

## MULTIVARIATE GARCH MODELS: A SURVEY

LUC BAUWENS,<sup>a\*</sup> SÉBASTIEN LAURENT<sup>b</sup> AND JEROEN V. K. ROMBOUTS<sup>a</sup>

<sup>a</sup> *CORE and Department of Economics, Université catholique de Louvain, Belgium*

<sup>b</sup> *CeReFim, Université de Namur and CORE, Université catholique de Louvain, Belgium*

### SUMMARY

This paper surveys the most important developments in multivariate ARCH-type modelling. It reviews the model specifications and inference methods, and identifies likely directions of future research. Copyright © 2006 John Wiley & Sons, Ltd.

### 1. INTRODUCTION

Understanding and predicting the temporal dependence in the second-order moments of asset returns is important for many issues in financial econometrics. It is now widely accepted that financial volatilities move together over time across assets and markets. Recognizing this feature through a multivariate modelling framework leads to more relevant empirical models than working with separate univariate models. From a financial point of view, it opens the door to better decision tools in various areas, such as asset pricing, portfolio selection, option pricing, hedging and risk management. Indeed, unlike at the beginning of the 1990s, several institutions have now developed the necessary skills to use the econometric theory in a financial perspective.

Since the seminal paper of Engle (1982), traditional time series tools such as autoregressive moving average (ARMA) models (Box and Jenkins, 1970) for the mean have been extended to essentially analogous models for the variance. Autoregressive conditional heteroscedasticity (ARCH) models are now commonly used to describe and forecast changes in the volatility of financial time series. For a survey of ARCH-type models, see Bollerslev *et al.* (1992, 1994), Bera and Higgins (1993), Pagan (1996), Palm (1996) and Shephard (1996), among others.

The most obvious application of MGARCH (multivariate GARCH) models is the study of the relations between the volatilities and co-volatilities of several markets.<sup>1</sup> Is the volatility of a market leading the volatility of other markets? Is the volatility of an asset transmitted to another asset directly (through its conditional variance) or indirectly (through its conditional covariances)? Does a shock on a market increase the volatility on another market, and by how much? Is the impact the same for negative and positive shocks of the same amplitude? A related issue is whether the correlations between asset returns change over time.<sup>2</sup> Are they higher during periods of higher volatility (sometimes associated with financial crises)? Are they increasing in the long run, perhaps because of the globalization of financial markets? Such issues can be studied directly by using a multivariate model, and raise the question of the specification of the dynamics of covariances or correlations. In a slightly different perspective, a few papers have used MGARCH models to

---

\* Correspondence to: Luc Bauwens, CORE, 34 Voie du Roman Pays, B-1348 Louvain-La-Neuve, Belgium.

E-mail: bauwens@core.ucl.ac.be

<sup>1</sup> Kearney and Patton (2000) and Karolyi (1995) exemplify such studies.

<sup>2</sup> See Bollerslev (1990) and Longin and Solnik (1995).

assess the impact of volatility in financial markets on real variables like exports and output growth rates, and the volatility of these growth rates.<sup>3</sup>

Another application of MGARCH models is the computation of time-varying hedge ratios. Traditionally, constant hedge ratios are estimated by OLS as the slope of a regression of the spot return on the futures return, because this is equivalent to estimating the ratio of the covariance between spot and futures over the variance of the futures. Since a bivariate MGARCH model for the spot and futures returns directly specifies their conditional variance–covariance matrix, the hedge ratio can be computed as a byproduct of estimation and updated by using new observations as they become available. See Lien and Tse (2002) for a survey on hedging and additional references.

Asset pricing models relate returns to ‘factors’, such as the market return in the capital asset pricing model. A specific asset excess return (in excess of the risk-free return) may be expressed as a linear function of the market return. Assuming its constancy, the slope, or  $\beta$  coefficient, may be estimated by OLS. Like in the hedging case, since the  $\beta$  is the ratio of a covariance to a variance, an MGARCH model can be used to estimate time-varying  $\beta$  coefficients. See Bollerslev *et al.* (1988), De Santis and Gérard (1998), Hafner and Herwartz (1998) for examples.

Given an estimated univariate GARCH model on a return series, one knows the return conditional distribution, and one can forecast the value-at-risk (VaR) of a long or short position. When considering a portfolio of assets, the portfolio return can be computed directly from the asset shares and returns. A GARCH model can be fit to the portfolio returns for given weights. If the weight vector changes, the model has to be estimated again. On the contrary, if a multivariate GARCH model is fitted, the multivariate distribution of the returns can be used directly to compute the implied distribution of any portfolio. There is no need to re-estimate the model for different weight vectors. In the present state of the art, it is probably simpler to use the univariate framework if there are many assets, but we conjecture that using a multivariate specification may become a feasible alternative. Whether the univariate ‘repeated’ approach is more adequate than the multivariate one is an open question. The multivariate approach is illustrated by Giot and Laurent (2003) using a trivariate example with a time-varying correlation model.

MGARCH models were initially developed in the late 1980s and the first half of the 1990s, and after a period of tranquillity in the second half of the 1990s, this area seems to be experiencing again a quick expansion phase. MGARCH models are partly covered in Franses and van Dijk (2000), Gouriéroux (1997) and most of the surveys on ARCH models cited above, but none of them presents, as this one, a comprehensive and up-to-date survey of the field, including the most recent findings.

The paper is organized in the following way. In Section 2, we review existing MGARCH specifications. Section 3 is devoted to estimation problems and Section 4 to diagnostic tests. Finally, we offer our conclusions and ideas for further developments in Section 5.

## 2. OVERVIEW OF MODELS

Consider a vector stochastic process  $\{y_t\}$  of dimension  $N \times 1$ . As usual, we condition on the sigma field, denoted by  $I_{t-1}$ , generated by the past information (here the  $y_t$ ’s) until time  $t - 1$ . We denote by  $\theta$  a finite vector of parameters and write:

$$y_t = \mu_t(\theta) + \varepsilon_t \quad (1)$$

<sup>3</sup> See Kim (2000).

where  $\mu_t(\theta)$  is the conditional mean vector and

$$\varepsilon_t = H_t^{1/2}(\theta)z_t \quad (2)$$

where  $H_t^{1/2}(\theta)$  is a  $N \times N$  positive definite matrix. Furthermore, we assume the  $N \times 1$  random vector  $z_t$  to have the following first two moments:

$$\begin{aligned} E(z_t) &= 0 \\ \text{Var}(z_t) &= I_N \end{aligned} \quad (3)$$

where  $I_N$  is the identity matrix of order  $N$ . We still have to explain what  $H_t^{1/2}$  is (for convenience we leave out  $\theta$  in the notation). To make this clear we calculate the conditional variance matrix of  $y_t$ :

$$\begin{aligned} \text{Var}(y_t|I_{t-1}) &= \text{Var}_{t-1}(y_t) = \text{Var}_{t-1}(\varepsilon_t) \\ &= H_t^{1/2} \text{Var}_{t-1}(z_t) (H_t^{1/2})' \\ &= H_t \end{aligned} \quad (4)$$

Hence  $H_t^{1/2}$  is any  $N \times N$  positive definite matrix such that  $H_t$  is the conditional variance matrix of  $y_t$ , e.g.  $H_t^{1/2}$  may be obtained by the Cholesky factorization of  $H_t$ . Both  $H_t$  and  $\mu_t$  depend on the unknown parameter vector  $\theta$ , which can in most cases be split into two disjoint parts, one for  $\mu_t$  and one for  $H_t$ .<sup>4</sup> A case where this is not true is that of GARCH-in-mean models, where  $\mu_t$  is functionally dependent on  $H_t$ . In this section, we take no account of the conditional mean vector for notational ease. It is usually specified as a function of the past, through a vectorial autoregressive moving average (VARMA) representation for the level of  $y_t$ .

In the following subsections we review different specifications of  $H_t$ . They differ in various aspects. We distinguish three nonmutually exclusive approaches for constructing multivariate GARCH models; (i) direct generalizations of the univariate GARCH model of Bollerslev (1986); (ii) linear combinations of univariate GARCH models; (iii) nonlinear combinations of univariate GARCH models. In the first category we have VEC, BEKK and factor models. Related models like the flexible MGARCH, Riskmetrics, Cholesky and full factor GARCH models are also in this category. In the second category we have (generalized) orthogonal models and latent factor models. The last category contains constant and dynamic conditional correlation models, the general dynamic covariance model and copula-GARCH models. To keep the notational burden low, we present the models in their '(1,1)' form rather than in their general '(p,q)' form.

## 2.1. Generalizations of the Univariate Standard GARCH Model

The models in this category are multivariate extensions of the univariate GARCH model. When we consider VARMA models for the conditional mean of several time series the number of parameters increases rapidly. The same happens for multivariate GARCH models as straightforward extensions of the univariate GARCH model. Furthermore, since  $H_t$  is a variance matrix, positive definiteness has to be ensured. To make the model tractable for applied purposes, additional structure may be

<sup>4</sup> Note that although the GARCH parameters do not affect the conditional mean, the conditional mean parameters generally enter the conditional variance specification through the residuals.

imposed, for example in the form of factors or diagonal parameter matrices. This class of models lends itself to relatively easy theoretical derivations of stationarity and ergodicity conditions, and unconditional moments (see e.g. He and Teräsvirta, 2002a).

#### VEC and BEKK Models

A general formulation of  $H_t$  has been proposed by Bollerslev *et al.* (1988). In the general VEC model, each element of  $H_t$  is a linear function of the lagged squared errors and cross-products of errors and lagged values of the elements of  $H_t$ .

**Definition 1** The VEC(1, 1) model is defined as:

$$h_t = c + A\eta_{t-1} + Gh_{t-1} \quad (5)$$

where

$$h_t = \text{vech}(H_t) \quad (6)$$

$$\eta_t = \text{vech}(\varepsilon_t \varepsilon_t') \quad (7)$$

and  $\text{vech}(\cdot)$  denotes the operator that stacks the lower triangular portion of a  $N \times N$  matrix as a  $N(N+1)/2 \times 1$  vector.  $A$  and  $G$  are square parameter matrices of order  $(N+1)N/2$  and  $c$  is a  $(N+1)N/2 \times 1$  parameter vector.

The number of parameters is  $N(N+1)(N(N+1)+1)/2$  (e.g. for  $N=3$  it is equal to 78), which implies that in practice this model is used only in the bivariate case. To overcome this problem some simplifying assumptions have to be imposed. Bollerslev *et al.* (1988) suggest the diagonal VEC (DVEC) model in which the  $A$  and  $G$  matrices are assumed to be diagonal, each element  $h_{ijt}$  depending only on its own lag and on the previous value of  $\varepsilon_{it}\varepsilon_{jt}$ . This restriction reduces the number of parameters to  $N(N+5)/2$  (e.g. for  $N=3$  it is equal to 12). But even under this diagonality assumption, large-scale systems are still highly parameterized and difficult to estimate in practice.

Necessary and sufficient conditions on the parameters to ensure that the conditional variance matrices in the DVEC model are positive definite almost surely are most easily derived by expressing the model in terms of Hadamard products (denoted by  $\odot$ ).<sup>5</sup> In particular, let us define the symmetric  $N \times N$  matrices  $A^\circ$ ,  $G^\circ$  and  $C^\circ$  as the matrices implied by the relations  $A = \text{diag}[\text{vech}(A^\circ)]$ ,<sup>6</sup>  $G = \text{diag}[\text{vech}(G^\circ)]$  and  $c = \text{vech}(C^\circ)$ . The diagonal model can thus be written as follows:

$$H_t = C^\circ + A^\circ \odot (\varepsilon_{t-1} \varepsilon_{t-1}') + G^\circ \odot H_{t-1} \quad (8)$$

It is straightforward to show (see Attanasio, 1991) that  $H_t$  is positive definite for all  $t$  provided that  $C^\circ$ ,  $A^\circ$ ,  $G^\circ$  and the initial variance matrix ( $H_0$ ) are positive definite. Moreover, these conditions are easily imposed through a Cholesky decomposition of the parameter matrices in (8). Note that even simpler versions of the DVEC model constrain the  $A^\circ$  and  $G^\circ$  matrices to be rank one

<sup>5</sup> If  $A = (a_{ij})$  and  $B = (b_{ij})$  are both  $m \times n$  matrices, then  $A \odot B$  is the  $m \times n$  matrix containing elementwise products  $(a_{ij}b_{ij})$ .

<sup>6</sup> If  $v$  is a vector of dimension  $m$  then  $\text{diag}(v)$  is the  $m \times m$  diagonal matrix with  $v$  on the main diagonal.

matrices, or a positive scalar times a matrix of ones, also called a scalar model (see Ding and Engle, 2001).

