duplication matrix defined by the property $D_N \text{vech}(A) = \text{vec}(A)$ for any $(N \times N)$ matrix A , and D_N^+ is its generalized inverse, i.e., $D_N^+ = (D_N' D_N)^{-1} D_N'$. These expressions can now easily be transformed into the matrices $\dot{H}_{t,i}$.

Turning to the second derivatives, we first define the following $(N^4 \times N^4)$ matrices.

$$\begin{aligned} C_1 &= I_N \otimes K_{NN} \otimes I_N \\ C_2 &= 2(I_{N^2} \otimes D_N D_N^+) \\ C_3 &= C_2 C_1, \end{aligned}$$

where K_{NN} is the $(N^2 \times N^2)$ commutation matrix, see e.g. Lütkepohl (1996). We then obtain

$$\begin{aligned} H_{AAAt} &= C_3 [\text{vec}(I_N) \otimes (\varepsilon_{t-1} \varepsilon'_{t-1} \otimes I_N) K_{NN}] + [I_{N^2} \otimes (B \otimes B)'] H_{AAAt-1} \\ H_{ABAt} &= (H'_{At-1} \otimes I_{N^2}) C_1 [K_{NN} \otimes \text{vec}(B') + \text{vec}(B') \otimes K_{NN}] \\ &\quad + [I_{N^2} \otimes (B \otimes B)'] H_{ABAt-1} \\ H_{ACAt}^\Delta &= 0 \end{aligned}$$

For example, the result for H_{ABAt} can be derived from Lütkepohl (1996, 10.5.2 (1)e). Furthermore,

$$\begin{aligned} H_{BAAt} &= C_3 [\text{vec}(I_N) \otimes (I_N \otimes B') H_{At-1}] + [I_{N^2} \otimes (B \otimes B)'] H_{BAAt-1} \\ H_{BBAt} &= C_3 [\text{vec}(I_N) \otimes \{(H_{t-1} \otimes I_N) K_{NN} + (I_N \otimes B') H_{Bt-1}\}] \end{aligned}$$

$$\begin{aligned}
& + (H'_{Bt-1} \otimes I_{N^2})C_1[K_{NN} \otimes \text{vec}(B') + \text{vec}(B') \otimes K_{NN}] \\
& + [I_{N^2} \otimes (B \otimes B)']H_{BBt-1} \\
H_{BCt}^{\Delta} & = C_3[\text{vec}(I_N) \otimes \{(I_N \otimes B')H_{Ct-1}\}] + [I_{N^2} \otimes (B \otimes B)']H_{BCt-1}^{\Delta}
\end{aligned}$$

and

$$\begin{aligned}
H_{CAt}^{\Delta} & = 0 \\
H_{CBt}^{\Delta} & = (H'_{Ct-1} \otimes I_{N^2})C_1[K_{NN} \otimes \text{vec}(B') + \text{vec}(B') \otimes K_{NN}] \\
& + [I_{N^*} \otimes (B \otimes B)']H_{CBt-1}^{\Delta} \\
H_{CCt}^{\Delta\Delta} & = 2(L_N \otimes D_N D_N^+)C_1[I_{N^2} \otimes \text{vec}(I_N)]L'_N + [I_{N^*} \otimes (B \otimes B)']H_{CCt-1}^{\Delta\Delta}.
\end{aligned}$$

Again, these expressions can easily be transformed into the matrices $\ddot{H}_{t,i,j}$. Note that there are explicit formulae to compute the matrices D_N and K_{NN} so that even in high dimensions the above expressions remain computationally feasible. In any case, using these analytic expressions should be preferred to using numerical derivatives. This will become obvious in the next section, where we present a simulation experiment.

4 Monte Carlo Analysis

To illustrate the empirical properties of competing devices to evaluate the asymptotic covariance matrix of QML-estimators we simulate bivariate GARCH-processes of the BEKK-form ($K = p = q = 1$) according to the following choice of parameter matrices:

$$C = \begin{pmatrix} 1.10 & 0.30 \\ 0 & 0.90 \end{pmatrix}, \quad A = \begin{pmatrix} 0.25 & 0.05 \\ -0.05 & 0.25 \end{pmatrix}, \quad B = \begin{pmatrix} 0.9 & -0.05 \\ 0.05 & 0.9 \end{pmatrix}.$$

GARCH innovations $\xi_t = H^{-1/2}\varepsilon_t$ are drawn alternatively from a bivariate Gaussian distribution or as standardized and independent innovations from a t -distribution with 8 degrees of freedom.

