



TI 2010-032/2

Tinbergen Institute Discussion Paper

# A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations

*Drew Creal<sup>a</sup>*

*Siem Jan Koopman<sup>b,d</sup>*

*André Lucas<sup>c,d</sup>*

<sup>a</sup> *Booth School of Business, University of Chicago;*

<sup>b</sup> *Dept. of Econometrics, VU University Amsterdam;*

<sup>c</sup> *Dept. of Finance, VU University Amsterdam, Duisenberg school of finance;*

<sup>d</sup> *Tinbergen Institute.*

### **Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

### **Tinbergen Institute Amsterdam**

Roetersstraat 31  
1018 WB Amsterdam  
The Netherlands  
Tel.: +31(0)20 551 3500  
Fax: +31(0)20 551 3555

### **Tinbergen Institute Rotterdam**

Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at  
<http://www.tinbergen.nl>.

# A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations\*

Drew Creal<sup>a</sup>, Siem Jan Koopman<sup>b,d</sup>, André Lucas<sup>c,d</sup>

<sup>(a)</sup> *Booth School of Business, University of Chicago*

<sup>(b)</sup> *Department of Econometrics, VU University Amsterdam*

<sup>(c)</sup> *Department of Finance, VU University Amsterdam, Duisenberg school of finance*

<sup>(d)</sup> *Tinbergen Institute, Amsterdam*

March 15, 2010

## Abstract

We propose a new class of observation-driven time-varying parameter models for dynamic volatilities and correlations to handle time series from heavy-tailed distributions. The model adopts generalized autoregressive score dynamics to obtain a time-varying covariance matrix of the multivariate Student's  $t$  distribution. The key novelty of our proposed model concerns the weighting of lagged squared innovations for the estimation of future correlations and volatilities. When we account for heavy tails of distributions, we obtain estimates that are more robust to large innovations. The model also admits a representation as a time-varying heavy-tailed copula which is particularly useful if the interest focuses on dependence structures. We provide an empirical illustration for a panel of daily global equity returns.

**Keywords:** dynamic dependence, multivariate Student's  $t$  distribution, copula.

**JEL classification codes:** C10, C22, C32, C51.

---

\*We would like to thank Ruey Tsay for his comments on an earlier draft of the paper. Corresponding author: Drew Creal, University of Chicago, Booth School of Business, 5807 S. Woodlawn Ave, Chicago, IL 60637, phone: 773.834.5249, email: dcreal@chicagobooth.edu.

# 1 Introduction

We contribute to the literature on multivariate modeling of volatilities and correlations by introducing a class of observation-driven time-varying parameter models with heavy tailed distributions. In particular, we consider a multivariate Student's  $t$  model with time-varying volatilities and correlations for which the multivariate Gaussian model is a special case. To anticipate the needs of different users, we introduce two levels of flexibility in the model. First, we propose a copula version of the model which treats the marginal distributions of each of the series separately from the dependence structure. The copula model enables the optimization problem to be broken into more manageable pieces. Secondly, we modify the model to accommodate alternative covariance matrix specifications. For example, we can consider the square root of the correlation matrix in terms of hyperspherical coordinates. The general model formulation enable us to impose a factor structure on either or both of the time-varying volatilities and correlations.

Modeling the conditional distribution of a large group of assets is an important challenge in modern financial time series analysis. Empirical evidence indicates that both the conditional volatilities and correlations of assets change over time. Time-varying volatilities and correlations among assets have practical implications for risk management and asset pricing. To capture these features of the data, two classes of models are generally considered in the literature. The first class comprises observation-driven models which include multivariate extensions of the univariate generalized autoregressive conditional heteroskedastic (GARCH) family of models introduced by Engle (1982) and Bollerslev (1986). The second class are parameter-driven models such as the multivariate stochastic volatility models of Chib, Nardari, and Shephard (2006) and Gouriéroux, Jasiak, and Sufana (2009). This paper focuses on observation-driven models for time-varying correlations. Time-varying correlation GARCH models were originally developed by Ding and Engle (2001), Engle (2002), Engle and Sheppard (2001), and Tse and Tsui (2002). Bauwens, Laurent, and Rombouts (2006) present a survey on multivariate GARCH models covering time-varying correlation models as well as models for time-varying covariances.

When modeling the time-varying covariance matrix of a multivariate time series, it is well-known that the number of static as well as time-varying parameters grows quickly as more series are added. The increasing dimensionality of the parameter space creates challenging numerical

problems and some of the additional parameters may be statistically insignificant. Various trade-offs have been recognized between building models that lead to a better fit of the data statistically and balancing practical considerations. Many alternative models and estimation procedures are proposed to address these challenges.

A popular device is the use of time-varying multivariate Gaussian copulas in which the variances/standard deviations are modeled separately from the correlations. The numerical optimization problem is then separated into more manageable pieces. Another key device is to impose restrictions on the parameter space and to limit the number of parameters that control the dynamics of the correlation matrix. Both of these strategies are taken by Engle (2002) in the successful dynamic conditional correlation (DCC) model. The modeling approach is motivated by pragmatic reasons as the DCC is intended to scale well when the cross-sectional dimension increases. In related models a factor structure is imposed on the volatilities and correlations; see, for example, Tsay (2005) and Fan, Wang, and Yao (2008). A factor structure reduces the number of time-varying parameters and potentially allows the user to extract more information from the data. Factor structures also allow us to pose interesting questions such as which series share common features and what economic factors drive correlations. For example, common macroeconomic shocks as well as arbitrage opportunities generally force some common dynamics on groups of assets. Ultimately, the appropriateness of a model and its associated estimation procedure depends upon the application.

This paper presents the details of how time-varying volatilities and correlations can be incorporated in the multivariate Student's  $t$  density using the generalized autoregressive score (GAS) framework of Creal, Koopman, and Lucas (2010). The resulting model is shown to be effective in treating different dynamic features simultaneously in a unified way. In our empirical illustration, we analyze daily return data for a group of global equity indices over a period of 15 years. We show that our Student's  $t$  GAS model accounts for outliers when updating the correlations and volatilities over time.

The remainder of the paper is organized as follows. In Sections 2 and 3 we present the basic model specification and updating equation. Section 4 proposes alternative specifications and factor model structures. Section 5 discusses maximum likelihood estimation and carries out a Monte Carlo study to compare its performance with other dynamic correlation models. Section 6 contains our illustration. Section 7 concludes.

