Università degli Studi di Ancona

DIPARTIMENTO DI ECONOMIA

Analytic Score for Multivariate GARCH Models

Riccardo Lucchetti

QUADERNI DI RICERCA n. 119

Ottobre 1999

Abstract

Multivariate GARCH models constitute the workhorse of empirical applications in several fields, a notable example being financial econometrics. Unfortunately, ML (or quasi-ML) estimation of such models, although relatively straightforward in theory, is often made difficult by the fact that available software relies on numerical methods for computing the first derivatives of the log-likelihood; the fact that these models often include several dozens of parameters makes it impractical to estimate even medium-sized models. In this paper, closed-form expressions for the score of the BEKK model of Engle and Kroner (1995) are obtained, and strategies for efficient computation are discussed.

Indirizzo: Dipartimento di Economia
Università di Ancona
P.le Martelli, 8 — 60121 Ancona

Analytic Score for Multivariate GARCH Models*

Riccardo Lucchetti

1 Introduction

Multivariate GARCH models constitute the workhorse of empirical applications in several fields, a notable example being the econometrics of portfolio allocation. Unfortunately, maximum likelihood estimation of such models, although relatively straightforward in theory, is often made difficult by the fact that available software relies on numerical methods for computing the first derivatives of the log-likelihood; apart from the inevitable loss in precision that this practice entails, the fact that these models often include several dozens of parameters makes it impractical to estimate even medium-sized models. Analytic derivatives for the GARCH log-likelihood have been considered by Fiorentini et al. (1996); in that paper, however, only univariate models without *in-mean* effects were considered. The present paper overcomes both limitations.

The unavailability of algorithms for analytic derivatives makes estimation of multivariate GARCH models slow and prone to numerical errors[1]. This is one of the reasons that motivated several authors to consider models with a reduced number of parameters. Various strategies have been developed for reducing the number of parameters involved in multivariate GARCH models: for example, Bollerslev et al. (1988), Diebold and Nerlove (1989), Bollerslev (1990) and Giannini and Rossi (1999); Bollerslev et al. (1994) provide a fairly complete exposition.

The most popular in the applied literature has proven the so-called BEKK model, which balances generality with parsimony. In the original article that put the BEKK model forward (Engle and Kroner (1995)), the issue of analytic derivatives was issued, but not deemed to be worth pursuing in view

I would like to thank Eduardo Rossi and Carlo Giannini for their useful comments without implicating them in whatever errors or inaccuracies might be in the paper.

[1]The general problem of using numerical rather than analytic derivatives in optimization problems is discussed at length in Quandt (1983).

of the algebraic complications; the authors advocated the use of numerical derivatives instead. In this paper, closed-form expressions for the derivatives of the log-likelihood for the BEKK model are obtained, and strategies for efficient computation are discussed.

Enhancing the performance (both in terms of speed and accuracy) of existing software for the estimation of multivariate GARCH models could be crucial in several contexts: first, existing models could be estimated and updated more easily: this may be particularly important in applied finance, where multivariate GARCH models are used for asset allocation or risk management; larger models could be estimated (or equivalently, models with less restrictions); finally, simulation-based techniques for inference on stochastic volatility (SV) models, which have relied so far mainly on univariate GARCH models, could benefit from the use of multivariate GARCH, as an auxiliary model as shorter estimation times make massive simulation feasible[2].

The plan of the paper is as follows: in section 2, the univariate GARCH model is discussed, mainly to establish notation. Section 3 generalises to the multivariate case. The extension to models with the variance entering the conditional mean function (the so-called *in-mean* models) is analysed in section 4. Section 5 contains a discussion of some practical issues for efficient evaluation of the formulas given in section 4. Finally, section 6 ends the paper.

2 The univariate case

In the univariate case, assume that an observable scalar stochastic process y_t is distributed, conditionally to an increasing sequence of σ-fields $\mathcal{I}_{t-1}$ (henceforth referred to as the *information set* at time $t-1$), as[3]

$$y_t|\mathcal{I}_{t-1} \sim N(\mu_t, h_t) \tag{1}$$

In the GARCH model (put forward by Bollerslev (1986) building on work by Engle (1982)) the expressions for the conditional mean and variance are

$$\mu_t = x_t'\beta \tag{2}$$

$$h_t = c + a e_{t-1}^2 + b h_{t-1} \tag{3}$$

[2]A standard reference for simulation-based inference is Gourieroux and Monfort (1996); for a recent contibution in the field of SV models estimation via simulation techniques, see Andersen et al. (1999).