Numerical derivatives are computed using the GAUSS procedures `gradp` and `hessp` (GAUSS 3.2.29). To indicate numerical evaluation we will refer to

the obtained estimates as $\bar{\mathcal{I}}$ and $\bar{\mathcal{J}}$. To summarize, five alternative devices are distinguished to estimate the covariance matrix of the estimated parameter vector:

1. $\widehat{\mathcal{I}}^{-1}$: Analytical estimation of the information matrix
2. $\widehat{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widehat{\mathcal{J}}^{-1}$: Analytical estimation of the sandwich matrix involving second derivatives of H_t
3. $\widetilde{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widetilde{\mathcal{J}}^{-1}$: Analytical estimation of the sandwich matrix without second derivatives of H_t
4. $\bar{\mathcal{I}}^{-1}$: Numerical estimation of the information matrix
5. $\bar{\mathcal{J}}^{-1}\bar{\mathcal{I}}\bar{\mathcal{J}}^{-1}$: Numerical estimation of the sandwich matrix

Two null hypotheses are tested. The first joint hypothesis concentrates on cross sectional dynamics, $H_{01} : a_{21} = 0.05, a_{12} = -0.05, b_{21} = -0.05, b_{12} = 0.05$. The second null hypothesis involves only diagonal elements of the ARCH parameter matrix, i.e. $H_{02} : a_{11} = 0.25, a_{22} = 0.25$. Since both null hypotheses correspond to the true data generating process the empirical rejection frequency should, under Gaussian innovations and when increasing the sample size, approach the nominal level of the test, which is 5% throughout. In presence of leptokurtic innovations, however, only the QML covariance evaluation can be expected to deliver unbiased empirical size estimates. In this case it might be interesting to get some evaluation of the magnitude of the empirical size distortions involved when evaluating the covariance matrix as $\widehat{\mathcal{I}}^{-1}$.

We simulate 2000 replications of the multivariate GARCH processes with alternative sample sizes $T = 1000, 2000, 4000, 8000$. Empirical rejection frequencies obtained from competing test devices are given in Table 1.

With respect to evaluating the cross product of first order derivatives it makes little difference if a numerical or an analytical approach is followed. In this case the empirical size estimates are very close under normally as well as leptokurtically distributed innovations. Using the QML-covariance estimate $\widehat{\mathcal{I}}^{-1}$ in case of conditional normality involves large size distortions in smaller

samples ($T = 1000, 2000$). Thus, in practice QML based p-values should only be preferred over ML based counterparts if diagnostic tests (e.g. Jarque Bera) indicate nonnormality of ξ_t . Empirical size distortions of QML-tests are more severe when testing off-diagonal coefficient estimates (H_{01}) in comparison to a situation in which "pure" ARCH-dynamics are tested (H_{02}). For instance under Gaussian innovations analytical evaluations of the QML covariance matrix deliver for the smallest sample size considered ($T = 1000$) empirical rejection frequencies of 35.8% (H_{01}) and 19.5% (H_{02}), respectively. Analogously, size estimates are 64.8% (!) (H_{01}) and 19.7% (H_{02}), respectively, if the corresponding covariances are determined numerically. Under leptokurtic innovations, maximum likelihood inference based on the normal distribution delivers severe size distortions which appear to stabilize for large sample sizes ($T = 4000, 8000$) around 26% (H_{01}) and 29% (H_{02}), respectively. In case of leptokurtic innovations QML-inference on diagonal parameters yields similar empirical size estimates if the relevant covariance matrix is numerically or analytically determined. In this case (H_{02}) analytically computed test statistics give size estimates that are slightly closer to the nominal test level. This picture changes dramatically if the hypothesis of interest formalizes restrictions for off diagonal dynamics. When testing H_{02} in presence of leptokurtic innovations via a numerically estimated covariance matrix the estimated size is at least 54% and it appears that this distortion will not vanish in even larger samples. The empirical size of the same test implemented with analytical derivatives is for processes of length $T = 8000$ 6.65% which is fairly close to the nominal level. The latter result has important implications for practical work. Implementing QML-based tests numerically might involve the risk of overstating the significance of cross sectional volatility dynamics.

Evaluating the Hessian matrix without using the second order derivatives of the covariances providing $\tilde{\mathcal{J}}$ turns out to outperform the more involved estimator $\hat{\mathcal{J}}$ in terms of the empirical significance level. In smaller samples ($T = 1000, 2000$) the size distortions under both, the normal and the conditionally leptokurtic process, are drastically reduced. In case of conditionally leptokurtic innovations and $T = 1000$ the empirical sizes involved with covari-