## 2 The basic model and general result

### 2.1 Multivariate Student's $t$ density

Let observation vector  $y_t \in \mathbb{R}^k$  follow a standardized Student's  $t$  distribution with  $\nu$  degrees of freedom. To simplify the notation, we set the location parameter  $\mu_t$  of  $y_t$  to zero. If  $\mu_t \neq 0$ ,  $y_t$  is replaced by  $y_t - \mu_t$  below. The model can thus easily be extended to allow for regressors and dynamics in the mean. The variance matrix of  $y_t$  is denoted by  $\Sigma_t$ . We assume that  $\nu > 2$ , such that the variance matrix exists. This assumption is made to conform our notation as much as possible to the volatility clustering literature. It is straightforward, however, to generalize the model below to the case  $0 < \nu \leq 2$  by taking the scaling matrix of the Student's  $t$  distribution, rather than its variance matrix as the key parameter.

The observation density of  $y_t$  is given by

$$p(y_t|\Sigma_t; \nu) = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) [(\nu-2)\pi]^{k/2} |\Sigma_t|^{1/2}} \left[1 + \frac{y_t' \Sigma_t^{-1} y_t}{(\nu-2)}\right]^{-(\nu+k)/2}, \quad (1)$$

where the variance matrix  $\Sigma_t$  is strictly positive-definite and can be a function of time-varying factors. For reference purposes below, we consider two special cases of the density (1). In case  $\nu^{-1} \rightarrow 0$ , the density reduces to the multivariate Gaussian density as given by

$$p_G(y_t|\Sigma_t) = \frac{1}{(2\pi)^{k/2} |\Sigma_t|^{1/2}} \exp\left(-\frac{1}{2} y_t' \Sigma_t^{-1} y_t\right). \quad (2)$$

In the univariate case  $k = 1$ , density (1) reduces to the Student's  $t$  density given by

$$p(y_t|\sigma_t^2; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{(\nu-2)\pi\sigma_t^2}} \left[1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right]^{-(\nu+1)/2}, \quad (3)$$

where  $\sigma_t^2 > 0$  is a time-varying scalar.

### 2.2 Generalized Autoregressive Score model

The generalized autoregressive score (GAS) model is an observation-driven model that allows parameters to change over time using information from the score of the observation density.

We collect the time-varying parameters of a model density in the vector  $f_t$  and we specify the autoregressive updating function by

$$f_{t+1} = \omega + \sum_{i=1}^p A_i s_{t-i+1} + \sum_{j=1}^q B_j f_{t-j+1}, \quad (4)$$

where  $\omega$  is a vector of constants, coefficient matrices  $A_i$  and  $B_j$  have appropriate dimensions for  $i = 1, \dots, p$  and  $j = 1, \dots, q$ , while  $s_t$  is an appropriately scaled function. The unknown coefficients in (4) are functions of the parameter vector  $\theta$ , that is  $\omega = \omega(\theta)$ ,  $A_i = A_i(\theta)$ , and  $B_j = B_j(\theta)$  for  $i = 1, \dots, p$  and  $j = 1, \dots, q$ . The coefficient matrices  $B_1, \dots, B_q$  determine the persistence of the vector  $f_t$  over time.

The GAS framework is developed by Creal, Koopman, and Lucas (2010) who let the driving mechanism  $s_t$  be the scaled derivative of the density function at time  $t$  with respect to the parameter vector  $f_t$ , that is

$$s_t = S_t \nabla_t, \quad \nabla_t = \frac{\partial \log p(y_t | f_t, \mathcal{F}_{t-1}; \theta)}{\partial f_t}, \quad (5)$$

where  $p(y_t | f_t, \mathcal{F}_{t-1}; \theta)$  is the observation density function,  $\mathcal{F}_t$  collects all relevant information up to time  $t$  (including covariates,  $y_j$  and  $f_j$  for  $j = 1, \dots, t$ ) and  $S_t$  is a scaling matrix of appropriate dimension. For a given density function, the equations (4) and (5) constitute our generalized autoregressive score model of orders  $p$  and  $q$ . We abbreviate it by GAS( $p, q$ ).

Different choices for the scaling matrix  $S_t$  can be considered and will lead to different GAS models. An intuitive choice is to base the scaling on the curvature of the logarithm of the observation density (1) at time  $t$ . For example, we can have  $S_t$  equal to the inverse of the Fisher information matrix, that is

$$S_t = \mathcal{I}_{t|t-1}^{-1}, \quad \mathcal{I}_{t|t-1} = \mathbb{E}_{t-1} [\nabla_t \nabla_t'], \quad (6)$$

such that  $\text{Var}(s_t) = \mathcal{I}_{t|t-1}^{-1}$ . The resulting recursion in (4) can then be interpreted as a Gauss-Newton algorithm for estimating  $f_t$  through time.

Different choices for  $S_t$  are possible. For example, Nelson and Foster (1994) derive optimal filtering properties for the GARCH updating equation when the scaling is based on  $\mathcal{I}_{t|t-1}^{-1/2}$  rather

than  $\mathcal{I}_{t|t-1}^{-1}$ . Our choice to set  $S_t = \mathcal{I}_{t|t-1}^{-1}$ , however, brings our GAS model closest to the familiar standard GARCH specification. For example, consider the case of a univariate series  $y_t$  that is normally distributed,  $y_t \sim \mathcal{N}(0, \sigma_t^2)$ . We let  $f_t = \sigma_t^2$  and adopt the scaling (6) such that  $\nabla_t = -\frac{1}{2}f_t^{-1} + \frac{1}{2}f_t^{-2}y_t^2$  and  $S_t = 2f_t^2$ . When we scale by the inverse information matrix, the updating equation (4) with  $p = q = 1$  becomes

$$f_{t+1} = \omega + A_1(y_t^2 - f_t) + B_1f_t, \quad (7)$$

which is equivalent to the GARCH model of Bollerslev (1986). The GARCH(1,1) model is usually presented by  $f_{t+1} = \omega + \alpha_1y_t^2 + \beta_1f_t$  so that  $\alpha_1 = A_1$  and  $\beta_1 = B_1 - A_1$  in relation to (7). A similar argument holds in the multivariate context.