[3]Although various distributions have been suggested, only the case of normality will be considered here.

where x_t is a vector of weakly exogenous observable variables, measurable with respect to $\mathcal{I}_{t-1}$. By using a sequential factorization argument, the log-likelihood for the t-th observation in the sample can be written as

$$\ell_t = \text{const} - \frac{1}{2}\left[\ln h_t + \frac{e_t^2}{h_t}\right] \tag{4}$$

where

$$e_t = y_t - x_t'\beta \tag{5}$$

The score vector is

$$s_t(\theta) = \frac{\mathrm{d}\,\ell_t}{\mathrm{d}\,\theta}$$

where the parameters β, c, a and b have been gathered (in this order) in the vector θ. Explicit calculation of $s_t(\theta)$ is made easier by applying the chain rule:

$$s_t(\theta) = \frac{\partial \ell_t}{\partial e_t}\frac{\mathrm{d}\,e_t}{\mathrm{d}\,\theta} + \frac{\partial \ell_t}{\partial h_t}\frac{\mathrm{d}\,h_t}{\mathrm{d}\,\theta} \tag{6}$$

As will be seen, this expression not only is theoretically useful, but it also leads to an advantageous computational strategy.

Let us consider the elements of (6) one by one: the derivatives of the log-likelihood itself are

$$\frac{\partial \ell_t}{\partial e_t} = -\frac{e_t}{h_t} \tag{7}$$

$$\frac{\partial \ell_t}{\partial h_t} = \frac{1}{2h_t}\left[\frac{e_t^2}{h_t} - 1\right] \tag{8}$$

Derivatives with respect to the elements of the vector θ are more complex: the first one is simple, since it is apparent from (5) that $\frac{\partial e_t}{\partial \beta} = -x_t$, while derivatives with respect to the other parameters are all 0. Calculation of $\frac{\mathrm{d}\,h_t}{\mathrm{d}\,\theta}$ is not so straightforward, as the recursive nature of (3) must be taken into account. Considering, for example, the parameter c, we have

$$\frac{\partial h_t}{\partial c} = 1 + b\frac{\partial h_{t-1}}{\partial c}$$

which, after defining $h_t^c \equiv \frac{\partial h_t}{\partial c}$, could be written as

$$h_t^c = b h_{t-1}^c + 1 \tag{9}$$

so that the "autoregressive" nature of the equation becomes evident; an analogous argument leads to

$$h_t^a = b h_{t-1}^a + e_{t-1}^2 \tag{10}$$

$$h_t^b = b h_{t-1}^b + h_{t-1} \tag{11}$$

Computation of $\frac{\partial h_t}{\partial \beta}$ yields:

$$\frac{\partial h_t}{\partial \beta} = a\frac{\partial e_{t-1}^2}{\partial \beta} + b\frac{\partial h_{t-1}}{\partial \beta} = -2ae_{t-1}x_{t-1} + b\frac{\partial h_{t-1}}{\partial \beta}$$

which can be written as

$$h_t^\beta = b h_{t-1}^\beta - 2ae_{t-1}x_{t-1} \tag{12}$$

so that one ends up with

$$h_t^\theta = \begin{bmatrix} h_{t-1}^\beta \\ h_{t-1}^c \\ h_{t-1}^a \\ h_{t-1}^b \end{bmatrix} = b h_{t-1}^\theta + \begin{bmatrix} -2ae_{t-1}x_{t-1} \\ 1 \\ e_{t-1}^2 \\ h_{t-1} \end{bmatrix} \tag{13}$$

and finally, combining all intermediate results into (6) yields:

$$s_t(\theta) = \begin{bmatrix} \frac{e_t x_t}{h_t} \\ 0 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{2h_t}\left[\frac{e_t^2}{h_t} - 1\right] h_t^\theta \tag{14}$$

Obviously, since expression (14) contains a recursive term, an initialisation rule is required. Such a rule, however, is also made necessary for the evaluation of the log-likelihood itself by the recursive term in (3). A common[4] choice is initialising h_t with the unconditional variance, ie letting

$$h_0 = \frac{\sum_{t=1}^T e_t^2}{T}$$

In this case, initial values for the recurrence relations in (13) are simply

$$\frac{\mathrm{d}\, h_0}{\mathrm{d}\, \theta} = \frac{2\sum_{t=1}^T e_t \frac{\mathrm{d}\, e_t}{\mathrm{d}\, \theta}}{T}$$

by which

$$h_0^\theta = \begin{bmatrix} -\frac{2\sum_{t=1}^T e_t x_t}{T} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{15}$$

[4]This choice is indeed common, although its asymptotic implications for the resulting estimators are unclear.

is obtained.