T	cov	Normal model				t-model			
		1000	2000	4000	8000	1000	2000	4000	8000
H_{01}	$\widehat{\mathcal{I}}^{-1}$.083	.076	.063	.064	.240	.295	.262	.264
	$\widehat{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widehat{\mathcal{J}}^{-1}$.358	.165	.085	.072	.450	.210	.096	.067
	$\widetilde{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widetilde{\mathcal{J}}^{-1}$.120	.091	.069	.070	.124	.116	.085	.062
	$\bar{\mathcal{I}}^{-1}$.083	.076	.063	.064	.240	.295	.262	.264
	$\bar{\mathcal{J}}^{-1}\bar{\mathcal{I}}\bar{\mathcal{J}}^{-1}$.648	.586	.584	.596	.661	.595	.544	.550
H_{02}	$\widehat{\mathcal{I}}^{-1}$.052	.082	.062	.046	.217	.272	.293	.294
	$\widehat{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widehat{\mathcal{J}}^{-1}$.195	.101	.067	.054	.204	.118	.069	.054
	$\widetilde{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widetilde{\mathcal{J}}^{-1}$.087	.097	.067	.054	.090	.092	.087	.067
	$\bar{\mathcal{I}}^{-1}$.051	.082	.062	.046	.217	.272	.293	.294
	$\bar{\mathcal{J}}^{-1}\bar{\mathcal{I}}\bar{\mathcal{J}}^{-1}$.197	.102	.065	.046	.210	.121	.074	.046

Table 1: *Empirical relative rejection frequencies at nominal level 5% of the null hypotheses H_{01} (off-diagonal elements of A and B) and H_{02} (diagonal elements of A and B). The estimate $\widehat{\mathcal{J}}$ involves computation of second derivatives of H_t , $\widetilde{\mathcal{J}}$ does not. $\bar{\mathcal{I}}$ and $\bar{\mathcal{J}}$ are using numerical derivatives to calculate the score and the Hessian.*

ance estimators $\widehat{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widehat{\mathcal{J}}^{-1}$ and $\widetilde{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widetilde{\mathcal{J}}^{-1}$ are 45.0% and 12.4% (20.4% and 9.0%), respectively, when testing H_{01} (H_{02}). Testing only the diagonal ARCH parameters H_{02} in large samples ($T = 4000, 8000$) under leptokurtic innovations, however, slightly better size estimates are obtained when using the more involved covariance estimator $\widehat{\mathcal{J}}$.

5 Conclusions and outlook

A clear conclusion of this paper and the simulation study can be drawn: Numerical derivatives should be avoided when computing the score and the Hessian in multivariate volatility models. Concerning the alternative analytical ways to estimate the expected Hessian, it seems that in small and medium samples the method that avoids calculation of second derivatives of H_t is

preferred, whereas the method involving calculation of second derivatives is slightly preferable in large samples.

A more elaborate method to obtain valid tests in small and medium samples is based on the bootstrap, see e.g. Hafner and Herwartz (2000) and Hafner and Herwartz (2002). As this is computationally challenging already in univariate AR and in VAR type models, we refrained from applying this in a multivariate volatility framework, but leave it to future research.

References

- Alexander, C.O. (2001), *Orthogonal GARCH*, in *Mastering Risk*, Volume II, edited by C.O. Alexander, pp. 21–38, Prentice Hall.
- Bauwens, L., S. Laurent, and J.V.K. Rombouts (2003), Multivariate GARCH models: A survey, CORE Discussion paper 31/2003, Université catholique de Louvain, Louvain-la-Neuve, Belgium.
- 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. and O. Lieberman (2003), Asymptotic Theory for Multivariate GARCH Processes, *Journal of Multivariate Analysis* 84, 61–84.
- Engle, R.F. and K.F. Kroner (1995), Multivariate Simultaneous Generalized ARCH. *Econometric Theory* 11, 122–150.
- Engle, R.F., V.M. Ng and Rothschild (1990), Asset Pricing with a Factor-ARCH Structure: Empirical Estimates for Treasury Bills. *Journal of Econometrics* 45, 213–237.
- Fiorentini, G., E. Sentana and G. Calzolari (2003), Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroskedastic Dynamic Regression Models with Student t Innovations, *Journal of Business and Economic Statistics*, forthcoming.

- Hafner, C.M. and H. Herwartz (2000), Testing for Linear Autoregressive Dynamics under Heteroskedasticity, *The Econometrics Journal* 3, 177–197.
- Hafner, C.M. and H. Herwartz (2002), Testing for vector autoregressive dynamics under heteroskedasticity, Econometric Institute Report 36, Erasmus University Rotterdam.
- Jeantheau, T. (1998), Strong Consistency of Estimators for Multivariate ARCH Models, *Econometric Theory* 14, 70–86.
- Ling, S. and M. McAleer (2003), Asymptotic Theory for a vector ARMA-GARCH Model, *Econometric Theory* 19, 280–309.
- Lucchetti, R. (2001), Analytical score for multivariate GARCH models, *Computational Economics* 19, 133–143.
- Lütkepohl, H. (1996), *Handbook of Matrices*, New York: Wiley.
- van der Weide, R. (2002), GO-GARCH: A multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* 17, 549–564.