Riskmetrics (1996) uses the exponentially weighted moving average model (EWMA) to forecast variances and covariances. Practitioners who study volatility processes often observe that their model is very close to the unit root case. To take this into account, Riskmetrics defines the variances and covariances as IGARCH-type models (Engle and Bollerslev, 1986):

$$h_{ij,t} = (1 - \lambda)\varepsilon_{i,t-1}\varepsilon_{j,t-1} + \lambda h_{ij,t-1} \quad (9)$$

In terms of the VEC model in (5) we have

$$h_t = (1 - \lambda)\eta_{t-1} + \lambda h_{t-1} \quad (10)$$

which is a scalar VEC model. The decay factor  $\lambda$  proposed by Riskmetrics is equal to 0.94 for daily data and 0.97 for monthly data. The decay factor is not estimated but suggested by Riskmetrics. In this respect, the model is easy to work with in practice. However, imposing the same dynamics on every component in a multivariate GARCH model, no matter which data are used, is difficult to justify.

Because it is difficult to guarantee the positivity of  $H_t$  in the VEC representation without imposing strong restrictions on the parameters,<sup>7</sup> Engle and Kroner (1995) propose a new parametrization for  $H_t$  that easily imposes its positivity, i.e. the BEKK model (the acronym comes from synthesized work on multivariate models by Baba, Engle, Kraft and Kroner).

**Definition 2** The BEKK(1, 1, K) model is defined as:

$$H_t = C^{*'}C^* + \sum_{k=1}^K A_k^{*'}\varepsilon_{t-1}\varepsilon_{t-1}'A_k^* + \sum_{k=1}^K G_k^{*'}H_{t-1}G_k^* \quad (11)$$

where  $C^*$ ,  $A_k^*$  and  $G_k^*$  are  $N \times N$  matrices but  $C^*$  is upper triangular.

The summation limit  $K$  determines the generality of the process. The parameters of the BEKK model do not represent directly the impact of the different lagged terms on the elements of  $H_t$ , like in the VEC model. The BEKK model is a special case of the VEC model. We refer to Engle and Kroner (1995) for propositions and proofs about VEC and BEKK models. For example, to avoid observationally equivalent structures they provide sufficient conditions to identify BEKK models with  $K = 1$ . These conditions are that  $A_{k,11}^*$ ,  $G_{k,11}^*$  and the diagonal elements of  $C^*$  are restricted to be positive.

The number of parameters in the BEKK(1,1,1) model is  $N(5N + 1)/2$ . To reduce this number, and consequently to reduce the generality, one can impose a diagonal BEKK model, i.e.  $A_k^*$  and  $G_k^*$  in (11) are diagonal matrices. This model is also a DVEC model but it is less general, although it is guaranteed to be positive definite while the DVEC is not. This can again be easily checked in the bivariate model: the DVEC model contains 9 parameters while the BEKK model contains only 7 parameters. This happens because the parameters governing the dynamics of the covariance equation in the BEKK model are the products of the corresponding parameters of the two variance

<sup>7</sup> Gouriéroux (1997, section 6.1) gives sufficient conditions for the positivity of  $H_t$ . These conditions are obtained by writing the model for the matrix  $H_t$  itself rather than for its vectorized version.

equations in the same model. Another way to reduce the number of parameters is to use a scalar BEKK model, i.e.  $A_k^*$  and  $G_k^*$  are equal to a scalar times a matrix of ones.

For the VEC model in Definition 1 to be covariance-stationary it is required that the eigenvalues of  $A + G$  are less than one in modulus. The unconditional variance matrix  $\Sigma$ , equal to  $E(H_t)$ , is given by  $\text{vech}(\Sigma) = [I_{N^*} - A - G]_c^{-1}$ , where  $N^* = N(N + 1)/2$ . Similar expressions can be obtained for the BEKK model. Hafner (2003) provides analytical expressions of the fourth-order moments of the general VEC model; see also Nijman and Sentana (1996).

Besides the BEKK model, another option to guarantee the positivity of  $H_t$  in the VEC representation is given by Kawakatsu (2003), who proposes the Cholesky factor GARCH model. Instead of specifying a functional form for  $H_t$ , he specifies a model on  $L_t$  where  $H_t = L_t L_t'$ . The advantage of this specification is that  $H_t$  is always positive definite without any restrictions on the parameters. The disadvantage is that identification restrictions are needed, which implies that the order of the series in  $y_t$  is relevant and that the interpretation of the parameters is difficult. A similar model based on the Cholesky decomposition can be found in Gallant and Tauchen (2001) and in Tsay (2002).

The difficulty when estimating a VEC or even a BEKK model is the high number of unknown parameters, even after imposing several restrictions. It is thus not surprising that these models are rarely used when the number of series is larger than 3 or 4. Factor and orthogonal models circumvent this difficulty by imposing a common dynamic structure on all the elements of  $H_t$ , which results in less parameterized models.

#### Factor Models

Engle *et al.* (1990b) propose a parameterization of  $H_t$  using the idea that co-movements of the stock returns are driven by a small number of common underlying variables, which are called factors. Bollerslev and Engle (1993) use this parametrization to model common persistence in conditional variances. The factor model can be seen as a particular BEKK model. We take the definition of Lin (1992).

**Definition 3** The BEKK(1, 1, K) model in Definition 2 is a factor GARCH model, denoted by F-GARCH(1, 1, K), if for each  $k = 1, \dots, K$ ,  $A_k^*$  and  $G_k^*$  have rank one and have the same left and right eigenvectors,<sup>8</sup>  $\lambda_k$  and  $w_k$ , i.e.

$$A_k^* = \alpha_k w_k \lambda_k' \quad \text{and} \quad G_k^* = \beta_k w_k \lambda_k', \quad (12)$$

where  $\alpha_k$  and  $\beta_k$  are scalars, and  $\lambda_k$  and  $w_k$  (for  $k = 1, \dots, K$ ) are  $N \times 1$  vectors satisfying

$$w_k' \lambda_i = \begin{cases} 0 & \text{for } k \neq i \\ 1 & \text{for } k = i \end{cases} \quad (13)$$

$$\sum_{n=1}^N w_{kn} = 1 \quad (14)$$

<sup>8</sup> A  $1 \times m$  vector  $v \neq 0$  satisfying  $vA = \lambda v$  for an  $m \times m$  matrix  $A$  and a complex number  $\lambda$  is a left eigenvector of  $A$  corresponding to the eigenvalue  $\lambda$ , see Lütkepohl (1996, p. 256).

If we substitute (12) and (13) into (11) and define  $\Omega = C^{*'}C^*$ , we get

$$H_t = \Omega + \sum_{k=1}^K \lambda_k \lambda_k' (\alpha_k^2 w_k' \varepsilon_{t-1} \varepsilon_{t-1}' w_k + \beta_k^2 w_k' H_{t-1} w_k) \quad (15)$$

Restriction (14) is an identification restriction. The  $K$ -factor GARCH model implies that the time-varying part of  $H_t$  has reduced rank  $K$ , but  $H_t$  remains of full rank because  $\Omega$  is assumed positive definite. The vector  $\lambda_k$  and the scalar  $w_k' \varepsilon_t$  (denoted by  $f_{kt}$  hereafter) are also called the  $k$ th factor loading and the  $k$ th factor, respectively. The number of parameters in the F-GARCH(1, 1, 1) is  $N(N+5)/2$ . In (15), the expression between brackets can be replaced by other univariate GARCH specifications.

For example, the conditional variance matrix of the F-GARCH (1, 1, 2) model is:

$$H_t = \Omega + \lambda_1 \lambda_1' [\alpha_1^2 w_1' \varepsilon_{t-1} \varepsilon_{t-1}' w_1 + \beta_1^2 w_1' H_{t-1} w_1] + \lambda_2 \lambda_2' [\alpha_2^2 w_2' \varepsilon_{t-1} \varepsilon_{t-1}' w_2 + \beta_2^2 w_2' H_{t-1} w_2] \quad (16)$$

where the parameter vectors  $\lambda_k = (\lambda_{k1}, \lambda_{k2}, \dots, \lambda_{kN})'$  and  $w_k$  are of dimension  $N \times 1$  while  $\alpha_k^2$  and  $\beta_k^2$  are scalar parameters. Denoting  $\sigma_{k,t}^2 = w_k' H_t w_k$  we can write (16) in a more familiar way as:

$$h_{ijt} = \tau_{ij} + \lambda_{1i} \lambda_{1j} \sigma_{1t} + \lambda_{2i} \lambda_{2j} \sigma_{2t} \quad \forall i, j = 1, \dots, N \quad (17)$$

$$\sigma_{k,t}^2 = \omega_k + \alpha_k^2 f_{k,t-1}^2 + \beta_k^2 \sigma_{k,t-1}^2 \quad k = 1, 2 \quad (18)$$

where  $\tau_{ij} = \omega_{ij} - \lambda_{1i} \lambda_{1j} \omega_1 - \lambda_{2i} \lambda_{2j} \omega_2$  and  $\omega_k = w_k' \Omega w_k$ . Hence  $\sigma_{kt}$  is defined as a univariate GARCH(1, 1) model. The persistence of the conditional variance in (16) is measured by  $\sum_{k=1}^2 (\alpha_k^2 + \beta_k^2)$  and can also be interpreted as common persistence. In other words, the dynamics of the elements of  $H_t$  is the same. We can write  $H_t$  as:

$$H_t = \Omega^* + \lambda_1 \lambda_1' \sigma_{1,t}^2 + \lambda_2 \lambda_2' \sigma_{2,t}^2 \quad (19)$$

where  $\Omega^* = \Omega - \lambda_1 \lambda_1' \omega_1 - \lambda_2 \lambda_2' \omega_2$ . Note that  $E_{t-1}(f_{1t} f_{2t}) = \omega_1' \Omega w_2$  because  $\omega_k' \lambda_l = 0$  for  $k \neq l$ , see (13). This implies that in the case of more than one factor we have the result that any pair of factors has a time-invariant conditional covariance.

Alternatively, the two-factor model described in (19) can be obtained from

$$\varepsilon_t = \lambda_1 f_{1t} + \lambda_2 f_{2t} + e_t \quad (20)$$

where  $e_t$  represents an idiosyncratic shock with constant variance matrix and uncorrelated with the two factors. Each factor  $f_{kt}$  has zero conditional mean and conditional variance like a GARCH(1,1) process, see (18). The  $K$ -factor model can be written as

$$\varepsilon_t = \Lambda f_t + e_t \quad (21)$$

where  $\Lambda$  is a matrix of dimension  $N \times K$  and  $f_t$  is a  $K \times 1$  vector. A factor is observable if it is specified as a function of  $\varepsilon_t$ , like in (16). See Section 2.6 for a brief discussion of latent factor models.

Several variants of the factor model are proposed in the literature. For example, Vrontos *et al.* (2003) introduce the full-factor multivariate GARCH model.



**Definition 4** The FF-MGARCH model is defined as

$$H_t = W \Sigma_t W' \quad (22)$$

where  $W$  is a  $N \times N$  triangular parameter matrix with ones on the diagonal and the matrix  $\Sigma_t = \text{diag}(\sigma_{1,t}^2, \dots, \sigma_{N,t}^2)$  where  $\sigma_{i,t}^2$  is the conditional variance of the  $i$ th factor, i.e. the  $i$ th element of  $W^{-1} \varepsilon_t$ , which can be separately defined as any univariate GARCH model.

By construction,  $H_t$  is always positive definite. Note that  $H_t$  has a structure that depends on the ordering of the time series in  $y_t$ , because of the triangular structure of  $W$ . The restriction of having ones on the diagonal of  $W$  avoids superfluous parameters if each  $\sigma_{i,t}^2$  has a free constant term.

Rigobon and Sack (2003) start from a system of simultaneous equations in structural form where the conditional variances of the innovations are jointly specified. By deriving the reduced form model one obtains innovations with a conditional variance matrix that can be compared with other unrestricted reduced-form MGARCH models. The structural model imposes a number of restrictions on the functional form of the conditional variance of the reduced-form innovations, resulting in less parameters than in a VEC model.

## 2.2. Linear Combinations of Univariate GARCH Models

In this category, we consider models, like orthogonal models and latent factor models (briefly discussed in Section 2.6), that are linear combinations of several univariate models, each of which is not necessarily a standard GARCH (e.g. the EGARCH model of Nelson (1991), the APARCH model of Ding *et al.* (1993), the fractionally integrated GARCH of Baillie *et al.* (1996), the contemporaneous asymmetric GARCH model of El Babsiri and Zakoian (2001) or the quadratic ARCH model of Sentana (1995)).

In the orthogonal GARCH model, the observed data are assumed to be generated by an orthogonal transformation of  $N$  (or a smaller number of) univariate GARCH processes. The matrix of the linear transformation is the orthogonal matrix (or a selection) of eigenvectors of the population unconditional covariance matrix of the standardized returns. In the generalized version, this matrix must only be invertible. The orthogonal models can also be considered as factor models, where the factors are univariate GARCH-type processes.

In the orthogonal GARCH model of Kariya (1988) and Alexander and Chibumba (1997), the  $N \times N$  time-varying variance matrix  $H_t$  is generated by  $m \leq N$  univariate GARCH models.