A second possibility is initialising h_t by setting h_0 at its lowest possible value, $h_0 = c$. This choice is probably less realistic, but it certainly is computationally more effective. If the GARCH model is not integrated (which happens if $a + b < 1$), this choice is inconsequential, as the sample size increases, anyway. In this case, h_0^θ does not depend on e_t (and therefore on β), but is a function of c only, so that

$$h_0^\theta = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \tag{16}$$

Yet another possibility is choosing for h_0 its unconditional value and setting

$$h_0 = \frac{c}{1 - a - b}$$

so that

$$h_0^\theta = \frac{1}{1 - a - b} \begin{bmatrix} 0 \\ 1 \\ h_0 \\ h_0 \end{bmatrix} \tag{17}$$

This choice is by far the most attractive from a theoretical standpoint; however, in practice it may well happen that, whatever maximization algorithm is used, the log-likelihood and its derivative get evaluated in a region of the parameter space where $a + b \geq 1$, and the expressions above (notably (17)) are inapplicable. It is chiefly for this reason that this choice is very rarely made.

3 The BEKK case

In a multivariate setting, the observable process y_t is a vector process with n elements. This implies that extension of the univariate model must take into account time variability not only of a vector of conditional means, but also of the $n \times n$ covariance matrix H_t. While the conditional mean is almost invariably parametrized as a linear function, several choices exist for the covariance matrix. In this paper, we will only consider a special case of the so-called BEKK parametrisation (see Engle and Kroner (1995)), which has proved the most popular. In this model, the conditional variance matrix is specified as

$$H_t = CC' + \sum_{s=1}^{p}\sum_{i=1}^{k} A_i e_{t-s} e'_{t-s} A'_i + \sum_{s=1}^{q}\sum_{i=1}^{k} B_i H_{t-s} B'_i$$

where $1 \le k \le n^2$ is called the *degree of generality* of the model. This formulation achieves several objectives: it retains high generality (several other multivariate models are special cases of the BEKK model), while ensuring positive semi-definiteness of the matrix H_t and identification of the parameters under rather mild conditions that are easily checked in practice.

It can be shown that, for an appropriate choice of k, this parametrization spans the whole space of positive semidefinite symmetric matrices. In what follows, though, generality will be sacrificed for the sake of conciseness and only the case where $k = p = q = 1$ will be analysed. On the other hand, it is worthwhile noting that the extension to the general case is algebraically messy but conceptually simple; moreover, the vast majority of empirical applications uses the same setup as ours.

For the BEKK parametrisation, the (conditional) log-likelihood for the $t-$th observation is

$$\ell_t = \text{const} - \frac{1}{2}\left[\ln|H_t| + e'_t H_t^{-1} e_t\right] \tag{18}$$

where

$$e_t = y_t - \Pi x_t \tag{19}$$

$$H_t = CC' + A e_{t-1} e'_{t-1} A' + B H_{t-1} B' \tag{20}$$

and equation (20) can be written in vector form as follows:

$$h_t = (C \otimes C)\operatorname{vec} I + (A \otimes A)(e_{t-1} \otimes e_{t-1}) + (B \otimes B) h_{t-1} \tag{21}$$

where $h_t = \operatorname{vec} H_t$. It is useful to define $P_t = H_t^{-1}$ and its vector form $p_t = \operatorname{vec} P_t$. Between h_t and p_t the following relation holds:

$$\frac{\partial p_t}{\partial h_t} = -H_t^{-1} \otimes H_t^{-1} = -P_t \otimes P_t$$

As in the univariate case, the parameters that have to be estimated are Π, C, A and B, or, equivalently, their vectorised versions π, c, a e b. In this notation, the log-likelihood for the t-th observation can be written as

$$\ell_t = \text{const} - \frac{1}{2}\left[-\ln|P_t| + (e'_t \otimes e'_t) p_t\right] \tag{22}$$