When we consider the Student's  $t$  density (3) for  $y_t$  and let  $f_t = \sigma_t^2$  with scaling (6), we obtain the updating equation

$$f_{t+1} = \omega + A_1 \cdot (1 + 3\nu^{-1}) \cdot \left( \frac{(1 + \nu^{-1})}{(1 - 2\nu^{-1})(1 + \nu^{-1}y_t^2 / ((1 - 2\nu^{-1}) \cdot f_t))} y_t^2 - f_t \right) + B_1f_t, \quad (8)$$

which is different from the Student's  $t$ -GARCH(1,1) model of Bollerslev (1987). The standard  $t$ -GARCH(1,1) model is also based on (3) but has updating equation (7). The denominator of the second term in the right-hand side of (8) causes a more moderate increase in the variance for a large realization of  $|y_t|$  as long as  $\nu$  is finite. Hence a large absolute realization of  $y_t$  does not always result in a substantial increase in the variance. The intuition for this specification is clear. If the density of  $y_t$  is heavy-tailed, a large value of  $y_t^2$  is not necessarily due to an increase in variance. It may also be due to the heavy-tailed feature of the distribution. The functional form in (8) automatically corrects for this. Large values of  $y_t^2$  have a bounded influence on  $f_{t+1}$ . This feature is not imposed explicitly, but it follows from the choice of the GAS model to excite the factor recursions by the scaled density score. Similar univariate models have been derived by, for example, Nelson and Foster (1994) and Harvey and Chakravarty (2008). The models presented in the current paper, however, are inherently multivariate and applicable to a range of different specifications. Other extensions such as considering asymmetric multivariate Gaussian or Student's  $t$  distributions for the modeling of leverage in financial return series are also possible.



Creal et al. (2010) demonstrate that the GAS framework based on the Fisher information matrix (6) also nests many other successful econometric models, including the autoregressive conditional duration model for the exponential distribution of Engle and Russell (1998) and the multiplicative error model for the gamma distribution of Engle and Gallo (2006).

### 2.3 Matrix notation and definitions

To develop our results below, we adopt the following matrix notation and definitions. The Kronecker product is denoted by  $A \otimes B$  for two matrices  $A$  and  $B$ . In case  $B = A$ , we define  $A_{\otimes} = A \otimes A$ . The operator  $\text{vec}(A)$  vectorizes matrix  $A$  into a column vector while  $\text{vech}(A)$  vectorizes the lower triangular part of matrix  $A$  into a column vector. The unit matrix is denoted as  $I$ . We also define the operator  $\oplus$  for two matrices  $A$  and  $B$  as

$$A \oplus B = (A \otimes B) + (B \otimes A).$$

The duplication matrix  $\mathcal{D}_k$  and the elimination matrix  $\mathcal{B}_k$  are defined as

$$\mathcal{D}_k \text{vech}(A) = \text{vec}(A), \quad \mathcal{B}_k \text{vec}(A) = \text{vech}(A), \quad (9)$$

for a symmetric  $k \times k$  matrix  $A$  and with  $\mathcal{B}_k = (\mathcal{D}'_k \mathcal{D}_k)^{-1} \mathcal{D}'_k$ . A duplication matrix consists of ones and zeros only and has more rows than columns. The elimination matrix clearly has more columns than rows. The commutation matrix  $\mathcal{C}_k$  is defined as

$$\text{vec}(B) = \mathcal{C}_k \text{vec}(B'), \quad (10)$$

for any  $k \times k$  matrix  $B$ . An introductory but detailed treatment of matrix calculus in econometrics is presented by Abadir and Magnus (2005).

To introduce the GAS updating function for the density in (1) we specify the variance matrix  $\Sigma_t$  in (1) as a function  $\Sigma_t(f_t)$  of the time-varying factor  $f_t$  in (4). The precise functional form of the link between  $f_t$  and  $\Sigma_t$  is elaborated below. Typically, it differs between models that concentrate on correlations only (as in the copula specification), and model specifications for correlations and volatilities jointly.

## 2.4 General result

The theorem presented below forms the basis for the model specifications in Sections 3 and 4.

Define

$$\Psi_t = \Psi(f_t) = \frac{\partial \text{vech}(\Sigma_t)}{\partial f_t'}, \quad \Sigma_t = \Sigma_t(f_t). \quad (11)$$

**Theorem 1** *For density (1) and time-varying factor  $f_t$  in (4), we have*

$$\nabla_t = \frac{\partial p_t(y_t | \Sigma_t; \nu)}{\partial f_t} = \frac{1}{2} \Psi_t' \mathcal{D}'_k \Sigma_t^{-1} [w_t y_t \otimes - \text{vec}(\Sigma_t)], \quad (12)$$

$$\mathcal{I}_{t|t-1} = \text{E}[\nabla_t \nabla_t'] = \frac{1}{4} \Psi_t' \mathcal{D}'_k J_t' \otimes [gG - \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})'] J_t \otimes \mathcal{D}_k \Psi_t, \quad (13)$$

with matrix  $\mathcal{D}_k$  defined in (9), scalar  $w_t = (\nu + k) / (\nu - 2 + y_t' \Sigma_t^{-1} y_t)$ , matrix  $J_t$  defined implicitly as  $\Sigma_t^{-1} = J_t' J_t$ , scalar  $g = (\nu + k) / (\nu + 2 + k)$  and matrix  $G$  defined in (35) of the Appendix. The square root matrix  $J_t$  can be obtained from any convenient matrix decomposition procedure.

**Proof:** See Appendix.

Theorem 1 reveals a number of important features. First, irrespective of the model specification and whether we use the model as a copula or as a complete density, the dynamics of  $f_t$  are driven by the deviations of the (vectorized) weighted outer product  $w_t y_t y_t'$  from the local covariance matrix  $\Sigma_t$ . For the normal distribution, the weights  $w_t$  collapse to 1 and we obtain the familiar driving mechanism of a multivariate GARCH model.

Second, Theorem 1 shows that potentially different specifications of  $\Sigma_t$  are accounted for by the matrix function  $\Psi_t$ , which gathers the derivatives of the full covariance matrix  $\Sigma_t$  with respect to the dynamic factors  $f_t$ . The core of the updating scheme implied by Theorem 1 is not affected when a different specification is chosen for the correlations and/or volatilities; only the definition of  $\Psi_t$  needs to change in that case. The GAS framework therefore can accommodate a wide class of models for time-varying covariance matrices or, alternatively, for time-varying correlation matrices under a copula specification with known variances. Some further illustrations of alternative correlation specifications are discussed in Sections 3 and 4.