**Definition 5** The O-GARCH(1,1,  $m$ ) model is defined as:

$$V^{-1/2} \varepsilon_t = u_t = \Lambda_m f_t \quad (23)$$

where  $V = \text{diag}(v_1, v_2, \dots, v_N)$ , with  $v_i$  the population variance of  $\varepsilon_{it}$ , and  $\Lambda_m$  is a matrix of dimension  $N \times m$  given by:

$$\Lambda_m = P_m \text{diag}(l_1^{1/2} \dots l_m^{1/2}) \quad (24)$$

$l_1 \geq \dots \geq l_m > 0$  being the  $m$  largest eigenvalues of the population correlation matrix of  $u_t$ , and  $P_m$  the  $N \times m$  matrix of associated (mutually orthogonal) eigenvectors. The vector  $f_t = (f_{1t} \dots f_{mt})'$



is a random process such that:

$$E_{t-1}(f_t) = 0 \quad \text{and} \quad \text{Var}_{t-1}(f_t) = \Sigma_t = \text{diag}(\sigma_{f1t}^2, \dots, \sigma_{fmt}^2) \quad (25)$$

$$\sigma_{f_{it}}^2 = (1 - \alpha_i - \beta_i) + \alpha_i f_{i,t-1}^2 + \beta_i \sigma_{f_{i,t-1}}^2 \quad i = 1, \dots, m \quad (26)$$

Consequently,

$$H_t = \text{Var}_{t-1}(\varepsilon_t) = V^{1/2} V_t V^{1/2} \quad \text{where} \quad V_t = \text{Var}_{t-1}(u_t) = \Lambda_m \Sigma_t \Lambda_m' \quad (27)$$

The parameters of the model are  $V$ ,  $\Lambda_m$  and the parameters of the GARCH factors ( $\alpha_i$ 's and  $\beta_i$ 's). The number of parameters is  $N(N+5)/2$  (if  $m = N$ ). In practice,  $V$  and  $\Lambda_m$  are replaced by their sample counterparts, and  $m$  is chosen by principal component analysis applied to the standardized residuals  $\hat{u}_t$ . Alexander (2001, section 7.4.3) illustrates the use of the O-GARCH model. She emphasizes that using a small number of principal components compared to the number of assets is the strength of the approach (in one example, she fixes  $m$  at 2 for 12 assets). However, note that the conditional variance matrix has reduced rank (if  $m < N$ ), which may be a problem for applications and for diagnostic tests which depend on the inverse of  $H_t$ .

In van der Weide (2002) the orthogonality condition assumed in the O-GARCH model is relaxed by assuming that the matrix  $\Lambda$  in the relation  $u_t = \Lambda f_t$  is square and invertible, rather than orthogonal. The matrix  $\Lambda$  has  $N^2$  parameters and is not restricted to be triangular like in the model of Vrontos *et al.* (2003), see Definition 4.

**Definition 6** The GO-GARCH(1, 1) model is defined as in Definition 5, where  $m = N$  and  $\Lambda$  is a nonsingular matrix of parameters. The implied conditional correlation matrix of  $\varepsilon_t$  can be expressed as:

$$R_t = J_t^{-1} V_t J_t^{-1} \quad \text{where} \quad J_t = (V_t \odot I_m)^{1/2} \text{ and } V_t = \Lambda \Sigma_t \Lambda' \quad (28)$$

In van der Weide (2002), the singular value decomposition of the matrix  $\Lambda$  is used as a parametrization, i.e.  $\Lambda = PL^{1/2}U$ , where the matrix  $U$  is orthogonal, and  $P$  and  $L$  are defined as above (from the eigenvectors and eigenvalues). The O-GARCH model (when  $m = N$ ) corresponds then to the particular choice  $U = I_N$ . More generally, van der Weide expresses  $U$  as the product of  $N(N-1)/2$  rotation matrices:

$$U = \prod_{i < j} G_{ij}(\delta_{ij}) \quad -\pi \leq \delta_{ij} \leq \pi, \quad i, j = 1, 2, \dots, n \quad (29)$$

where  $G_{ij}(\delta_{ij})$  performs a rotation in the plane spanned by the  $i$ th and  $j$ th vectors of the canonical basis of  $\mathbb{R}^N$  over an angle  $\delta_{ij}$ . For example, in the trivariate case

$$G_{12} = \begin{pmatrix} \cos \delta_{12} & \sin \delta_{12} & 0 \\ -\sin \delta_{12} & \cos \delta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad G_{13} = \begin{pmatrix} \cos \delta_{13} & 0 & -\sin \delta_{13} \\ 0 & 1 & 0 \\ -\sin \delta_{13} & 0 & \cos \delta_{13} \end{pmatrix} \quad (30)$$

and  $G_{23}$  has the block with  $\cos \delta_{23}$  and  $\sin \delta_{23}$  functions in the lower right corner. The  $N(N-1)/2$  rotation angles are parameters to be estimated.

For estimation, van der Weide (2002) replaces in a first step  $P$  and  $L$  by their sample counterparts and the remaining parameters (those of  $U$ ) are estimated together with the parameters of the GARCH factors in a second step. Note that such a two-step estimation method is not applicable if an MGARCH-in-mean effect is included (this is also the case for the O-GARCH model). More generally, as pointed out by a referee, the elements in the matrix  $\Lambda$  could be estimated together with the GARCH parameters of the factors, in a single step.

The orthogonal models are particular F-GARCH models and thus are nested in the BEKK model. As a consequence, their properties follow from those of the BEKK model. In particular, it is obvious that the (G)O-GARCH model is covariance-stationary if the  $m$  univariate GARCH processes are themselves stationary.

### 2.3. Nonlinear Combinations of Univariate GARCH Models

This section collects models that may be viewed as nonlinear combinations of univariate GARCH models. This allows for models where one can specify separately, on the one hand, the individual conditional variances, and on the other hand, the conditional correlation matrix or another measure of dependence between the individual series (like the copula of the conditional joint density). For models of this category, theoretical results on stationarity, ergodicity and moments may not be so straightforward to obtain as for models presented in the preceding sections. Nevertheless, they are less greedy in parameters than the models of the first category, and therefore they are more easily estimable.

#### *Conditional Correlation Models*

The conditional variance matrix for this class of models is specified in a hierarchical way. First, one chooses a GARCH-type model for each conditional variance. For example, some conditional variances may follow a conventional GARCH model while others may be described as an EGARCH model. Second, based on the conditional variances one models the conditional correlation matrix (imposing its positive definiteness  $\forall t$ ).

Bollerslev (1990) proposes a class of MGARCH models in which the conditional correlations are constant and thus the conditional covariances are proportional to the product of the corresponding conditional standard deviations. This restriction greatly reduces the number of unknown parameters and thus simplifies the estimation.

**Definition 7** *The CCC model is defined as:*

$$H_t = D_t R D_t = (\rho_{ij} \sqrt{h_{iit} h_{jjt}}) \quad (31)$$

where

$$D_t = \text{diag}(h_{11t}^{1/2} \dots h_{NNt}^{1/2}) \quad (32)$$

$h_{iit}$  can be defined as any univariate GARCH model, and

$$R = (\rho_{ij}) \quad (33)$$

is a symmetric positive definite matrix with  $\rho_{ii} = 1, \forall i$ .

$R$  is the matrix containing the constant conditional correlations  $\rho_{ij}$ . The original CCC model has a GARCH(1, 1) specification for each conditional variance in  $D_t$ :

$$h_{iit} = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{iit-1} \quad i = 1, \dots, N \quad (34)$$

This CCC model contains  $N(N+5)/2$  parameters.  $H_t$  is positive definite if and only if all the  $N$  conditional variances are positive and  $R$  is positive definite. The unconditional variances are easily obtained, as in the univariate case, but the unconditional covariances are difficult to calculate because of the nonlinearity in (31). He and Teräsvirta (2002b) use a VEC-type formulation for  $(h_{11t}, h_{22t}, \dots, h_{NNt})'$ , to allow for interactions between the conditional variances. They call this the extended CCC model.

The assumption that the conditional correlations are constant may seem unrealistic in many empirical applications. Christodoulakis and Satchell (2002), Engle (2002) and Tse and Tsui (2002) propose a generalization of the CCC model by making the conditional correlation matrix time-dependent. The model is then called a dynamic conditional correlation (DCC) model. An additional difficulty is that the time-dependent conditional correlation matrix has to be positive definite  $\forall t$ . The DCC models guarantee this under simple conditions on the parameters.

The DCC model of Christodoulakis and Satchell (2002) uses the Fisher transformation of the correlation coefficient. The specification of the correlation coefficient is  $\rho_{12,t} = (e^{2r_t} - 1)/(e^{2r_t} + 1)$ , where  $r_t$  can be defined as any GARCH model using  $\varepsilon_{1t}\varepsilon_{2t}/\sqrt{h_{11t}h_{22t}}$  as innovation. This model is easy to implement because the positive definiteness of the conditional correlation matrix is guaranteed by the Fisher transformation. However, it is only a bivariate model. The DCC models of Tse and Tsui (2002) and Engle (2002) are genuinely multivariate and are useful when modelling high-dimensional data sets.

**Definition 8** The DCC model of Tse and Tsui (2002) or  $DCC_T(M)$  is defined as:

$$H_t = D_t R_t D_t \quad (35)$$

where  $D_t$  is defined in (32),  $h_{iit}$  can be defined as any univariate GARCH model and

$$R_t = (1 - \theta_1 - \theta_2)R + \theta_1 \Psi_{t-1} + \theta_2 R_{t-1} \quad (36)$$

In (36),  $\theta_1$  and  $\theta_2$  are non-negative parameters satisfying  $\theta_1 + \theta_2 < 1$ ,  $R$  is a symmetric  $N \times N$  positive definite parameter matrix with  $\rho_{ii} = 1$  and  $\Psi_{t-1}$  is the  $N \times N$  correlation matrix of  $\varepsilon_\tau$  for  $\tau = t - M, t - M + 1, \dots, t - 1$ . Its  $i, j$ th element is given by:

$$\psi_{ij,t-1} = \frac{\sum_{m=1}^M u_{i,t-m} u_{j,t-m}}{\sqrt{\left(\sum_{m=1}^M u_{i,t-m}^2\right) \left(\sum_{m=1}^M u_{j,t-m}^2\right)}} \quad (37)$$

where  $u_{it} = \varepsilon_{it}/\sqrt{h_{iit}}$ . The matrix  $\Psi_{t-1}$  can be expressed as:

$$\Psi_{t-1} = B_{t-1}^{-1} L_{t-1} L_{t-1}' B_{t-1}^{-1} \quad (38)$$

where  $B_{t-1}$  is a  $N \times N$  diagonal matrix with  $i$ th diagonal element given by  $(\sum_{h=1}^M u_{i,t-h}^2)^{1/2}$  and  $L_{t-1} = (u_{t-1}, \dots, u_{t-M})$  is a  $N \times M$  matrix, with  $u_t = (u_{1t} u_{2t} \dots u_{Nt})'$ .

A necessary condition to ensure the positivity of  $\Psi_{t-1}$ , and therefore also of  $R_t$ , is that  $M \geq N$ .<sup>9</sup> Then  $R_t$  is itself a correlation matrix if  $R_{t-1}$  is also a correlation matrix (notice that  $\rho_{iit} = 1 \forall i$ ).

Alternatively, Engle (2002) proposes a different DCC model (see also Engle and Shephard, 2001).

**Definition 9** The DCC model of Engle (2002) or  $DCC_E(1, 1)$  is defined as in (35) with

$$R_t = \text{diag}(q_{11,t}^{-1/2} \dots q_{NN,t}^{-1/2}) Q_t \text{diag}(q_{11,t}^{-1/2} \dots q_{NN,t}^{-1/2}) \quad (39)$$

where the  $N \times N$  symmetric positive definite matrix  $Q_t = (q_{ij,t})$  is given by:

$$Q_t = (1 - \alpha - \beta) \bar{Q} + \alpha u_{t-1} u_{t-1}' + \beta Q_{t-1} \quad (40)$$

with  $u_t$  as in Definition 8.  $\bar{Q}$  is the  $N \times N$  unconditional variance matrix of  $u_t$ , and  $\alpha$  and  $\beta$  are non-negative scalar parameters satisfying  $\alpha + \beta < 1$ .

The elements of  $\bar{Q}$  can be estimated or alternatively set to their empirical counterpart to render the estimation even simpler (see Section 3). To show more explicitly the difference between  $DCC_T$  and  $DCC_E$ , we write the expression of the correlation coefficient in the bivariate case: for the  $DCC_T(M)$

$$\rho_{12t} = (1 - \theta_1 - \theta_2) \rho_{12} + \theta_2 \rho_{12,t-1} + \theta_1 \frac{\sum_{m=1}^M u_{1,t-m} u_{2,t-m}}{\sqrt{\left( \sum_{m=1}^M u_{1,t-m}^2 \right) \left( \sum_{m=1}^M u_{2,t-m}^2 \right)}} \quad (41)$$

and for the  $DCC_E(1, 1)$

$$\rho_{12t} = \frac{(1 - \alpha - \beta) \bar{q}_{12} + \alpha u_{1,t-1} u_{2,t-1} + \beta q_{12,t-1}}{\sqrt{((1 - \alpha - \beta) \bar{q}_{11} + \alpha u_{1,t-1}^2 + \beta q_{11,t-1})((1 - \alpha - \beta) \bar{q}_{22} + \alpha u_{2,t-1}^2 + \beta q_{22,t-1})}} \quad (42)$$

Unlike in the  $DCC_T$  model, the  $DCC_E$  model does not formulate the conditional correlation as a weighted sum of past correlations. Indeed, the matrix  $Q_t$  is written like a GARCH equation, and then transformed to a correlation matrix. However, for both the  $DCC_T$  and  $DCC_E$  models, one can test  $\theta_1 = \theta_2 = 0$  or  $\alpha = \beta = 0$ , respectively to check whether imposing constant conditional correlations is empirically relevant.