Since for symmetric invertible matrices $\frac{\partial \ln|A|}{\partial \text{vec}\, A} = \text{vec}\,(A^{-1})'$ holds, we have

$$\frac{\partial \ell_t}{\partial e_t} = -e_t' H_t^{-1} = -u_t' \tag{23}$$

$$\frac{\partial \ell_t}{\partial h_t} = \frac{\partial \ell_t}{\partial p_t} \frac{\partial p_t}{\partial h_t} = \frac{1}{2} \left[(e_t' \otimes e_t') - h_t' \right] (P_t \otimes P_t) = \frac{1}{2} \left[(u_t' \otimes u_t') - p_t' \right] \tag{24}$$

where $u_t \equiv H_t^{-1} e_t$.

It is now possible to exploit the relations (19) and (20) in the same way as in the univariate case, thus obtaining

$$\frac{\partial e_t}{\partial \pi} = -x_t' \otimes I$$

with the other partial derivatives with respect to e_t are 0. Now, the rule[5]

$$\frac{\partial \text{vec}\,(XAX')}{\partial \text{vec}\, X} = (I + \textcircled{T})(XA \otimes I)$$

gives rise to

$$\frac{\partial h_t}{\partial c} = (I + \textcircled{T})(C \otimes I) + (B \otimes B) \frac{\partial h_{t-1}}{\partial c}$$

Actually, the last expression overlooks the fact that in the BEKK setup C is a lower triangular matrix. This constraint, however, can be handled in a straightforward way by writing

$$\text{vec}\, C = c = D\gamma$$

where D is a matrix (sometimes known as the 'elimination matrix' — see Magnus (1988)) with n^2 rows and $n(n+1)/2$ columns that deletes the superdiagonal elements of C. As an example, for $n = 2$, D is equal to

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

[5]The symbol $\textcircled{T}$ is defined in Pollock (1979)) as the permutation matrix (called the *commutation matrix*) defined by the property

$$\textcircled{T} \text{vec}\, A = \text{vec}\, A'$$

An additional property of $\textcircled{T}$ that will be used in what follows is that

$$\textcircled{T}\,(A \otimes B) = (B \otimes A)\,\textcircled{T}$$

Other authors (for instance Magnus (1988) or Magnus and Neudecker (1988)) use the symbol K_{mp} instead.

Since $\frac{\partial c}{\partial \gamma} = D$, application of the chain rule yields

$$\frac{\partial h_t}{\partial \gamma} = (I + \textcircled{\scriptsize T})(C \otimes I)D + (B \otimes B)\frac{\partial h_{t-1}}{\partial \gamma}$$

In practice, this linear transformation boils down to selecting the appropriate columns of $\frac{\partial h_t}{\partial c}$. The same operation is unnecessary for A and B, which are normally unrestricted $n \times n$ matrices.

$$\frac{\partial h_t}{\partial a} = (I + \textcircled{\scriptsize T})(Ae_{t-1}e'_{t-1} \otimes I) + (B \otimes B)\frac{\partial h_{t-1}}{\partial a}$$
$$\frac{\partial h_t}{\partial b} = (I + \textcircled{\scriptsize T})(BH_t \otimes I) + (B \otimes B)\frac{\partial h_{t-1}}{\partial b}$$

It is possible to write these results in a more compact (and computationally more covenient) way:

$$\begin{aligned} \left[\ \frac{\partial h_t}{\partial c} \,\middle|\, \frac{\partial h_t}{\partial a} \,\middle|\, \frac{\partial h_t}{\partial b}\ \right] &= (I + \textcircled{\scriptsize T})\left\{\left[\ C \,\middle|\, Ae_{t-1}e'_{t-1} \,\middle|\, BH_{t-1}\ \right] \otimes I\right\} + \\ &+ (B \otimes B)\left[\ \frac{\partial h_{t-1}}{\partial c} \,\middle|\, \frac{\partial h_{t-1}}{\partial a} \,\middle|\, \frac{\partial h_{t-1}}{\partial b}\ \right] \end{aligned} \tag{25}$$

where, again, $\frac{\partial h_t}{\partial \gamma} = \frac{\partial h_t}{\partial c} D$.