Third, the weight  $w_t$  in (12) also appears in the univariate specification of (8) with  $k = 1$ . If the density of the observations  $y_t$  is heavy-tailed ( $\nu^{-1} > 0$ ), large values in  $y_t y_t'$  (in absolute terms) do not automatically lead to dramatic changes in the elements of  $\Sigma_t$ . Such large values

may be due to the heavy-tailed feature of the distribution of  $y_t$  rather than to an increase in volatility. The weight  $w_t$  in (12) automatically accounts for extreme values since it decreases if  $y_t' \Sigma_t^{-1} y_t$  is large.

Fourth, as  $\nabla_t$  is the score of the density function with respect to  $f_t$ , it follows immediately that  $E_{t-1}(s_t) = 0$  and that  $s_t$  forms a martingale difference sequence. The process  $f_t$  can then be written as an infinite moving average of martingale differences. In case of the GAS(1,1) model, we have

$$f_t = \omega + \sum_{i=1}^{\infty} B_1^{i-1} A_1 s_{t-i}. \quad (14)$$

If the conditional variance of  $s_t$  is constant over time, the process of  $f_t$  is covariance stationary when the roots of  $B_1$  lying inside the unit circle. Generally, however, with inverse information matrix scaling, that is  $S_t = \mathcal{I}_{t|t-1}^{-1}$ , the conditional variance of  $s_t$  will not be constant and formulating conditions for a stationary process is much harder. Creal et al. (2010) solve this issue by scaling with  $S_t = \mathcal{I}_{t|t-1}^{-1/2}$  instead. In the developments below, we like to stay close to the multivariate GARCH framework which requires the score to be scaled by  $\mathcal{I}_{t|t-1}^{-1}$  as shown for the univariate case in section 2.2.

Fifth, the model uses unexpected deviations  $w_t y_t y_t' - \Sigma_t$  to drive the evolution of volatilities and correlations. Unexpected large or small cross-products are taken into account as well as unexpected large or small squared observations. The intuition is as follows. Consider the bivariate case  $k = 2$  with fixed unit variances and a time-varying correlation  $f_t = \rho_t$  for the normal distribution ( $\nu^{-1} = 0$ ). We have  $\text{vec}(\Sigma_t) = (1, \rho_t, \rho_t, 1)'$ . The key component in updating the time-varying correlation is the score  $\nabla_t$  in Theorem 1 for which it follows that  $\Psi_t' \mathcal{D}'_k = (0, 1, 1, 0)$ . It does not imply that only the second and third elements of  $w_t y_t \otimes - \text{vec}(\Sigma_t)$  (the cross-product terms) are taken into account. Since  $\Psi_t' \mathcal{D}'_k$  is first post-multiplied by  $\Sigma_t^{-1} \otimes \Sigma_t^{-1}$ , we obtain

$$s_t = \frac{2}{(1 - \rho_t^2)^2} \left[ (1 + \rho_t^2)(y_{1t} y_{2t} - \rho_t) - \rho_t (y_{1t}^2 + y_{2t}^2 - 2) \right], \quad (15)$$

where  $y_{it}$  is the  $i$ th element of  $y_t$  for  $i = 1, 2$ . The first term in (15) enforces an increase in the correlation when  $y_{1t} y_{2t}$  exceeds  $\rho_t$ . The second term is less intuitive. This term with expectation zero implies that if  $y_{1t}^2 + y_{2t}^2$  is large compared to its expected value of 2, the correlation should move towards zero due to the multiplication of this difference by  $-\rho_t$ . To illustrate the effect, assume that  $\rho_t = 0.5$ . Consider two different scenarios where we observe

$(y_{1t} = 1, y_{2t} = 1)$  or alternatively  $(y_{1t} = 0.25, y_{2t} = 4)$ . In both scenarios, the first term in (15) is the same and implies that the correlation should increase in the next period. For the first observation, the second term in (15) equals zero. For the second observation, however, the second term is negative and completely off-sets the effect of the first term. Hence, the second observation carries a much stronger signal that the observations are uncorrelated. By contrast, if the observation vector is  $(y_{1t} = 4, y_{2t} = 4)$ , the first term off-sets the second step and the correlation increases even though  $y_{1t}^2 + y_{2t}^2$  is large.

### 3 Update equations for volatilities and correlations

We develop the GAS updating equations for density (1) with different specifications for the time-varying variance  $\Sigma_t$ . In all cases, Theorem 1 applies since a different specification only affects how  $\Sigma_t$  depends on  $f_t$ . In this section we decompose the variance matrix in (1) by

$$\Sigma_t = D_t R_t D_t, \tag{16}$$

where  $D_t$  is the diagonal standard deviation matrix and  $R_t$  is the (symmetric) correlation matrix. Either one of or both the matrices  $D_t$  and  $R_t$  can be time-varying depending on the needs of the user. The decomposition (16) closely follows the DCC model of Engle (2002), though other decompositions might be followed as well, see Section 4.

#### 3.1 Time-varying volatilities and correlations

We consider here the multivariate Student's  $t$  density (1) with time varying  $D_t$  as well as  $R_t$ . We decompose the correlation matrix  $R_t$  as

$$R_t = \Delta_t^{-1} Q_t \Delta_t^{-1}, \tag{17}$$

where  $Q_t$  is a symmetric positive definite matrix,  $\Delta_t$  is a diagonal matrix whose non-zero elements equal the square root of the diagonal elements of  $Q_t$ . The transformation (17) ensures that the correlation matrix  $R_t$  is positive definite and symmetric with off-diagonal elements between  $[-1, 1]$ . The specification (17) is the same as for the DCC model of Engle (2002) and