A drawback of the DCC models is that  $\theta_1, \theta_2$  in  $DCC_T$  and  $\alpha, \beta$  in  $DCC_E$  are scalars, so that all the conditional correlations obey the same dynamics. This is necessary to ensure that  $R_t$  is positive definite  $\forall t$  through sufficient conditions on the parameters. If the conditional variances are specified as GARCH(1,1) models then the  $DCC_T$  and  $DCC_E$  models contain  $(N + 1)(N + 4)/2$  parameters.

<sup>9</sup> Note that when  $M = 1$ ,  $\Psi_{t-1}$  is equal to a matrix of ones.

Interestingly, DCC models can be estimated consistently in two steps (see Section 3.2), which makes this approach feasible when  $N$  is high. Of course, when  $N$  is large, the restriction of common dynamics gets tighter, but for large  $N$  the problem of maintaining tractability also gets harder. In this respect, several variants of the DCC model are proposed in the literature. For example, Billio *et al.* (2003) argue that constraining the dynamics of the conditional correlation matrix to be the same for all the correlations is not desirable. To solve this problem, they propose a block-diagonal structure where the dynamics is constrained to be identical only within each block. The price to pay for this additional flexibility is that the block members have to be defined *a priori*, which may be cumbersome in some applications. Pelletier (2003) proposes a model where the conditional correlations follow a switching regime driven by an unobserved Markov chain so that the correlation matrix is constant in each regime but may vary across regimes. Another extension proposed by Engle (2002) consists of changing (40) into

$$Q_t = \bar{Q} \odot (ii' - A - B) + A \odot u_{t-1}u_{t-1}' + B \odot Q_{t-1} \quad (43)$$

where  $i$  is a vector of ones and  $A$  and  $B$  are  $N \times N$  matrices of parameters. This increases the number of parameters considerably, but the matrices  $A$  and  $B$  could be defined to depend on a small number of parameters (e.g.  $A = aa'$ ).

To conclude, DCC models open the door to using flexible GARCH specifications in the variance part. Indeed, as the conditional variances (together with the conditional means) can be estimated using  $N$  univariate models, one can easily extend the DCC-GARCH models to more complex GARCH-type structures (as mentioned at the beginning of Section 2.2). One can also extend the bivariate CCC FIGARCH model of Brunetti and Gilbert (2000) to a model of the DCC family.

#### General Dynamic Covariance Model

A model somewhat different from the previous ones but that nests several of them is the general dynamic covariance (GDC) model proposed by Kroner and Ng (1998). They illustrate that the choice of a multivariate volatility model can lead to substantially different conclusions in an application that involves forecasting dynamic variance matrices. We extend the definition of Kroner and Ng (1998) to cover models with dynamic conditional correlations.

**Definition 10** *The GDC model is defined as:*

$$H_t = D_t R_t D_t + \Phi \odot \Theta_t \quad (44)$$

where

$$\begin{aligned} D_t &= (d_{ijt}), \quad d_{iit} = \sqrt{\theta_{iit}} \quad \forall i, \quad d_{ijt} = 0 \quad \forall i \neq j \\ \Theta_t &= (\theta_{ijt}) \\ R_t &\text{ is specified as } DCC_T(M), \text{ see (36) and (37), or as } DCC_E(1, 1), \text{ see (39) and (40)} \\ \Phi &= (\phi_{ij}), \quad \phi_{ii} = 0 \quad \forall i, \quad \phi_{ij} = \phi_{ji} \\ \theta_{ijt} &= \omega_{ij} + a_i' \varepsilon_{t-1} \varepsilon_{t-1}' a_j + g_i' H_{t-1} g_j \quad \forall i, j \\ a_i, g_i, \quad i &= 1, \dots, N \text{ are } (N \times 1) \text{ vectors of parameters, and } \Omega = (\omega_{ij}) \text{ is positive definite} \\ &\text{and symmetric.} \end{aligned} \quad (45)$$

Elementwise we have:

$$h_{iit} = \theta_{iit} \quad \forall i \text{ and } h_{ijt} = \rho_{ijt} \sqrt{\theta_{iit} \theta_{jjt}} + \phi_{ij} \theta_{ijt} \quad \forall i \neq j \quad (46)$$

where the  $\theta_{ijt}$  are given by the BEKK formulation in (45). The GDC model contains several MGARCH models as special cases. To show this we adapt a proposition from Kroner and Ng (1998). Consider the following set of conditions:

- (ia)  $\theta_1 = \theta_2 = 0$  (DCC<sub>T</sub>) or  $\alpha = \beta = 0$  (DCC<sub>E</sub>);
- (ib)  $R = I_N$  (DCC<sub>T</sub>) or  $\bar{Q} = I_N$  (DCC<sub>E</sub>);
- (ii)  $a_i = \alpha_i l_i$  and  $g_i = \beta_i l_i \quad \forall i$ , where  $l_i$  is the  $i$ th column of an  $(N \times N)$  identity matrix, and  $\alpha_i$  and  $\beta_i, i = 1, \dots, N$  are scalars;
- (iii)  $\phi_{ij} = 0 \quad \forall i \neq j$ ;
- (iv)  $\phi_{ij} = 1 \quad \forall i \neq j$ ;
- (v)  $A = \alpha(w\lambda')$  and  $G = \beta(w\lambda')$  where  $A = [a_1, \dots, a_N]$  and  $G = [g_1, \dots, g_N]$  are  $N \times N$  matrices,  $w$  and  $\lambda$  are  $N \times 1$  vectors, and  $\alpha$  and  $\beta$  are scalars.

The GDC model reduces to different multivariate GARCH models under different combinations of these conditions. Specifically, the GDC model becomes:

- the DCC<sub>T</sub> or the DCC<sub>E</sub>(1, 1) model with GARCH(1,1) conditional variances under conditions (ii) and (iii);
- the CCC model with GARCH(1,1) conditional variances under conditions (ia), (ii) and (iii);
- a restricted DVEC(1,1) model under conditions (i) and (ii);
- the BEKK(1,1,1) model under conditions (i) and (iv);
- the F-GARCH(1,1,1) model under conditions (i), (iv) and (v).

Condition (ib) serves as an identification restriction for the VEC, BEKK and F-GARCH models. As we can see, the GDC model is an encompassing model. This requires a large number of parameters (i.e.  $[N(7N - 1) + 4]/2$ ). For example, in the bivariate case there are 11 parameters in  $\Theta_t$ , 3 in  $R_t$  and 1 in  $\Phi$ , which makes a total of 15. This is less than for an unrestricted VEC model (21 parameters), but more than for the BEKK model (11 parameters).

### Copula-MGARCH Models

Another approach for modelling the conditional dependence is known as the copula-GARCH model. This approach makes use of the theorem due to Sklar (1959) stating that any  $N$ -dimensional joint distribution function may be decomposed into its  $N$  marginal distributions, and a copula function that completely describes the dependence between the  $N$  variables. See Nelsen (1999) for a comprehensive introduction to copulas. Patton (2000) and Jondeau and Rockinger (2001) have proposed copula-GARCH models. These models are specified by GARCH equations for the conditional variances (possibly with each variance depending on the lag of the other variances and of the other shocks), marginal distributions for each series (e.g.  $t$ -distributions) and a conditional copula function. Both papers highlight the need to allow for time-variation in the conditional copula, extending in some sense the DCC models to other specifications of the conditional dependence. The copula function is rendered time-varying through its parameters, which can be functions of past data. In this respect, like the DCC model of Engle (2002), copula-GARCH

models can be estimated using a two-step maximum likelihood approach (see Section 3.2) which solves the dimensionality problem. An interesting feature of copula-GARCH models is the ease with which very flexible joint distributions may be obtained in the bivariate case. Their application to higher dimensions is a subject for further research.

## 2.4. Leverage Effects in MGARCH Models

For stock returns, negative shocks may have a larger impact on their volatility than positive shocks of the same absolute value (this is most often interpreted as the leverage effect unveiled by Black, 1976). In other words, the news impact curve, which traces the relation between volatility and the previous shock, is asymmetric. Univariate models that allow for this effect are the EGARCH model of Nelson (1991), the GJR model of Glosten *et al.* (1993) and the threshold ARCH model of Zakoian (1994), among others. For multivariate series the same argument applies: the variances and covariances may react differently to a positive than to a negative shock. In the multivariate case, a shock can be defined in terms of  $\varepsilon_t$  or  $z_t$ . Note that the signs of  $\varepsilon_{it}$  and  $z_{it}$  do not necessarily coincide, see (2).<sup>10</sup>

The MGARCH models reviewed in the previous subsections define the conditional variance matrix as a function of lagged values of  $\varepsilon_t \varepsilon_t'$ . For example, each conditional variance in the VEC model is a function of its own squared error but it is also a function of the squared errors of the other series as well as the cross-products of errors. A model that takes explicitly the sign of the errors into account is the asymmetric dynamic covariance (ADC) model of Kroner and Ng (1998). The only difference with Definition 10 is an extra term based on the vector  $v_t = \max[0, -\varepsilon_t]$  in  $\theta_{ijt}$  to take into account the sign of  $\varepsilon_{it}$ :

$$\theta_{ijt} = \omega_{ij} + a_i' \varepsilon_{t-1} \varepsilon_{t-1}' a_j + g_i' H_{t-1} g_j + b_i' v_{t-1} v_{t-1}' b_j \quad (47)$$

The ADC model nests some natural extensions of MGARCH models that incorporate the leverage effect. Kroner and Ng (1998) apply the model to large and small firm returns. They find that bad news about large firms can cause additional volatility in both small-firm and large-firm returns. Furthermore, this bad news increases the conditional covariance. Small firm news has only minimal effects.

Hansson and Hordahl (1998) add the term  $D \odot v_{t-1} v_{t-1}'$ , in a DVEC model like (8), where  $D$  is a diagonal matrix of parameters. To incorporate the leverage effect in the (bivariate) BEKK model, Hafner and Herwartz (1998) add the terms  $D_1' \varepsilon_{t-1} \varepsilon_{t-1}' D_1 D_{\{\varepsilon_{1,t-1} < 0\}} + 1_2' \varepsilon_{t-1} \varepsilon_{t-1}' D_2 D_{\{\varepsilon_{2,t-1} < 0\}}$ , where  $D_1$  and  $D_2$  are  $2 \times 2$  matrices of parameters and  $1_{\{\dots\}}$  is the indicator function. This generalizes the univariate GJR specification.

## 2.5. Transformations of MGARCH Models

Not all MGARCH models are invariant with respect to linear transformations. By invariance of a model, we mean that it stays in the same class if a linear transformation is applied to  $y_t$ , say  $\tilde{y}_t = F y_t$ , where  $F$  is a matrix of constants (for simplicity we assume  $F$  is square). If  $y_t$  is a vector of returns, a linear transformation corresponds to new assets (portfolios combining the original assets). It seems sensible that a model should be invariant, otherwise the question arises which

<sup>10</sup> Remember that  $H_t^{-1/2}$  and hence  $z_t$  is not unique.



basic assets should be modelled. In some cases (stocks), these are naturally defined, in other cases, like exchange rates, they are not, since a reference currency must be chosen (see Gouriéroux and Jasiak, 2001, p. 140). Lack of invariance of a model does not imply that the model is not suitable at all for use in empirical work. Implications of invariance, or lack of invariance, are an open issue. For example, if the model is invariant, one can estimate it with some number of basic assets, as well as with a smaller number of portfolios of the basic assets. Estimates of the larger model imply estimates of the smaller models, which could be compared to the direct estimates of the latter. Very different estimates may lead us to question the specification.

Lack of invariance occurs whenever a diagonal matrix in the equation defining  $H_t$  is premultiplied by the matrix  $F$  defined above. The general VEC and BEKK models are invariant, but their diagonal versions are not. Conditional correlation models are not invariant, since  $FD_t$  is not diagonal when  $D_t$  is diagonal, see (31).

A related question is the marginalization of MGARCH processes: starting from a strong MGARCH model for  $y_t$ , can we characterize the implied marginal process of a subvector of  $y_t$ , in particular of the scalar  $y_{it}$ ? Nijman and Sentana (1996) provide an answer to that question.<sup>11</sup> To take a simple case, for a bivariate VEC(1,1) model, the implied process for  $y_{1t}$  is at most a weak GARCH (3,3) process.<sup>12</sup> In the DVEC(1,1) case, the marginal process of  $y_{1t}$  remains a strong GARCH process. In proving such results, they use the VARMA(1,1) representation of the VEC(1,1) model  $h_t = c + A\eta_{t-1} + G h_{t-1}$ , given by  $\eta_t = c + (A + G)\eta_{t-1} + \omega_t - G\omega_{t-1}$  where  $\omega_t = \eta_t - E_{t-1}(\eta_t)$  is a martingale difference. Hence it is clear that this approach cannot be applied to the conditional correlation models and the GDC model. Marginalization results for the latter models are not known.