Computation of the derivative of h_t with respect to π can proceed along the following line: equation (21) makes it possible to write it as the sum of two components, so that

$$\frac{\partial h_t}{\partial \pi} = (A \otimes A)\frac{\partial(e_{t-1} \otimes e_{t-1})}{\partial \pi} + (B \otimes B)\frac{\partial h_{t-1}}{\partial \pi}$$

and the only element left to evaluate is $\frac{\partial(e_{t-1} \otimes e_{t-1})}{\partial \pi}$; this can be done by applying the chain rule again

$$\frac{\partial(e_{t-1} \otimes e_{t-1})}{\partial \pi} = \frac{\partial(e_{t-1} \otimes e_{t-1})}{\partial e_{t-1}} \cdot \frac{\partial e_{t-1}}{\partial \pi} = (I + \textcircled{\scriptsize T})(e_{t-1} \otimes I) \cdot (-x'_{t-1} \otimes I)$$

and putting everything back together yields

$$\frac{\partial h_t}{\partial \pi} = -(I + \textcircled{\scriptsize T})(Ae_{t-1}x'_{t-1} \otimes A) + (B \otimes B)\frac{\partial h_{t-1}}{\partial \pi} \tag{26}$$

The complete score can now be written as

$$s_t(\theta) = \frac{\partial \ell_t}{\partial e_t}\left[\ \frac{\partial e_t}{\partial \pi} \,\middle|\, \frac{\partial e_t}{\partial \gamma} \,\middle|\, \frac{\partial e_t}{\partial a} \,\middle|\, \frac{\partial e_t}{\partial \pi}\ \right] + \frac{\partial \ell_t}{\partial h_t}\left[\ \frac{\partial h_t}{\partial \pi} \,\middle|\, \frac{\partial h_t}{\partial c} \,\middle|\, \frac{\partial h_t}{\partial a} \,\middle|\, \frac{\partial h_t}{\partial b}\ \right]$$

so that

$$s_t(\theta) = -u_t' \left[\begin{array}{c|c|c|c} -x_t' \otimes I & 0 & 0 & 0 \end{array} \right] + \frac{1}{2} [(u_t' \otimes u_t') - p_t'] \left[\begin{array}{c|c|c|c} \frac{\partial h_t}{\partial \pi} & \frac{\partial h_t}{\partial c} & \frac{\partial h_t}{\partial a} & \frac{\partial h_t}{\partial b} \end{array} \right] \quad (27)$$

where the derivatives of h_t are evaluated recursively by using (25) and (26). From a computational point of view, it might be advantageous to exploit the following relations:

$$-u_t'(-x_t' \otimes I) = \text{vec}\,(u_t x_t')' \quad (28)$$

$$(u_t' \otimes u_t') - p_t' = \text{vec}\left(u_t u_t' - H_t^{-1}\right)' \quad (29)$$

The problem of initialising the recursions is analogous to that already considered for the univariate case. Initialising $H_0 = CC'$ yields simply

$$\frac{\partial h_0}{\partial \theta} = \left[\begin{array}{c|c|c|c} 0 & (I + ⓣ)(C \otimes I)D & 0 & 0 \end{array} \right]$$

whereas setting H_0 to the unconditional sample variance, given by $T^{-1} \sum_{t=1}^{T} e_t e_t'$ leads to

$$\frac{\partial h_0}{\partial \theta} = \left[\begin{array}{c|c|c|c} -\frac{1}{T}(I + ⓣ)(I \otimes \sum_{t=1}^{T} e_t x_t') & 0 & 0 & 0 \end{array} \right]$$

4 GARCH in mean

In the GARCH-in-mean (GARCH-M) case things are more complex, since the conditional mean function includes a term which accounts for the effects of the contemporaneous variance matrix[6].

As before, let us consider the univariate case first: in this case, equation (5) includes a term $h_t \phi$, while (3) remains unchanged; therefore:

$$e_t = y_t - x_t'\beta - h_t \phi \quad (30)$$

$$h_t = c + a e_{t-1}^2 + b h_{t-1} \quad (31)$$

Since (4) and the equations (6)–(8) are still valid, the recursive relations for evaluating the derivatives of e_t and h_t with respect to θ can be synthesized in the following matrix equation:

$$\begin{bmatrix} 1 & \phi \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_t^\theta \\ h_t^\theta \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2ae_{t-1} & b \end{bmatrix} \begin{bmatrix} e_{t-1}^\theta \\ h_{t-1}^\theta \end{bmatrix} + \begin{bmatrix} -x_t \frac{\partial \beta}{\partial \theta} - h_t \frac{\partial \phi}{\partial \theta} \\ \frac{\partial c}{\partial \theta} + e_{t-1}^2 \frac{\partial a}{\partial \theta} + h_{t-1} \frac{\partial b}{\partial \theta} \end{bmatrix}$$

[6] In the univariate case, the possibility of using the standard deviation instead of the variance has also been explored. At this stage, however, considering the general case would introduce unnecessary complications.