Engle and Sheppard (2001), see also Tse and Tsui (2002). In case of the DCC model, the updating equation for the correlations is specified directly in terms of  $Q_t$  and is given by

$$Q_{t+1} = \Omega_{\text{dcc}}(\mathbf{I} - A_{\text{dcc}} - B_{\text{dcc}}) + A_{\text{dcc}} \odot y_t y_t' + B_{\text{dcc}} \odot Q_t, \quad (18)$$

where  $\odot$  is the Hadamard product (element by element multiplication) and  $k \times k$  matrices  $\Omega_{\text{dcc}}$ ,  $A_{\text{dcc}}$  and  $B_{\text{dcc}}$  are fixed and unknown. It can be shown that matrix  $Q_{t+1}$  remains positive definite when matrices  $\Omega_{\text{dcc}}$ ,  $A_{\text{dcc}}$  and  $B_{\text{dcc}}$  are positive definite. It is common practice to reduce the parameter space by replacing  $\Omega_{\text{dcc}}$  with the sample correlation matrix of  $y_t$  and by restricting  $A_{\text{dcc}} = a_{\text{dcc}} \cdot \iota \iota'$  and  $B_{\text{dcc}} = b_{\text{dcc}} \cdot \iota \iota'$  where  $a_{\text{dcc}}$  and  $b_{\text{dcc}}$  are scalars, and  $\iota$  is the vector of ones.

We specify our factor as

$$f_t = \begin{pmatrix} \text{diag}(D_t^2) \\ \text{vech}(Q_t) \end{pmatrix}. \quad (19)$$

with  $Q_t$  from (17). Our proposed updating equation (4) is given in Theorem 1, with

$$\begin{aligned} \Psi_t &= \frac{\partial \text{vech}(\Sigma_t)}{\partial f_t'} = \mathcal{B}_k \frac{\partial \text{vec}(\Sigma_t)}{\partial f_t'} = \mathcal{B}_k \frac{\partial \text{vec}(D_t R_t D_t)}{\partial f_t'} \\ &= \mathcal{B}_k (\mathbf{I} \oplus D_t R_t) \frac{\partial \text{vec}(D_t)}{\partial f_t'} + \mathcal{B}_k D_{t \otimes} \frac{\partial \text{vec}(\Delta_t^{-1} Q_t \Delta_t^{-1})}{\partial f_t'} \\ &= \mathcal{B}_k (\mathbf{I} \oplus D_t R_t) \frac{\partial \text{vec}(D_t)}{\partial f_t'} + \mathcal{B}_k D_{t \otimes} \Delta_{t \otimes}^{-1} \left( \mathcal{D}_k \frac{\partial \text{vech}(Q_t)}{\partial f_t'} - (\Delta_t \oplus Q_t) \Delta_{t \otimes}^{-1} \frac{\partial \text{vec}(\Delta_t)}{\partial f_t'} \right). \end{aligned}$$

Define the matrices  $W_{\Delta_t}$ ,  $W_{D_t}$ ,  $\mathcal{S}_\Delta$ ,  $\mathcal{S}_D$ , and  $\mathcal{S}_Q$  as follows. The matrix  $W_{\Delta_t}$  is constructed by having a  $k^2 \times k^2$  diagonal matrix with diagonal elements  $0.5 \text{vec}(\Delta_t^{-1})$  and then dropping the columns containing only zeros. Similarly, the matrix  $W_{D_t}$  is constructed from the  $k^2 \times k^2$  diagonal matrix with diagonal elements  $0.5 \text{vec}(D_t^{-1})$  and then dropping the columns containing only zeros. The matrices  $\mathcal{S}_\Delta$ ,  $\mathcal{S}_D$ , and  $\mathcal{S}_Q$  are selection matrices containing only ones and zeros such that  $\text{diag}(\Delta_t^2) = \mathcal{S}_\Delta \text{vech}(Q_t)$ ,  $\text{diag}(D_t^2) = \mathcal{S}_D f_t$ , and  $\text{vech}(Q_t) = \mathcal{S}_Q f_t$ . Using these matrix definitions, we obtain

$$\Psi_t = \mathcal{B}_k (\mathbf{I} \oplus D_t R_t) \cdot W_{D_t} \mathcal{S}_D + \mathcal{B}_k D_{t \otimes} \Delta_{t \otimes}^{-1} [\mathcal{D}_k - (\Delta_t \oplus Q_t) \Delta_{t \otimes}^{-1} W_{\Delta_t} \mathcal{S}_\Delta] \mathcal{S}_Q. \quad (20)$$

The updating equation (4) follows from Theorem 1 with  $\Psi_t$  given by (20). If only the correlations in  $R_t$  are time-varying,  $\mathcal{S}_D = 0$  and the first term in (20) drops out. Similarly, if only the volatilities are time-varying and the correlations are constant, (20) gives the appropriate specification by setting  $\mathcal{S}_Q = 0$  and the second term in (20) vanishes.

### 3.2 Comparisons with DCC

Our correlation updating based on the GAS model shares some features with the DCC updating based on (18). For example, the number of parameters in the model can be limited substantially by setting  $\omega$  in (4) equal to

$$\omega = (\mathbf{I} - B_1 - \dots - B_q) \cdot \text{vech}(\bar{Q}), \quad (21)$$

with  $\bar{Q}$  being equal to the (unconditional) sample correlation matrix. The matrix  $\bar{Q}$  is typically computed from the standardized (pre-filtered) residuals. In addition, the number of parameters can be reduced by having coefficient matrices in (4) as scaled identity matrices or diagonal matrices. The GAS model, however, does not reduce to the DCC model. This is mainly due to the presence of the weighting term  $w_t$  in (12). It limits the impact of observations corresponding to large values of  $y_t' \Sigma_t^{-1} y_t$  on the updating of correlations. Such a mechanism is absent in the DCC model. Our current specification therefore differs substantially from the DCC specification with Student's  $t$  distributed error terms. Also, when we consider the normal distribution ( $\nu^{-1} = 0$  and  $w_t \equiv 1$ ), our updating equation for the current specification  $f_t = \text{vech}(Q_t)$  does not reduce to the DCC model.