Another question is that of temporal aggregation of MGARCH processes. Hafner (2004) shows that, like Drost and Nijman (1993) in the univariate case, the class of weak multivariate GARCH processes is closed under temporal aggregation. Weak multivariate GARCH models are characterized by a weak VARMA structure of  $\eta_t$  in (7). Fourth moment characteristics turn out to be crucial for deriving the low-frequency dynamics. The issue of estimation of the parameters of the low-frequency model is difficult because the probability law of the innovation vector is unknown, since it is only assumed to be a weak white noise. See Hafner and Rombouts (2003) for more details.

## 2.6. Alternative Approaches to Multivariate Volatility

There are at least two other approaches to multivariate volatility than MGARCH models: stochastic volatility (SV) models and realized volatility.

Multivariate stochastic volatility models (see e.g. Harvey *et al.*, 1994) specify that the conditional variance matrix depends on some unobserved or latent processes rather than on past observations. A multivariate SV model is typically specified as  $N$  univariate SV models for the conditional variances (see Ghysels *et al.*, 1996, for a survey of SV models):

$$\varepsilon_{it} = \sigma_i z_{it} \exp(0.5 h_{it}) \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (48)$$

<sup>11</sup> Nijman and Sentana (1996) and Meddahi and Renault (1996) study the issue of contemporaneous aggregation, i.e., the aggregation of independent univariate GARCH processes.

<sup>12</sup> In a weak GARCH process, the dynamic equation for  $h_t$  defines the best linear predictor of  $\varepsilon_t^2$  given the past of  $\varepsilon_t$ . In a strong GARCH,  $h_t$  is the conditional variance. See Drost and Nijman (1993).

where  $\sigma_i$  is a parameter. The innovation vector  $z_t = (z_{1t}, \dots, z_{Nt})'$  has  $E(z_t) = 0$  and  $\text{Var}(z_t) = \Sigma_z$ , while the vector of volatilities  $h_t = (h_{1t}, \dots, h_{Nt})'$  follows a VAR(1) process  $h_t = \Phi h_{t-1} + \eta_t$  where  $\eta_t$  is i.i.d.  $\sim N(0, \Sigma_\eta)$ . In this model, the dynamics of the covariances depends on the dynamics of the corresponding conditional variances, in other words, there is no direct specification of changing covariances or correlations. A drawback of SV models is the complexity of estimation.

Because the main emphasis of this survey is on 'data-driven' MGARCH models, a thorough discussion of the vast literature on latent factor models is beyond the scope of this paper. The factor model in (21) becomes a latent model if  $F_t$  is latent, which means that it is not included in  $I_t$ , implying that the conditional variance matrix, see for example (19), is not measurable any more. This is in contrast with Section 2.1, where the conditional variance of the factors is specified as a function of the past data ( $\varepsilon_t$ ). Therefore, latent factor models can be classified as stochastic volatility models as mentioned in Shephard (1996). The elements of  $F_t$  typically follow dynamic heteroscedastic processes, for example Diebold and Nerlove (1989) use ARCH models. The fact that the factor is considered as nonobservable complicates inference considerably, since the likelihood function must be marginalized with respect to it (see Gouriéroux, 1997, section 6.3). The conditional covariance between the factors is usually assumed to be equal to zero. See Sentana and Fiorentini (2001) and Fiorentini *et al.* (2004) for more details on identification and estimation of factor models. Sentana (1998) shows that the observed factor model is observationally equivalent (up to conditional second moments) to a class of conditionally heteroscedastic factor models including latent factor models. Doz and Renault (2003) elaborate on this result and draw the conclusions in terms of model specification and identification, and in terms of inference methodologies.

The second alternative has been proposed by Andersen *et al.* (2003). In this case, a daily measure of variances and covariances is computed as an aggregate measure from intraday returns. More specifically, a daily realized variance for day  $t$  is computed as the sum of the squared intraday equidistant returns for the given trading day and a daily realized covariance is obtained by summing the products of intraday returns. Once such daily measures have been obtained, they can be modelled, e.g. for a prediction purpose. A nice feature of this approach is that unlike MGARCH and multivariate stochastic volatility models, the  $N(N-1)/2$  covariance components of the conditional variance matrix (or, rather, the components of its Choleski decomposition) can be forecasted independently, using as many univariate models. As shown by Andersen *et al.* (2003), although the use of the realized covariance matrix facilitates rigorous measurement of conditional volatility in much higher dimensions than is feasible with MGARCH and multivariate SV models, it does not allow the dimensionality to become arbitrarily large. Indeed, to ensure the positive definiteness of the realized covariance matrix, the number of assets ( $N$ ) cannot exceed the number of intraday returns for each trading day. The main drawback is that intraday data remain relatively costly and are not readily available for all assets. Furthermore, a large amount of data handling and computer programming is usually needed to retrieve the intraday returns from the raw data files supplied by the exchanges or data vendors. On the contrary, working with daily data is relatively simple and the data are broadly available.

Which approach is best, for example in terms of forecasting, is beyond the scope of the paper and an interesting topic for future theoretical and empirical research.

### 3. ESTIMATION

In the previous section we have defined existing specifications of conditional variance matrices that enter the definition either of a data generating process (DGP) or of a model to be estimated. In Section 3.1 we discuss maximum likelihood (ML) estimation of these models, and in Section 3.2 we explain a two-step approach for estimating conditional correlation models. Finally, we review briefly various issues related to practical estimation in Section 3.3.

#### 3.1. Maximum Likelihood

Suppose the vector stochastic process  $\{y_t\}$  (for  $t = 1, \dots, T$ ) is a realization of a DGP whose conditional mean, conditional variance matrix and conditional distribution are respectively  $\mu_t(\theta_0)$ ,  $H_t(\theta_0)$  and  $p(y_t|\zeta_0, I_{t-1})$ , where  $\zeta_0 = (\theta_0 \ \eta_0)$  is a  $r$ -dimensional parameter vector and  $\eta_0$  is the vector that contains the parameters of the distribution of the innovations  $z_t$  (there may be no such parameter). Importantly, to justify the choice of the estimation procedure, we assume that the model to be estimated encompasses the true formulations of  $\mu_t(\theta_0)$  and  $H_t(\theta_0)$ .

The procedure most often used in estimating  $\theta_0$  involves the maximization of a likelihood function constructed under the auxiliary assumption of an i.i.d. distribution for the standardized innovations  $z_t$ . The i.i.d. assumption may be replaced by the weaker assumption that  $z_t$  is a martingale difference sequence with respect to  $I_{t-1}$ , but this type of assumption does not translate into the likelihood function. The likelihood function for the i.i.d. case can then be viewed as a quasi-likelihood function.

Consequently, one has to make an additional assumption on the innovation process by choosing a density function, denoted  $g(z_t(\theta)|\eta)$ , where  $\eta$  is a vector of nuisance parameters. The problem to solve is thus to maximize the sample loglikelihood function  $L_T(\theta, \eta)$  for the  $T$  observations (conditional on some starting values for  $\mu_0$  and  $H_0$ ), with respect to the vector of parameters  $\zeta = (\theta, \eta)$ , where

$$L_T(\zeta) = \sum_{t=1}^T \log f(y_t|\zeta, I_{t-1}) \quad (49)$$

with

$$f(y_t|\zeta, I_{t-1}) = |H_t|^{-1/2} g(H_t^{-1/2}(y_t - \mu_t)|\eta) \quad (50)$$

and the dependence with respect to  $\theta$  occurs through  $\mu_t$  and  $H_t$ . The term  $|H_t|^{-1/2}$  is the Jacobian that arises in the transformation from the innovations to the observables. Note that unless  $g(\cdot)$  belongs to the class of elliptical distributions, i.e. is a function of  $z_t'z_t$ , the ML estimator depends on the choice of decomposition of  $H_t^{1/2}$ , since  $z_t'z_t = (y_t - \mu_t)'H_t^{-1}(y_t - \mu_t)$ .

The most commonly employed distribution in the literature is the multivariate normal, uniquely determined by its first two moments (so that  $\zeta = \theta$  since  $\eta$  is empty). In this case, the sample loglikelihood is (up to a constant):

$$L_T(\theta) = -\frac{1}{2} \sum_{t=1}^T \log |H_t| - \frac{1}{2} \sum_{t=1}^T (y_t - \mu_t)' H_t^{-1} (y_t - \mu_t) \quad (51)$$

It is well known that the normality of the innovations is rejected in most applications dealing with daily or weekly data. In particular, the kurtosis of most financial asset returns is larger than

three, which means that they have too many extreme values to be normally distributed. Moreover, their unconditional distribution often has fatter tails than what is implied by a conditional normal distribution: the increase of the kurtosis coefficient brought by the dynamics of the conditional variance is not usually sufficient to match adequately the unconditional kurtosis of the data.

However, as shown by Bollerslev and Wooldridge (1992), a consistent estimator of  $\theta_0$  may be obtained by maximizing (51) with respect to  $\theta$  even if the DGP is not conditionally Gaussian. This estimator, called (Gaussian) quasi-maximum likelihood (QML) or pseudo-maximum likelihood (PML) estimator, is consistent provided the conditional mean and the conditional variance are specified correctly. Jeantheau (1998) proves the strong consistency of the Gaussian QML estimator of multivariate GARCH models. He also provides sufficient identification conditions for the CCC model. See Gourieroux (1997) for a detailed description of the QML method in an MGARCH context and its asymptotic properties. For these reasons and as far as the purpose of the analysis is to estimate consistently the first two conditional moments, estimating MGARCH models by QML is justified.

Nevertheless, in certain situations it is desirable to search for a better distribution for the innovation process. For instance, when one is interested in obtaining density forecasts (see Diebold *et al.*, 1998, in the univariate case and Diebold *et al.*, 1999, in the multivariate case) it is natural to relax the normality assumption, keeping in mind the risk of inconsistency of the estimator (see Newey and Steigerwald, 1997).

A natural alternative to the multivariate Gaussian density is the Student density, see Harvey *et al.* (1992) and Fiorentini *et al.* (2003). The latter has an extra scalar parameter, the degrees of freedom parameter, denoted  $\nu$  hereafter. When this parameter tends to infinity, the Student density tends to the normal density. When it tends to zero, the tails of the density become thicker and thicker. The parameter value indicates the order of existence of the moments, e.g. if  $\nu = 2$ , the second-order moments do not exist, but the first-order moments exist. For this reason, it is convenient (although not necessary) to assume that  $\nu > 2$ , so that  $H_t$  is always interpretable as a conditional covariance matrix. Under this assumption, the Student density can be defined as:

$$g(z_t|\theta, \nu) = \frac{\Gamma\left(\frac{\nu + N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) [\pi(\nu - 2)]^{\frac{N}{2}}} \left[1 + \frac{z_t' z_t}{\nu - 2}\right]^{-\frac{N + \nu}{2}} \quad (52)$$

where  $\Gamma(\cdot)$  is the Gamma function. Note that in this case  $\eta = \nu$ . The density function of  $y_t$  is easily obtained by applying (50).

The relevance of the Student distribution may be questioned when the innovations are found to be skewed. To account for both the skewness and the excess kurtosis in returns, an MGARCH model can be combined with a multivariate density for the innovations, which is skewed and has fat tails. Densities used in this context are mixtures of multivariate normal densities (see Vlaar and Palm, 1993), the generalized hyperbolic distribution (see Barndorff-Nielsen and Shephard, 2001, for the density and Mencía and Sentana, 2003, for a recent application to an MGARCH context) and a multivariate skew-Student density (see Bauwens and Laurent, 2002). The latter authors show, in applications to several portfolios of stocks and currencies, that the multivariate skew-Student density improves the quality of out-of-sample Value-at-Risk forecasts, by comparison with a symmetric density.

Alternatively, Hafner and Rombouts (2004) propose a semi-parametric estimation technique, extending the previous work of Engle and González-Rivera (1991) and Drost and Klaassen

(1997) to MGARCH models. This consists of first estimating the model by QML, which provides consistent estimates of the innovations. In a second step, these are used to estimate the function  $g(\cdot)$  nonparametrically. Finally, the parameters of the GARCH model are estimated using  $\hat{g}(\cdot)$  to define the likelihood function.

The asymptotic properties of ML and QML estimators in multivariate GARCH models are not yet firmly established, and are difficult to derive from low level assumptions. As mentioned previously, consistency has been shown by Jeantheau (1998). Asymptotic normality of the QMLE is not established generally. Gouriéroux (1997, section 6.3) proves it for a general formulation using high level assumptions. Comte and Lieberman (2003) prove it for the BEKK formulation. Since F-GARCH and (G)O-GARCH models are special cases of the BEKK model, this result holds also for these two models (see van der Weide, 2002). Researchers who use MGARCH models have generally proceeded as if asymptotic normality holds in all cases. Asymptotic normality of the MLE and QMLE has been proved in the univariate case under low level assumptions, one of which is the existence of moments of order four or higher of the innovations (see Lee and Hansen, 1994; Lumsdaine, 1996; Ling and McAleer, 2003). However, Hall and Yao (2003) show that the asymptotic distribution of the QMLE in the univariate GARCH( $p, q$ ) model is not normal, but is a multivariate stable distribution (with fatter tails than the normal) if the innovations are in the domain of attraction of a stable law with exponent smaller than two (implying nonexisting fourth moments). Extension of this result to the multivariate case is a subject for further research.