Since

$$\begin{bmatrix} 1 & \phi \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -\phi \\ 0 & 1 \end{bmatrix}$$

one easily obtains

$$\begin{bmatrix} e_t^\theta \\ h_t^\theta \end{bmatrix} = \begin{bmatrix} 1 & -\phi \\ 0 & 1 \end{bmatrix} \left\{ \begin{bmatrix} 0 & 0 \\ 2ae_{t-1} & b \end{bmatrix} \begin{bmatrix} e_{t-1}^\theta \\ h_{t-1}^\theta \end{bmatrix} + \begin{bmatrix} -x_t \frac{\partial \beta}{\partial \theta} - h_t \frac{\partial \phi}{\partial \theta} \\ \frac{\partial c}{\partial \theta} + e_{t-1}^2 \frac{\partial a}{\partial \theta} + h_{t-1} \frac{\partial b}{\partial \theta} \end{bmatrix} \right\} \tag{32}$$

It should be noted that all the expressions like $\frac{\partial \cdot}{\partial \theta}$ are selection matrices, whose elements are 0 or 1. Although equation (32) may look messy, computational difficulties are limited, as will be discussed later.

The multivariate analogue to (19) is

$$e_t = y_t - \Pi x_t - \Phi D' h_t = y_t - (x_t' \otimes I)\pi - (h_t' D \otimes I)\phi \tag{33}$$

where Φ is an $n \times \frac{n(n+1)}{2}$ matrix whose identification is attained by inserting the D matrix, as $D'h_t = \text{vech}\,(H_t)$; expressions (20) and (21) stay unmodified. The derivatives for the conditional mean and variance respectively are therefore

$$\frac{\partial e_t}{\partial \theta} = -(x_t' \otimes I)\frac{\partial \pi}{\partial \theta} - (h_t' D \otimes I)\frac{\partial \phi}{\partial \theta} - \Phi D' \frac{\partial h_t}{\partial \theta} \tag{34}$$

$$\begin{aligned} \frac{\partial h_t}{\partial \theta} &= (I + \textcircled{\scriptsize T})(Ae_{t-1} \otimes A)\frac{\partial e_{t-1}}{\partial \theta} + (B \otimes B)\frac{\partial h_{t-1}}{\partial \theta} + \\ &+ (I + \textcircled{\scriptsize T})\left[(C \otimes I)D\frac{\partial \gamma}{\partial \theta} + (Ae_{t-1}e_{t-1}' \otimes I)\frac{\partial a}{\partial \theta} + (BH_{t-1} \otimes I)\frac{\partial b}{\partial \theta}\right] \end{aligned} \tag{35}$$

Now the multivariate equivalent to (32) can be written as

$$\begin{aligned} \begin{bmatrix} e_t^\theta \\ h_t^\theta \end{bmatrix} &= \begin{bmatrix} 1 & -\Phi D' \\ 0 & 1 \end{bmatrix} \left\{ \begin{bmatrix} 0 & 0 \\ (I + \textcircled{\scriptsize T})(Ae_{t-1} \otimes A) & (B \otimes B) \end{bmatrix} \begin{bmatrix} e_{t-1}^\theta \\ h_{t-1}^\theta \end{bmatrix} + \right. \\ &+ \left. \begin{bmatrix} -(x_t' \otimes I)\frac{\partial \pi}{\partial \theta} - (h_t' D \otimes I)\frac{\partial \phi}{\partial \theta} \\ (I + \textcircled{\scriptsize T})\left[(C \otimes I)D\frac{\partial \gamma}{\partial \theta} + (Ae_{t-1}e_{t-1}' \otimes I)\frac{\partial a}{\partial \theta} + (BH_{t-1} \otimes I)\frac{\partial b}{\partial \theta}\right] \end{bmatrix} \right\} \end{aligned} \tag{36}$$

which looks formidable, but whose implementation is not overly difficult (consider, again, that all the terms like $\frac{\partial \cdot}{\partial \theta}$ are selection matrices). Recursive evaluation of the preceeding expression makes it possible to evaluate the score by using

$$s(\theta) = \begin{bmatrix} \frac{\partial \ell_t}{\partial e_t} & \frac{\partial \ell_t}{\partial h_t} \end{bmatrix} \begin{bmatrix} e_t^\theta \\ h_t^\theta \end{bmatrix} = \begin{bmatrix} -u_t' & \frac{1}{2}\left[(u_t' \otimes u_t') - p_t'\right] \end{bmatrix} \begin{bmatrix} e_t^\theta \\ h_t^\theta \end{bmatrix}$$