To illustrate our different updating equation compared to the one of DCC, we consider  $D_t = \mathbf{I}$  and return to the bivariate case  $k = 2$  from Section 2.4 which has led to equation (15). We then have

$$\Psi'_t = \frac{1}{2\sqrt{Q_{11,t}Q_{22,t}}} \begin{pmatrix} -Q_{12,t}/Q_{11,t} \\ 2 \\ -Q_{12,t}/Q_{22,t} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \quad (22)$$

where  $Q_{ij,t}$  is the  $(i, j)$  element of  $Q_t$  for  $i, j = 1, 2$ . The updating for all elements in  $f_t = \text{vech}(Q_t)$  is driven by the single driver (15) that is multiplied by the three elements in the first vector in (22). By contrast, the drivers in the DCC model for  $Q_{11,t}$ ,  $Q_{12,t}$  and  $Q_{22,t}$  are given

by  $y_{1t}^2$ ,  $y_{1t}y_{2t}$ , and  $y_{2t}^2$ , respectively. We thus have three different drivers for the DCC instead of one for the GAS model. We justify our single driver by the notion that matrix  $Q_t$  is an auxiliary parameter and its diagonal elements are not of core interest. The driver of the GAS model takes this into account by postulating that there is only one function (rather than three separate functions) of the data containing information on changes in the correlation parameter. The numerical intuition was provided at the end of Section 2.4.

The current specification  $f_t$  with  $\text{vech}(Q_t)$  has obvious drawbacks that it shares with DCC. In particular, the dimension of  $\mathcal{S}_Q f_t$  is higher than the number of unique elements in  $R_t$ , i.e.  $f_t$  contains an additional  $k$  factors. The Fisher information matrix for (17) is therefore singular and we need to replace  $\mathcal{I}_{t|t-1}^{-1}$  in (6) by the Moore-Penrose pseudo-inverse of  $\mathcal{I}_{t|t-1}$ , that is  $\mathcal{I}_{t|t-1}^+$ . Since the diagonal elements of  $Q_t$  are effectively redundant, its corresponding elements in  $\omega$  in (4) are not identified. In particular, the elements of  $\omega$  corresponding to the diagonal elements of  $Q_t$  can be multiplied by any arbitrary positive number without changing the decomposition. In our empirical work, this is resolved by using the normalization (21).

### 3.3 A time-varying copula specification

It is straightforward to transform the general framework of Section 3.1 to a copula specification. Assume that a univariate time series model for each series in the vector  $y_t$  has been estimated separately. Let  $u_{it}$  denote the probability integral transform of  $y_{it}$ , which is the  $i$ th series in  $y_t$ . The density (1) then operates on the  $k \times 1$  vector  $\tilde{y}'_t = [P_\nu^{-1}(u_{1t}), \dots, P_\nu^{-1}(u_{kt})]'$  via Sklar's theorem in Sklar (1959), where  $P_\nu^{-1}(\cdot)$  is the univariate inverse Student's  $t$  distribution with  $\nu$  degrees of freedom. In the copula specification, the decomposition (16) simplifies because  $D_t \equiv \mathbf{I}$ . We obtain

$$\Psi_t = \mathcal{B}_k D_{t \otimes} \Delta_{t \otimes}^{-1} [\mathcal{D}_k - (\Delta_t \oplus Q_t) \Delta_{t \otimes}^{-1} W_{\Delta t} \mathcal{S}_\Delta] \mathcal{S}_Q, \quad (23)$$

where  $\mathcal{S}_Q = \mathbf{I}$  because  $f_t = \text{vech}(Q_t)$ . Note that the marginal densities drop out from the expression for the GAS step  $s_t$  in Theorem 1 as they do not depend on  $R_t$ . They therefore vanish when taking derivatives of the log-density with respect to  $f_t$ . To evaluate the likelihood and estimate the parameters, however, the marginal densities have to be considered as well. The main reason is that they depend on the unknown parameter  $\nu$ , which has to be estimated.

## 4 Alternative specifications for the covariance matrix

In this section, we consider alternative specifications for the time-varying covariance matrix  $\Sigma_t$ . We also extend the basic specification by imposing a factor structure on the covariance matrix. In all cases, the only change necessary for the GAS driving mechanism  $s_t$  provided in Theorem 1 is the matrix  $\Psi_t$ .

### 4.1 Time-varying log-volatilities

To impose positive standard deviations in the decomposition (16), we can, for example, define the vector of factors  $f_t$  as

$$f_t = \begin{pmatrix} \log(\text{diag}(D_t^2)) \\ \text{vech}(Q_t) \end{pmatrix}. \quad (24)$$

In the log-variance specification with correlation matrix given by (17), we obtain

$$\Psi_t = \mathcal{B}_k(\mathbf{I} \oplus D_t R_t) W_{D_t} D_t^2 \mathcal{S}_D + \mathcal{B}_k D_{t \otimes} \Delta_{t \otimes}^{-1} [\mathcal{D}_k - (\Delta_t \oplus Q_t) \Delta_{t \otimes}^{-1} W_{\Delta_t} \mathcal{S}_\Delta] \mathcal{S}_Q \quad (25)$$

which is the same as  $\Psi_t$  in (20) except its first term is multiplied by  $D_t^2$ . The multiplication by  $D_t^2$  accounts for the log transformation in the derivative of  $\text{diag}(D_t^2)$  with respect to  $f_t$ . In the setting where  $R_t = R$ ,  $\mathcal{S}_D$  is the identity matrix and  $\mathcal{S}_Q = 0$ .

### 4.2 Time-varying correlations based on hyperspherical coordinates

In this section we provide an alternative specification for  $R_t$ . The variance matrix decomposition  $\Sigma_t = D_t R_t D_t$  remains as in (17). The difficulty with specifying a correlation matrix is that three necessary conditions are needed: (i) the matrix  $R_t$  has to be positive (semi) definite; (ii) the off-diagonal elements of  $R_t$  all lie in the interval  $[-1, 1]$ ; and (iii) the diagonal elements of  $R_t$  are equal to one for all values of  $t$ . To satisfy (i), we can adopt the Cholesky decomposition of  $R_t$  rather than  $R_t$  itself. The Cholesky decomposition, however, by itself does not automatically satisfy (ii) and (iii). An alternative specification and decomposition that satisfies all conditions (i)-(iii) simultaneously is based on hyperspherical coordinates and is given by

$$R_t = X_t' X_t,$$



where  $X_t$  is an upper-triangular matrix given by

$$X_t(\phi_t) = \begin{pmatrix} 1 & c_{12t} & c_{13t} & \dots & c_{1kt} \\ 0 & s_{12t} & c_{23t}s_{13t} & \dots & c_{2kt}s_{1kt} \\ 0 & 0 & s_{23t}s_{13t} & \dots & c_{3kt}s_{2kt}s_{1kt} \\ 0 & 0 & 0 & \dots & c_{4kt}s_{3kt}s_{2kt}s_{1kt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & c_{k-1,kt} \prod_{\ell=1}^{k-2} s_{\ell kt} \\ 0 & 0 & 0 & \dots & \prod_{\ell=1}^{k-1} s_{\ell kt} \end{pmatrix}, \quad (26)$$