Finally, it is worth mentioning that the conditional mean parameters may be consistently estimated in a first stage, prior to the estimation of the conditional variance parameters, for example for a VARMA model, but not for a GARCH-in-mean model. Estimating the parameters simultaneously with the conditional variance parameters would increase the efficiency at least in large samples (unless the asymptotic covariance matrix is block diagonal between the mean and variance parameters), but this is computationally more difficult. For this reason, one usually takes either a very simple model for the conditional mean or one considers  $y_t - \hat{\mu}_t$  as the data for fitting the MGARCH model. A detailed investigation of the consequences of such a two-step procedure on properties of estimators has still to be conducted. Conditions for block diagonality of the asymptotic covariance matrix have also to be worked out (generalizing results of Engle, 1982 for the univariate case).

### 3.2. Two-Step Estimation

A useful feature of the DCC models presented earlier is that they can be estimated consistently using a two-step approach. Engle and Sheppard (2001) show that in the case of a DCC<sub>E</sub> model, the loglikelihood can be written as the sum of a mean and volatility part (depending on a set of unknown parameters  $\theta_1^*$ ) and a correlation part (depending on  $\theta_2^*$ ).

Indeed, recalling that the conditional variance matrix of a DCC model can be expressed as  $H_t = D_t R_t D_t$ , an inefficient but consistent estimator of the parameter  $\theta_1^*$  can be found by replacing  $R_t$  by the identity matrix in (51). In this case the quasi-loglikelihood function corresponds to the sum of loglikelihood functions of  $N$  univariate models:

$$QL1_T(\theta_1^*) = -\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^N \left[ \log(h_{iit}) + \frac{(y_{it} - \mu_{it})^2}{h_{iit}} \right] \quad (53)$$

Given  $\theta_1^*$  and under appropriate regularity conditions, a consistent, but inefficient, estimator of  $\theta_2^*$  can be obtained by maximizing:

$$QL2_T(\theta_2^*|\theta_1^*) = -\frac{1}{2} \sum_{t=1}^T (\log |R_t| + u_t' R_t^{-1} u_t) \quad (54)$$

where  $u_t = D_t^{-1}(y_t - \mu_t)$ . The sum of the likelihood functions in (53) and (54), plus half of the total sum of squared standardized residuals ( $\sum_t u_t' u_t / 2$ , which is almost equal to  $NT/2$ ), is equal to the loglikelihood in (51). It is thus possible to compare the loglikelihood of the two-step approach with that of the one-step approach and of other models.

Engle and Sheppard (2001) explain that the estimators  $\hat{\theta}_1^*$  and  $\hat{\theta}_2^*$ , obtained by maximizing (53) and (54) separately, are not fully efficient (even if  $z_t$  is normally distributed) since they are limited information estimators. However, one iteration of a Newton–Raphson algorithm applied to the total likelihood (51), starting at  $(\hat{\theta}_1^*, \hat{\theta}_2^*)$ , provides an estimator that is asymptotically efficient.

Another two-step approach for the diagonal VEC model is proposed by Ledoit *et al.* (2003). To avoid estimating  $c$ ,  $A$  and  $G$  jointly, they estimate each variance and covariance equation separately. The resulting estimates do not necessarily guarantee positive semi-definite  $H_t$ 's. Therefore, in a second step, the estimates are transformed in order to achieve the requirement, keeping the disruptive effects as small as possible. The transformed estimates are still consistent with respect to the parameters of the DVEC model.

### 3.3. Various Issues

#### *Analytical vs. Numerical Score*

Typically, for conditionally heteroscedastic models, numerical techniques are used to approximate the derivatives of the loglikelihood function (the score) with respect to the parameter vector. As shown by Fiorentini *et al.* (1996) and McCullough and Vinod (1999), in a univariate framework, using analytical scores in the estimation procedure improves the numerical accuracy of the resulting estimates and speeds-up ML estimation. According to Hafner and Rombouts (2004), the score vector corresponding to a term of (49) takes the form

$$s_t(\zeta) = \frac{\partial \text{vec}(H_t)'}{\partial \zeta} \left\{ -\frac{1}{2} \text{vec}(H_t^{-1}) - D_N D_N^+ (I_N \otimes H_t + H_t^{1/2} \otimes H_t^{1/2})^{-1} \text{vec} \left( \frac{\partial \log g(z_t)}{\partial z_t} z_t' \right) \right\} \quad (55)$$

where  $s_t(\zeta) = \partial \log f(y_t | \zeta, I_{t-1}) / \partial \zeta$ ,  $\text{vec}(\cdot)$  is the operator that stacks the columns of a  $M \times N$  matrix into a  $MN \times 1$  vector,  $D_N$  is the duplication matrix defined so that  $D_N \text{vech}(A) = \text{vec}(A)$  for every symmetric matrix  $A$  of order  $N$  and  $D_N^+$  is its generalized inverse. As pointed out by a referee, one sees from (55) that the choice of the square root matrix  $H_t^{1/2}$  has consequences for the exact form of the score.

In this respect, Lucchetti (2002) proposes a closed-form expression of the score vector for the BEKK model with a Gaussian loglikelihood. Hafner and Herwartz (2003) also provide analytical formulae for the score and the Hessian of a general MGARCH model in a QML framework and propose two methods to estimate the expectation of the Hessian. The authors show in a simulation study that analytical derivatives clearly outperform numerical methods.



### Variance Targeting

We have seen that what renders most MGARCH models difficult for estimation is their high number of parameters. A simple trick to ensure a reasonable value of the model-implied unconditional covariance matrix, which also helps to reduce the number of parameters in the maximization of the likelihood function, is referred to as variance targeting by Engle and Mezrich (1996). For example, in the VEC model (and all its particular cases), the conditional variance matrix may be expressed in terms of the unconditional variance matrix (see earlier) and other parameters. Doing so one can reparametrize the model using the unconditional variance matrix and replace it by a consistent estimator (before maximizing the likelihood). When doing this, one should correct the covariance matrix of the estimator of the other parameters for the uncertainty in the preliminary estimator. In DCC models, this can also be done with the constant matrix of the correlation part, e.g.  $\bar{Q}$  in (40). In this case, the two-step estimation procedure explained in Section 3.2 becomes a three-step procedure.

### Imposing or Not the Positivity Constraints

A key problem in MGARCH models is that the conditional variance matrix has to be positive definite almost surely for all  $t$ . As shown in the previous section this is done by constraining the parameter space (for instance by using a constrained optimization algorithm), assuming that the constraints are known. However, these constraints are usually sufficient but not necessary. For instance, we know since Nelson and Cao (1992) that imposing  $\omega_i > 0$  and  $\alpha_i, \beta_i \geq 0$  in (34) is overly restrictive and that negative values of  $\alpha_i$  and  $\beta_i$  are not incompatible with a positive conditional variance. If one imposes positivity restrictions to facilitate estimation, one incurs the risk of rejecting  $\theta_0$  from the parameter space.

### Software

Brooks *et al.* (2003) review the relatively small number of software packages that are currently available for estimating MGARCH models. It is obvious that the development of MGARCH models in standard econometric packages is still in its infancy, and that further developments would greatly help applied researchers who cannot afford to program the estimation of a particular model, but who would rather try several models and distributions.

## 4. DIAGNOSTIC CHECKING

Since estimating MGARCH models is time-consuming, both in terms of computations and their programming (if needed), it is desirable to check *ex ante* whether the data present evidence of multivariate ARCH effects. *Ex post*, it is also of crucial importance to check the adequacy of the MGARCH specification. However, compared to the huge body of diagnostic tests devoted to univariate models, few tests are specific to multivariate models.

In the current literature on MGARCH models, one can distinguish two kinds of specification tests, namely univariate tests applied independently to each series and multivariate tests applied to the vector series as a whole. We deliberately leave out the first kind of tests and refer interested readers to surveys of univariate ARCH processes (see Section 1). As emphasized by Kroner and Ng (1998), the existing literature on multivariate diagnostics is sparse compared to the univariate case. However, although univariate tests can provide some guidance, contemporaneous correlation of disturbances entails that statistics from individual equations are not independent. As a result,



combining test decisions over all equations raises size control problems, so the need for joint testing naturally arises (Dufour *et al.*, 2003).

Since the dynamics of the series is assumed to be captured by the model (at least in the first two conditional moments), the standardized error term  $z_t = H_t^{-1/2}\varepsilon_t$  should obey the following moment conditions (see Ding and Engle, 2001):<sup>13</sup>

- (A)  $E(z_t z_t') = I_N$
- (B)  $\text{Cov}(z_{it}^2, z_{jt}^2) = 0$ , for all  $i \neq j$
- (C)  $\text{Cov}(z_{it}^2, z_{j,t-k}^2) = 0$ , for  $k > 0$

While testing A has power to detect misspecification in the conditional mean, testing B is suited to check if the conditional distribution is Gaussian, which could be false even if  $H_t$  is correctly specified. In contrast, testing C aims at checking the adequacy of the dynamic specification of  $H_t$ , regardless of the validity of the assumption about the distribution of  $z_t$ . Ding and Engle (2001) show that if the true conditional distribution is the multivariate Student described in (52),  $\text{Cov}(z_{it}^2, z_{jt}^2) = 2v^2/[(v-4)(v-2)^2]$ , for  $i \neq j$ , which is different from 0 when  $1/v \neq 0$  (the non-Gaussian case). Moreover, starting from a conditionally homoscedastic multivariate regression model (i.e.  $H_t = H, \forall t$ ), testing C is equivalent to testing the presence of ARCH effects in the data. Provided that a sufficient number of moments exist (which is not always the case), testing conditions A–C could be done using the conditional moment test principle of Newey (1985) and Tauchen (1985).

A quite different approach aims at checking the overall adequacy of a model, i.e. the coincidence of the assumed density  $f(y_t|\theta, \eta, I_{t-1})$  and the true density  $p(y_t|\theta_0, \eta_0, I_{t-1})(\forall t)$ . Diebold *et al.* (1998) (in the univariate case) and Diebold *et al.* (1999) (in the multivariate case) propose an elegant and practical procedure based on the concept of density forecasts. For more details about density forecasts and their applications in finance, see the special issue of the *Journal of Forecasting* (Timmermann, 2000).

As mentioned by Tse (2002), diagnostics for conditional heteroscedasticity models applied in the literature can be divided into three categories: portmanteau tests of the Box–Pierce–Ljung type, residual-based diagnostics and Lagrange multiplier tests.

#### 4.1. Portmanteau Statistics

The most widely used diagnostics to detect ARCH effects are probably the Box–Pierce/Ljung–Box portmanteau tests. Following Hosking (1980), a multivariate version of the Ljung–Box test statistic is given by:

$$HM(M) = T^2 \sum_{j=1}^M (T-j)^{-1} \text{tr}\{C_{Y_t}^{-1}(0)C_{Y_t}(j)C_{Y_t}^{-1}(0)C_{Y_t}'(j)\} \quad (56)$$

where  $Y_t = \text{vech}(y_t y_t')$  and  $C_{Y_t}(j)$  is the sample autocovariance matrix of order  $j$ . Under the null hypothesis of no ARCH effects,  $HM(M)$  is distributed asymptotically as  $\chi^2(K^2 M)$ . Duchesne and Lalancette (2003) generalize this statistic using a spectral approach and obtain higher asymptotic

<sup>13</sup> The definition of the exact form of the square root matrix  $H_t^{1/2}$  is deliberately left unspecified by Ding and Engle (2001). It is not known to what extent a particular choice has a consequence on the tests presented in this section.

power by using a different kernel than the truncated uniform kernel used in  $HM(M)$ . This test is also used to detect misspecification in the conditional variance matrix  $H_t$ , by replacing  $y_t$  by  $\hat{z}_t = \hat{H}_t^{-1/2}\hat{\varepsilon}_t$ . The asymptotic distribution of the portmanteau statistics is, however, unknown in this case since  $\hat{z}_t$  has been estimated. Furthermore, ad hoc adjustments of degrees of freedom for the number of estimated parameters have no theoretical justification. In such a case, portmanteau tests should be interpreted with care even if simulation results reported by Tse and Tsui (1999) suggest that they provide a useful diagnostic in many situations.