5 Computational details

Any procedure for estimating BEKK models taking advantage of the analytic score must implement equations (18) and (36) as part of an iterative optimization scheme. The two expressions will typically be evaluated over a loop from 1 to T, with the initialization choices discussed above. Especially the latter expression looks rather complicated, but practical difficulties in software implementation can be mitigated considering that:

1. The matrices D, $\textcircled{\top}$, and all the $\frac{\partial \cdot}{\partial \theta}$ are selection matrices, so that pre-(post-)multiplication by any of these matrices is equivalent to selecting the appropriate rows (columns). In a matrix-oriented programming language, such as GAUSS, Ox or Matlab, this task can be accomplished very easily and efficiently at the same time, thus drastically reducing the number of floating-point operations necessary.

2. Many elements of (36), for example $(B \otimes B)$, are time-invariant, so they only have to be evaluated once per iteration, outside the main loop. This reduces the number of total multiplications and therefore increases the speed of the algorithm.

3. The same loop can be used for evaluating both the log-likelihood and the score. This would reduce computational time, as quantities like H_t^{-1} are evaluated only once per iteration. Put another way, with little extra computational effort it is possible to evaluate the score every time the log-likelihood is computed. This can enhance performances if combined with a maximization scheme like BHHH (without line search).

6 Conclusions

In this paper, explicit expressions for the score of the BEKK model with in-mean effects are obtained. Although the resulting formulas are rather complex (or, in Engle and Kroner's words, "cumbersome" (Engle and Kroner, 1995, p. 139)), they lend themselves to efficient translation into a computer algorithm. Implementation of such an algorithm would allow for efficient estimation of large scale multivariate GARCH models.

Both theoretical and applied research would benefit for improved software speed and accuracy: in theoretical research, the way would be opened to large-scale simulation, needed for estimation of SV models and Monte Carlo experiments (asymptotic properties of ML estimators for multivariate GARCH models are still the object of current research: see for instance

Jeantheau (1998)). Applied research, on the other hand, could consider estimation of models previously considered intractable due to the high number of parameters.

References

Torben G. Andersen, Hyung-Jin Chung, and Bent E. Sørensen. Efficient method of moments estimation of a stochastic volatility model: A monte carlo study. *Journal of Econometrics*, 91:61–87, 1999.

Tim Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, 72:498–505, 1990.

Tim Bollerslev, Robert F. Engle, and D.B. Nelson. ARCH models. In Robert F. Engle and D.L. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 49, pages 2959–3040. North-Holland, 1994.

Tim Bollerslev, Robert F. Engle, and Jeffrey M. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–31, 1988.

Francis X. Diebold and Mark Nerlove. The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics*, 4:1–21, 1989.

Robert F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1008, 1982.

Robert F. Engle and Kenneth F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11:122–150, 1995.

Gabriele Fiorentini, Giorgio Calzolari, and Lorenzo Panattoni. Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics*, 11:399–417, 1996.

Carlo Giannini and Eduardo Rossi. A principal components multivariate GARCH technique for medium size portfolio management. *Collana Studi del Credito Italiano*, 6/99, 1999.

Christian Gourieroux and Alain Monfort. *Simulation-Based Econometric Methods*. Oxford University Press, 1996.

Thierry Jeantheau. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14:70–86, 1998.

Jan Magnus. *Linear Structures*. Charles Griffin & Co., 1988.

Jan Magnus and Heinz Neudecker. *Matrix Differential Calculus*. John Wiley & Sons, 1988.

D.S.G. Pollock. *The Algebra of Econometrics*. John Wiley & Sons, 1979.

Richard E. Quandt. Computational problems and methods. In Zvi Griliches and Michael D. Intriligator, editors, *Handbook of Econometrics*, number 1, chapter 12, pages 699–764. North-Holland, 1983.