with scalar  $c_{ijt} = \cos(\phi_{ijt})$ , scalar  $s_{ijt} = \sin(\phi_{ijt})$ , and scalar  $\phi_{ijt}$  as the time-varying angle measured in radians. The vector  $\phi_t$  contains the  $k(k-1)/2$  angles  $\phi_{ijt}$  for  $i, j = 1, \dots, k$ . The columns of  $X_t$  consist of hyperspherical coordinates with unit length. This decomposition has been used by Jaeckel and Rebonato (2000) with constant angles over time as a device to model term structure data. The GAS framework is ideally suited to allow the angles, and therefore the correlations, to be time-varying. In this case, we define the vector of factors as  $f'_t = [\text{diag}(D_t^2)', \phi'_t]$ . The general structure of Theorem 1 automatically provides the functions of the data to drive the angles  $\phi_t$  through the GAS updating variable  $s_t$ .

As before, the only change needed compared to the previous specifications is the form of  $\Psi_t$ . Below we use the derivatives

$$\frac{\partial x_{ijt}}{\partial \phi_{\ell mt}} = \begin{cases} 0 & \text{if } i > j, \ell \geq m, \ell \geq i, \text{ or } j \neq m, \\ -x_{ijt} \cdot \tan(\phi_{ijt}) & \text{if } i = \ell \text{ and } i \neq j, \\ x_{ijt} / \tan(\phi_{\ell jt}) & \text{otherwise,} \end{cases} \quad (27)$$

for  $i, j, \ell, m = 1, \dots, k$  and where  $x_{ijt}$  is the  $(i, j)$  element of  $X_t$ . These derivatives induce the elements in the  $k^2 \times [k(k-1)/2]$  derivative matrix

$$Z_t = \frac{\partial \text{vec}(X_t)}{\partial \phi'_t}.$$

In the GAS framework, define the selection matrix  $\mathcal{S}_\phi$  such that  $\phi_t = \mathcal{S}_\phi f_t$ . Also note that

$$\frac{\partial \text{vec}(R_t)}{\partial f_t'} = \frac{\partial \text{vec}(X_t' X_t)}{\partial \phi_t'} \mathcal{S}_\phi = [(\mathbf{I} \otimes X_t') + (X_t' \otimes \mathbf{I}) \mathcal{C}_k] Z_t \mathcal{S}_\phi. \quad (28)$$

with the commutation matrix  $\mathcal{C}_k$  as defined in (10). Combining these results, we obtain

$$\Psi_t = \mathcal{B}_k (\mathbf{I} \oplus D_t R_t) W_{D_t} \mathcal{S}_D + \mathcal{B}_k D_{t \otimes} [(\mathbf{I} \otimes X_t') + (X_t' \otimes \mathbf{I}) \mathcal{C}_k] Z_t \mathcal{S}_\phi. \quad (29)$$

The decomposition of the correlation matrix into hyperspherical coordinates can also easily be combined with the log-variance specification of the volatility matrix discussed in section 4.1.

The hyperspherical specification has the advantage relative to the decomposition of  $R_t$  in (17) that the number of correlations in  $R_t$  is the same as the number of unique elements in  $X_t$ . As a result, all parameters in the vector  $\omega$  are identified and the scaling matrix  $S_t$  in (6) is nonsingular. Also, when the dimension  $k$  of  $y_t$  increases, the number of elements in  $f_t$  for the hyperspherical specification is smaller than for the specification (17). A possible disadvantage relative to (17) is the interpretation of the factors as angles. Each angle is only identified within the region  $[0, 2\pi]$ . This constraint can be imposed on the factors via a transformation. However, in practical cases our experience is that numerical problems do not occur when such constraints are omitted.

### 4.3 Common dynamic factors

In our time-varying specifications, the number of factors coincides with the number of time-varying volatilities and correlations in the model. In the correlation specification (17) the number of factors is even larger than the number of correlations. However, the time-varying features may be shared between different volatilities and different correlations (indeed, even between volatilities and correlations). To allow for common dynamic features, we can impose a dynamic factor structure on the volatilities and correlations. Consider partitioning  $f_t$  as in (19), that is  $f_t' = [\text{diag}(D_t^2)', \text{vech}(Q_t)']$ . With a slight abuse of notation, we have

$$\text{diag}(D_t^2) = a + \mathcal{S}_D f_t, \quad \text{vech}(Q_t) = b + \mathcal{S}_Q f_t, \quad (30)$$

with vectors  $a$  and  $b$ , and where  $\mathcal{S}_D$  and  $\mathcal{S}_Q$  are defined as full matrices rather than selection matrices. We assume that all vectors and matrices have the appropriate dimensions, are fixed, may depend on unknown coefficients, and are subject to identification restrictions. Extensions to non-linear factor structures are straightforward. The form of  $\Psi_t$  is now the same as in (20). For the hyperspherical coordinate specification, the result remains (29) with  $\mathcal{S}_\phi$  specified as a real-valued matrix rather than a selection matrix consisting of zeros and ones. For multivariate GARCH models, factor structures have been imposed by Tsay (2005) and by Bauwens et al. (2006) in their survey article.

A low-dimensional  $f_t$  reduces the number of time-varying factors but it typically raises the number of unknown coefficients in the factor loading matrices  $\mathcal{S}_D$  and  $\mathcal{S}_Q$  or  $\mathcal{S}_\phi$ . In the multivariate GARCH cases, it is common practice to estimate the constant vectors  $a$  and  $b$  initially using sample variances and sample correlations. Furthermore, it is possible to impose additional structure on the correlation matrix  $R_t$ . For example, a specification of  $R_t$  similar to the dynamic equicorrelation (DECO) model of Engle and Kelly (2009) can be considered. In our case the equicorrelation structure can be enforced by having a single factor driving all the correlations and having unit values in the matrix (here, vector)  $\mathcal{S}_Q$ . This structure imposed on  $R_t$  may introduce both statistical and computational advantages when modeling very high-dimensional systems.