Ling and Li (1997) propose an alternative portmanteau statistic for multivariate conditional heteroscedasticity. They define the sample lag- $h$  (transformed) residual autocorrelation as:

$$\tilde{R}(h) = \frac{\sum_{t=h+1}^T (\hat{\varepsilon}'_t \hat{H}_t^{-1} \hat{\varepsilon}_t - N)(\hat{\varepsilon}'_{t-h} \hat{H}_{t-h}^{-1} \hat{\varepsilon}_{t-h} - N)}{\sum_{t=h+1}^T (\hat{\varepsilon}'_t \hat{H}_t^{-1} \hat{\varepsilon}_t - N)^2} \quad (57)$$

Their test statistic is given by  $LL(M) = T \sum_{h=1}^M \tilde{R}^2(h)$  and is asymptotically distributed as  $\chi^2(M)$  under the null of no conditional heteroscedasticity. In the derivation of the asymptotic results, normality of the innovation process is not assumed. The statistic is thus robust with regard to the distribution choice. Tse and Tsui (1999) show that there is a loss of information in the transformation  $\hat{\varepsilon}'_t \hat{H}_t^{-1} \hat{\varepsilon}_t$  of the residuals and the test may suffer from a power reduction. Furthermore, Duchesne and Lalancette (2003) argue that if an inappropriate choice of  $M$  is selected, the resulting test statistic may be quite inefficient (the same comment applies to the residual-based tests presented below). For these reasons, these authors propose a more powerful version of the  $LL(M)$  test based on the spectral density of the stochastic process  $\{\varepsilon'_t H_t^{-1} \varepsilon_t, t \in \mathbb{Z}\}$ , which is i.i.d. under the null of homoscedasticity. Interestingly, since their test is based on a spectral density estimator, a data-dependent choice of  $M$  is available.

## 4.2. Residual-Based Diagnostics

These tests involve running regressions of the cross-products of the standardized residuals ( $\hat{u}_t$ ) on some explanatory variables and testing for the statistical significance of the regression coefficients. The key problem is that since the regressors (transformed residuals) are obtained after estimating a first model and so depend on estimated parameters with their own uncertainty, the usual OLS theory does not apply. The contribution of Tse (2002) is to establish the asymptotic distribution of the OLS estimator in this context. Let us define  $\hat{u}_{it} = \hat{\varepsilon}_{it} / \sqrt{\hat{h}_{iit}}$  as the  $i$ th ( $i = 1, \dots, N$ ) standardized residual at time  $t$  and  $\hat{\rho}_{ijt} = \hat{h}_{ijt} / \sqrt{\hat{h}_{iit}\hat{h}_{jtt}}$  as the estimated conditional correlation between  $y_{it}$  and  $y_{jt}$ . Tse (2002) proposes to run the following regressions:

$$\hat{u}_{it}^2 - 1 = \hat{d}'_{it} \delta_i + \zeta_{it}, \quad i = 1, \dots, N \quad (58)$$

$$\hat{u}_{it}\hat{u}_{jt} - \hat{\rho}_{ijt} = \hat{d}'_{ijt} \delta_{ij} + \zeta_{ijt}, \quad 1 \leq i < j \leq N \quad (59)$$

where  $\hat{d}_{it}$  and  $\hat{d}_{ijt}$  are the estimated counterparts of respectively  $d_{it} = (u_{i,t-1}^2, \dots, u_{i,t-M}^2)'$ ,  $d_{ijt} = (u_{i,t-1}u_{j,t-1}, \dots, u_{i,t-M}u_{j,t-M})'$ , and  $\delta_i$  and  $\delta_{ij}$  are the regression coefficients. The choice of the

regressors may be changed depending on the particular type of model inadequacy one wants to investigate. An advantage of the residual-based diagnostics is that they focus on several distinctive aspects of possible causes of 'remaining' ARCH effects.

Tse (2002) shows that under reasonable assumptions the statistics  $RB(M)_1 = T\hat{\delta}'_i\hat{L}_i\hat{\Omega}_i^{-1}\hat{L}_i\hat{\delta}_i$  and  $RB(M)_2 = T\hat{\delta}'_{ij}\hat{L}_{ij}\hat{\Omega}_{ij}^{-1}\hat{L}_{ij}\hat{\delta}_{ij}$  are each asymptotically distributed as  $\chi^2(M)$  under the null of correct specification of the first two conditional moments, where

$$L_i = \text{plim} \left( \frac{1}{T} \sum d_{it}d'_{it} \right) \quad L_{ij} = \text{plim} \left( \frac{1}{T} \sum d_{ijt}d'_{ijt} \right) \quad (60)$$

$$\Omega_i = E\{(u_{it}^2 - 1)L_i - Q_i G Q'_i\} \quad \Omega_{ij} = E\{(u_{ijt}^2 - 1)L_{ij} - Q_{ij} G Q'_{ij}\} \quad (61)$$

$$Q_i = \text{plim} \left\{ \frac{1}{T} \sum d_{it} \frac{\partial u_{it}^2}{\partial \theta'} \right\} \quad Q_{ij} = \text{plim} \left\{ \frac{1}{T} \sum d_{ijt} \frac{\partial (u_{ijt}^2 - \rho_{ijt})}{\partial \theta'} \right\} \quad (62)$$

and assuming that under certain conditions, the MLE estimate of  $\theta$  satisfies the condition  $\sqrt{T}(\hat{\theta} - \theta) \rightarrow N(0, G)$ . Naturally, to compute these statistics, one has to replace the unobservable components by their estimated counterparts.

### 4.3. Lagrange Multiplier Tests

Lagrange multiplier tests usually have a higher power than portmanteau tests when the alternative is correct (although they can be asymptotically equivalent in certain cases), but they may have low power against other alternatives. Bollerslev *et al.* (1988) and Engle and Kroner (1995), among others, have developed LM tests for MGARCH models. Recently, Sentana and Fiorentini (2001) have developed a simple preliminary test for ARCH effects in common factor models.

To reduce the number of parameters in the estimation of MGARCH models, it is usual to introduce restrictions. For instance, the CCC model of Bollerslev (1990) assumes that the conditional correlation matrix is constant over time. It is then desirable to test this assumption afterwards. Tse (2000) proposes a test for constant correlations. The null is  $h_{ijt} = \rho_{ij}\sqrt{h_{iit}h_{jtt}}$  where the conditional variances are GARCH (1,1), the alternative is  $h_{ijt} = \rho_{ijt}\sqrt{h_{iit}h_{jtt}}$ . The test statistic is a LM statistic which under the null is asymptotically  $\chi^2(N(N-1)/2)$ . Bera and Kim (2002) also develop a test for constancy of the correlation parameters in the CCC model of Bollerslev (1990). It is an information matrix-type test that besides constant correlations examines at the same time various features of the specified model. An alternative test has been proposed by Longin and Solnik (1995).

Engle and Sheppard (2001) propose another test of the constant correlation hypothesis, in the spirit of the DCC models presented earlier. The null  $H_0 : R_t = \bar{R}\forall t$  is tested against the alternative  $H_1 : \text{vech}(R_t) = \text{vech}(\bar{R}) + \beta_1^* \text{vech}(R_{t-1}) + \dots + \beta_p^* \text{vech}(R_{t-p})$ . The test is easy to implement since  $H_0$  implies that coefficients in the regression  $X_t = \beta_0^* + \beta_1^* X_{t-1} + \dots + \beta_p^* X_{t-p} + u_t^*$  are equal to zero, where  $X_t = \text{vech}^u(\hat{z}_t \hat{z}'_t - I_N)$ ,  $\text{vech}^u$  is like the vech operator but it only selects the elements under the main diagonal,  $\hat{z}_t = \hat{R}^{-1/2} \hat{D}_t^{-1} \hat{e}_t$  is the  $N \times 1$  vector of standardized residuals (under the null), and  $D_t = \text{diag}(h_{11t}^{1/2} \dots h_{NNt}^{1/2})$ .

## 5. CONCLUSION

The main purpose of this paper is to review MGARCH models. Since the seminal paper of Engle (1982), much progress has been made in understanding GARCH models and their multivariate extensions. As mentioned in Section 1, these models are increasingly used in applied financial econometrics. Given the large and increasing variety of existing models, an applied econometrician is confronted with the issue of choosing among them for each particular application. A related question is ‘which model is most appropriate under which circumstances?’.

Applied research has naturally followed theoretical research and used existing models. There are very few papers where a comparison of different MGARCH models for the same problem and data is done.<sup>14</sup> The discussion paper version of this article—see Bauwens *et al.* (2003)—contains a detailed review of the application fields and gives an idea of which models have been used for each field. This reveals that to a large extent, when a model came out in the literature, applied researchers soon started to try it, discarding the previous models. This partly reflects the fact that parsimonious models were introduced progressively to overcome the difficulty of estimating the VEC model.

In our opinion, the crucial point in MGARCH modelling is to provide a realistic but parsimonious specification of the variance matrix ensuring its positivity. There is a dilemma between flexibility and parsimony. BEKK models are flexible but require too many parameters for multiple time series of more than four elements. Diagonal VEC and BEKK models are much more parsimonious but very restrictive for the cross-dynamics. They are not suitable if volatility transmission is the object of interest, but they usually do a good job in representing the dynamics of variances and covariances. This may be sufficient for some applications like asset pricing models.

In contrast, factor GARCH models allow the conditional variances and covariances to depend on the past of all variances and covariances, but they imply common persistence in all these elements. In this respect, the DCC models allow for different persistence between variances and correlations, but impose common persistence in the latter (although this may be relaxed). They open the door to handling more than a very small number of series. They are an extension of the CCC model which is relatively easy to estimate.

One way to deal with flexible but heavily parametrized models is to keep a flexible functional form and to reduce the number of parameters by imposing restrictions. An example of this approach is provided in Engle *et al.* (1990a), who impose restrictions in a VEC model. We conjecture that researchers will propose new ways to impose restrictions. An idea is to base the restrictions on preliminary (easy to obtain) estimates, in the spirit of Ledoit *et al.* (2003) or Bauwens and Rombouts (2003).

Finally, here is a list of open issues/research topics (stated in a condensed way since they have been discussed in this survey):

1. Improving software for inference (this is a prerequisite for progress in applications).
2. Comparing the performance and assessing the financial value of different specifications in applications.<sup>15</sup>
3. Implications of stability or not of a model class with respect to linear transformations.
4. More flexible specifications for the dynamics of correlations of DCC models.

<sup>14</sup> Examples are Karolyi (1995), Bera *et al.* (1997), Kroner and Ng (1998), Engle and Sheppard (2001).

<sup>15</sup> See Rombouts and Verbeek (2004) and Fleming *et al.* (2003) for recent examples.

5. Unconditional moments of correlations/covariances, marginalization and temporal aggregation in DCC models.
6. Development of a copula tool for specification and inference.
7. Impact of choice of the square root decomposition of  $H_t$  on statistical procedures.
8. Conditions for two-step efficient estimation (MGARCH on residuals of the mean model).
9. Asymptotic properties of MLE (in particular low level, easy to check, sufficient conditions for asymptotic normality when it holds).
10. Further developments of multivariate diagnostic tests.

There is little doubt that progress on these issues would greatly contribute to the theory and practice of MGARCH models.

#### ACKNOWLEDGEMENTS

The authors would like to thank Christian Hafner and Roy van der Weide for useful comments. They are especially grateful to three referees and the editor (T. Bollerslev) for their detailed comments and numerous suggestions that helped to improve the paper. This text presents research results of the Belgian Programme on Interuniversity Poles of Attraction initiated by the Belgian State Prime Minister's Office, Science Policy Programming. Scientific responsibility is assumed by the authors.

#### REFERENCES

- Alexander CO. 2001. *Market Models*. John Wiley: New York.
- Alexander CO, Chibumba AM. 1997. Multivariate orthogonal factor GARCH. Mimeo, University of Sussex.
- Andersen TG, Bollerslev T, Diebold FX, Labys P. 2003. Modeling and forecasting realized volatility. *Econometrica* **71**: 579–625.
- Attanasio O. 1991. Risk, time-varying second moments and market efficiency. *Review of Economic Studies* **58**: 479–494.
- Baillie RT, Bollerslev T, Mikkelsen HO. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**: 3–30.
- Barndorff-Nielsen OE, Shephard N. 2001. Normal modified stable processes. *Theory of Probability and Mathematics Statistics* **65**: 1–19.
- Bauwens L, Laurent S. 2002. A new class of multivariate skew densities, with application to GARCH models. *Journal of Business and Economic Statistics*.
- Bauwens L, Rombouts JVK. 2003. Bayesian clustering of similar GARCH models. CORE DP 2003/87.
- Bauwens L, Laurent S, Rombouts JVK. 2003. Multivariate GARCH models: a survey. CORE DP 2003/31.
- Bera AK, Higgins ML. 1993. ARCH models: properties, estimation and testing. *Journal of Economic Surveys* **7**: 305–362.
- Bera AK, Kim S. 2002. Testing constancy of correlation and other specifications of the BGARCH model with an application to international equity returns. *Journal of Empirical Finance* **9**: 171–195.
- Bera AK, Garcia P, Roh JS. 1997. Estimation of time-varying hedging ratios for corns and soybeans: BGARCH and random coefficient approaches. *Sankhya* **59**: 346–368.
- Billio M, Caporin M, Gobbo M. 2003. Block dynamic conditional correlation multivariate GARCH models. Greta Working Paper 03.03.
- Black F. 1976. Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.
- Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.

- Bollerslev T. 1990. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics* **72**: 498–505.
- Bollerslev T, Chou RY, Kroner KF. 1992. ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* **52**: 5–59.
- Bollerslev T, Engle RF. 1993. Common persistence in conditional variances. *Econometrica* **61**: 167–186.
- Bollerslev T, Engle RF, Nelson DB. 1994. ARCH models. In *Handbook of Econometrics*, Engle R, McFadden D (eds). North Holland Press: Amsterdam.
- Bollerslev T, Engle RF, Wooldridge JM. 1988. A capital asset pricing model with time varying covariances. *Journal of Political Economy* **96**: 116–131.
- Bollerslev T, Wooldridge JM. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* **11**: 143–172.
- Box GEP, Jenkins GM. 1970. *Time Series Analysis, Forecasting and Control*. Holden-Day: San Francisco.
- Brooks C, Burke SP, Persaud G. 2003. Multivariate GARCH models: software choice and estimation issues. *Journal of Applied Econometrics* **18**: 725–734.
- Brunetti C, Gilbert CL. 2000. Bivariate FIGARCH and fractional cointegration. *Journal of Empirical Finance* **7**: 509–530.
- Christodoulakis GA, Satchell SE. 2002. Correlated ARCH: modelling the time-varying correlation between financial asset returns. *European Journal of Operations Research* **139**: 351–370.
- Comte F, Lieberman O. 2003. Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* **84**: 61–84.
- De Santis G, Gérard B. 1998. How big is the premium for currency risk? *Journal of Financial Economics* **49**: 375–412.
- Diebold FX, Nerlove M. 1989. The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics* **4**: 1–21.
- Diebold FX, Gunther TA, Tay AS. 1998. Evaluating density forecasts, with applications to financial risk management. *International Economic Review* **39**: 863–883.
- Diebold FX, Hahn J, Tay AS. 1999. Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange. *Review of Economics and Statistics* **81**: 863–883.
- Ding Z, Engle RF. 2001. Large scale conditional covariance matrix modeling, estimation and testing. Working Paper FIN-01-029, NYU Stern School of Business.
- Ding Z, Granger CWJ, Engle RF. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1**: 83–106.
- Doz C, Renault E. 2003. Conditionally heteroskedastic factor models: identification and instrumental variables estimation. Mimeo.
- Drost CD, Klaassen CAJ. 1997. Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* **81**: 193–221.
- Drost CD, Nijman T. 1993. Temporal aggregation of GARCH processes. *Econometrica* **61**: 909–927.
- Duchesne P, Lalancette S. 2003. On testing for multivariate ARCH effects in vector time series models. *La Revue Canadienne de Statistique* **31**: 275–292.
- Dufour J-M, Khalaf L, Beaulieu M-C. 2003. Finite-sample diagnostics for multivariate regression with applications to linear asset pricing models. Working Paper No. 06-2003, CIREQ.
- El Babsiri M, Zakoian J-M. 2001. Contemporaneous asymmetry in GARCH processes. *Journal of Econometrics* **101**: 257–294.
- Engle R, Kroner FK. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**: 122–150.
- Engle R, Mezrich J. 1996. GARCH for groups. *RISK* **9**: 36–40.
- Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007.
- Engle RF. 2002. Dynamic conditional correlation—a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* **20**: 339–350.
- Engle RF, Bollerslev T. 1986. Modeling the persistence of conditional variances. *Econometric Reviews* **5**: 1–50.
- Engle RF, González-Rivera G. 1991. Semiparametric ARCH model. *Journal of Business and Economic Statistics* **9**: 345–360.



- Engle RF, Sheppard K. 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Mimeo, UCSD.
- Engle RF, Ito T, Lin W-L. 1990a. Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica* **58**: 525–542.
- Engle RF, Ng VK, Rothschild M. 1990b. Asset pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics* **45**: 213–238.
- Fiorentini G, Calzolari G, Panattoni L. 1996. Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics* **11**: 399–417.
- Fiorentini G, Sentana E, Calzolari G. 2003. Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with Student  $t$  innovations. *Journal of Business and Economic Statistics* **21**: 532–546.
- Fiorentini G, Sentana E, Shephard N. 2004. Likelihood-based estimation of latent generalized ARCH structures. *Econometrica*, **72**: 1481–1517.
- Fleming J, Kirby C, Ostdiek B. 2003. The economic value of volatility timing using ‘realized’ volatility. *Journal of Financial Economics* **67**: 473–509.
- Franses PH, van Dijk D. 2000. *Non-Linear Series Models in Empirical Finance*. Cambridge University Press: Cambridge.
- Gallant RA, Tauchen GE. 2001. SNP: a program for nonparametric time series analysis. <http://www.econ.duke.edu/get/>.
- Ghysels E, Harvey A, Renault E. 1996. Stochastic volatility. In *Handbook of Statistics*, Maddala GS, Rao CR (eds). Elsevier Science: Amsterdam.
- Giot P, Laurent S. 2003. Value-at-Risk for long and short positions. *Journal of Applied Econometrics* **18**: 641–664.
- Glosten LR, Jagannathan R, Runkle DE. 1993. On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**: 1779–1801.
- Gourieroux C. 1997. *ARCH Models and Financial Applications*. Springer-Verlag: New York.
- Gourieroux C, Jasiak J. 2001. *Financial Econometrics*. Princeton University Press: Princeton, NJ.
- Hafner CM. 2003. Fourth moment structure of multivariate GARCH processes. *Journal of Financial Econometrics* **1**: 26–54.
- Hafner CM. 2004. Temporal aggregation of multivariate GARCH processes. Econometric Institute Report 29. Erasmus University Rotterdam.
- Hafner C, Herwartz H. 1998. Time-varying market price of risk in the CAPM. Approaches, empirical evidence and implications. *Finance* **19**: 93–112.
- Hafner CM, Herwartz H. 2003. Analytical quasi maximum likelihood inference in multivariate volatility models. Econometric Institute Report 21. Erasmus University Rotterdam.
- Hafner CM, Rombouts JVK. 2003. Estimation of temporally aggregated multivariate GARCH models. Core DP 2003/73.
- Hafner CM, Rombouts JVK. 2004. Semiparametric multivariate volatility models. Econometric Institute Report 21. Erasmus University Rotterdam.
- Hall P, Yao P. 2003. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**: 285–317.
- Hansson B, Hordahl P. 1998. Testing the conditional CAPM using multivariate GARCH-M. *Applied Financial Economics* **8**: 377–388.
- Harvey AC, Ruiz E, Shephard N. 1992. Unobservable component time series models with ARCH disturbances. *Journal of Econometrics* **52**: 129–158.
- Harvey AC, Ruiz E, Shephard N. 1994. Multivariate stochastic variance models. *Review of Economic Studies* **61**: 247–264.
- He C, Teräsvirta T. 2002a. An application of the analogy between vector ARCH and vector random coefficient autoregressive models. SSE/EFI Working Paper Series in Economics and Finance No. 516, Stockholm School of Economics.
- He C, Teräsvirta T. 2002b. An extended conditional correlation GARCH model and its fourth-moment structure. SSE/EFI Working Paper Series in Economics and Finance No. 509, Stockholm School of Economics.
- Hosking JRM. 1980. The multivariate portmanteau statistic. *Journal of American Statistical Association* **75**: 602–608.



- Jeantheau T. 1998. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**: 70–86.
- Jondeau E, Rockinger M. 2001. The copula-GARCH model of conditional dependencies: an international stock-market application. *Journal of International Money and Finance*.
- Kariya T. 1988. MTV model and its application to the prediction of stock prices. In *Proceedings of the Second International Tampere Conference in Statistics*, Pullila T, Puntanen S (eds). University of Tampere, Finland.
- Karolyi GA. 1995. A multivariate GARCH model of international transmission of stock returns and volatility: the case of the United States and Canada. *Journal of Business and Economic Statistics* **13**: 11–25.
- Kawakatsu H. 2003. Cholesky factor GARCH. Mimeo, Quantitative Micro Software, Irvine, CA.
- Kearney C, Patton AJ. 2000. Multivariate GARCH modelling of exchange rate volatility transmission in the European Monetary System. *Financial Review* **41**: 29–48.
- Kim J. 2000. The relationship between the monetary regime and output volatility: a multivariate GARCH-M model of the Japanese experience, 1919–1996. *Japan and the World Economy* **12**: 49–69.
- Kroner FK, Ng VK. 1998. Modelling asymmetric comovements of asset returns. *The Review of Financial Studies* **11**: 817–844.
- Ledoit O, Santa-Clara P, Wolf M. 2003. Flexible multivariate GARCH modeling with an application to international stock markets. *The Review of Economics and Statistics* **85**: 735–747.
- Lee SW, Hansen BE. 1994. Asymptotic properties of the maximum likelihood estimator and test of the stability of parameters of the GARCH and IGARCH models. *Econometric Theory* **10**: 29–52.
- Lien D, Tse YK. 2002. Some recent developments in futures hedging. *Journal of Economic Surveys* **16**: 357–396.
- Lin WL. 1992. Alternative estimators for factor GARCH models—a Monte Carlo comparison. *Journal of Applied Econometrics* **7**: 259–279.
- Ling S, Li WK. 1997. Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis* **18**: 447–464.
- Ling S, McAleer M. 2003. Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* **19**: 280–310.
- Longin F, Solnik B. 1995. Is the correlation in international equity returns constant: 1960–1990? *Journal of International Money and Finance* **14**: 3–26.
- Lucchetti R. 2002. Analytic score for multivariate GARCH models. *Computational Economics* **19**: 133–143.
- Lumsdaine RL. 1996. Asymptotic properties of the quasi maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. *Econometrica* **64**: 575–596.
- Lütkepohl H. 1996. *Handbook of Matrices*. John Wiley: New York.
- McCullough BD, Vinod HD. 1999. The numerical reliability of econometric software. *Journal of Economic Literature* **37**: 633–665.
- Meddahi N, Renault E. 1996. Aggregation and marginalization of GARCH and stochastic volatility models. Working Paper, Department of Economics, University of Montréal.
- Mencía FJ, Sentana E. 2003. Estimation and testing of dynamic models with generalised hyperbolic innovations. Mimeo, CEMFI.
- Nelsen RB. 1999. *An Introduction to Copulas*. Springer-Verlag: New York.
- Nelson DB, Cao CQ. 1992. Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics* **10**: 229–235.
- Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 349–370.
- Newey WK. 1985. Maximum likelihood specification testing and conditional moment tests. *Econometrica* **6**: 1047–1070.
- Newey WK, Steigerwald DS. 1997. Asymptotic bias for quasi maximum likelihood estimators in conditional heteroskedasticity models. *Econometrica* **3**: 587–599.
- Nijman T, Sentana E. 1996. Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* **71**: 71–87.
- Pagan A. 1996. The econometrics of financial markets. *Journal of Empirical Finance* **3**: 15–102.
- Palm FC. 1996. GARCH models of volatility. In *Handbook of Statistics*, Maddala GS, Rao CR (eds). Elsevier Science: Amsterdam.

- Patton A. 2000. Modelling time-varying exchange rate dependence using the conditional copula. University of California, San Diego, Discussion Paper 01–09.
- Pelletier D. 2003. Regime switching for dynamic correlations. *Journal of Econometrics*.
- Rigobon R, Sack B. 2003. Spillover across US financial markets. NBER Working Paper 9640.
- Riskmetrics. 1996. *Riskmetrics Technical Document*, 4th edn. J.P. Morgan: New York.
- Rombouts JVK, Verbeek M. 2004. Evaluating portfolio Value-At-Risk using semi-parametric GARCH models. Discussion Paper ERS-2004-107-F&A, Erasmus Research Institute of Management, Erasmus University Rotterdam.
- Sentana E. 1995. Quadratic ARCH models. *Review of Economic Studies* **62**: 639–661.
- Sentana E. 1998. The relation between conditionally heteroskedastic factor models and factor GARCH models. *Econometrics Journal* **1**: 1–9.
- Sentana E, Fiorentini G. 2001. Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics* **102**: 143–164.
- Shephard N. 1996. Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields*, Hinkley DV, Cox DR, Barndorff-Nielsen OE (eds). Chapman & Hall: London.
- Sklar A. 1959. Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* **8**: 229–231.
- Tauchen G. 1985. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* **30**: 415–443.
- Timmermann A. 2000. Density forecasting in economics and finance. *Journal of Forecasting* **19**: 120–123.
- Tsay RS. 2002. *Analysis of Financial Time Series*. John Wiley: New York.
- Tse YK. 2000. A test for constant correlations in a multivariate GARCH model. *Journal of Econometrics* **98**: 107–127.
- Tse YK. 2002. Residual-based diagnostics for conditional heteroscedasticity models. *Econometrics Journal* **5**: 358–373.
- Tse YK, Tsui AKC. 1999. A note on diagnosing multivariate conditional heteroscedasticity models. *Journal of Time Series Analysis* **20**: 679–691.
- Tse YK, Tsui AKC. 2002. A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* **20**: 351–362.
- van der Weide R. 2002. GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* **17**: 549–564.
- Vlaar PJG, Palm FC. 1993. The message in weekly exchange rates in the European Monetary System: mean reversion, conditional heteroskedasticity and jumps. *Journal of Business and Economic Statistics* **11**: 351–360.
- Vrontos ID, Dellaportas P, Politis DN. 2003. A full-factor multivariate GARCH model. *Econometrics Journal* **6**: 311–333.
- Zakoian J-M. 1994. Threshold heteroskedasticity models. *Journal of Economic Dynamics and Control* **15**: 931–955.