## 5 Estimation and Monte Carlo evidence

In this section we carry out a Monte Carlo study to investigate the performance of our modeling framework. We verify whether parameter estimation can be successful without knowing the evolution of the time-varying factor  $f_t$ . We first briefly discuss the estimation of fixed and static parameters in the model using the method of maximum likelihood. Then the Monte Carlo design is presented and our results are discussed.

## 5.1 Parameter estimation

Given a set of  $n$  realizations for the observation vector  $y_t$  with mean zero and for  $t = 1, \dots, n$ , the log-likelihood function for the multivariate Student's  $t$  model is given by

$$\begin{aligned} \mathcal{L} = \sum_{t=1}^n \left\{ \log \left[ \Gamma \left( \frac{\nu + k}{2} \right) \right] - \log \left[ \Gamma \left( \frac{\nu}{2} \right) \right] - \frac{1}{2} \log |\Sigma_t| \right. \\ \left. - \frac{k}{2} \log [(\nu - 2)\pi] - \frac{(\nu + k)}{2} \log \left[ 1 + \frac{y_t' \Sigma_t^{-1} y_t}{(\nu - 2)} \right] \right\}, \end{aligned} \quad (31)$$

where the time variation of  $\Sigma_t$  is determined by the GAS updating equations for  $D_t$  and  $R_t$ . For given values of  $\nu$  in (1) and of  $\omega$ ,  $A_i$  and  $B_i$  in (4), the log-likelihood function can be evaluated in a straightforward way. The unknown coefficients are collected in the parameter vector  $\theta$  and its estimation is based on the maximization of (31) with respect to  $\theta$ . Maximization can take place via a standard quasi-Newton numerical optimization procedure.

## 5.2 Monte Carlo study

The design of our Monte Carlo study is similar to the study conducted by Engle (2002). We simulate a series of  $n = 1000$  observations from a bivariate Student's  $t$  distribution with unit variance and time-varying correlation  $\rho_t$ . The following time-varying patterns for  $\rho_t$  will be considered:

1. Constant      0.9;
2. Sine             $0.5 + 0.4 \cos(2\pi t / 200)$ ;
3. Fast Sine       $0.5 + 0.4 \cos(2\pi t / 20)$ ;
4. Step             $0.9 - 0.5 (t > 500)$ ;
5. Ramp             $\text{mod}(t / 200)$ ;
6. Model            $\exp(h_t) / [1 + \exp(h_t)]$  where  
 $h_t = -0.4(1 - 0.99) + 0.99h_{t-1} + 0.14\eta_t, \quad \eta_t \sim \mathcal{N}(0, 1)$ .

Graphs depicting each of these patterns can be found in Engle (2002). Given one of these patterns, the data generation process is given by

$$y_t \sim p(y_t | \Sigma_t; \nu), \quad \Sigma_t = R_t = \frac{\nu - 2}{\nu} \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}, \quad \nu = 5, \quad (32)$$

where  $p(y_t|\Sigma_t; \nu)$  is the bivariate Student's  $t$  density (1) with  $k = 2$ . In the Monte Carlo study, we simulate 500 bivariate series for each of the six time-varying patterns  $\rho_t$ .

For a realized bivariate time series, we consider the model with density (1) and the GAS(1, 1) recursion (4) for  $f_t$ . We consider both the specification of  $R_t$  given by (17) and given by the hyperspherical coordinates in (26). For both models, we estimate the unknown parameters  $\nu$ ,  $A_1$  and  $B_1$  by maximum likelihood while  $\omega$  is estimated using the (unconditional) sample correlation. Given the parameter estimates, we obtain  $\hat{\rho}_t$  from  $f_t$  using the GAS(1, 1) updating equation (4). Our two measures of accuracy are the mean absolute error (MAE) and the mean squared error (MSE) as given by

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |\hat{\rho}_t - \rho_t|, \quad \text{MSE} = \frac{1}{n} \sum_{t=1}^n (\hat{\rho}_t - \rho_t)^2.$$

Once the two GAS specifications are treated for a simulated series  $y_t$ , we consider alternative approaches of estimating  $\rho_t$  for comparison purposes. In particular, we benchmark the two GAS models against the DCC model. Also, we consider as a benchmark the exponentially weighted moving average (EWMA) recursion for  $Q_t$  in (17) given by

$$Q_{t+1} = \lambda Q_t + (1 - \lambda) y_t y_t', \quad (33)$$

where the smoothing parameter is set to  $\lambda = 0.96$ . Given the insights obtained from the GAS updating and the impact of the weighting term  $w_t$ , our final benchmark is an adjusted EWMA scheme with an additional penalty term, that is

$$Q_{t+1} = \lambda Q_t + (1 - \lambda) \frac{(\nu^* + 2)}{\nu^* - 2 + y_t' \Sigma_t^{-1} y_t} y_t y_t', \quad (34)$$

with  $\nu^* = 10$ .

The results from this Monte Carlo study are presented in Table 1, where we take the DCC model as the benchmark and report all MAE and MSE values relative to it. The reported MAE and MSE values are based on the means of the MAE and MSE for each Monte Carlo repetition. The GAS(1, 1) model with the hyperspherical specification (26) appears to be preferred to the other models for 5 out of the 6 different correlation paths. The improvements in MSE or MAE

Table 1: *Mean absolute error and mean squared error results: in-sample*

The table contains the mean absolute error (MAE) and mean squared error (MSE) for the estimated dynamic correlation patterns for six different models and six different correlation processes. The details of the  $t$ -GAS method are in Section 3.1. The  $t$ -GAS-h method is based on the hyperspherical coordinates specification of Section 4.2. The results for  $t$ -DCC are obtained from the method of Engle (2002) using a Student's  $t$  likelihood function. The methods EWMA and adj-EWMA refer to equations (33) and (34), respectively. The MAE and MSE measures are presented relative to the  $t$ -DCC model.

	Constant	Sine	Fast Sine	Step	Ramp	Model
MAE						
$t$ -GAS	0.7909	0.9159	1.0320	0.8954	<b>0.9736</b>	0.9233
$t$ -GAS-h	<b>0.7455</b>	<b>0.9066</b>	1.0455	<b>0.8844</b>	0.97786	<b>0.9224</b>
$t$ -DCC	1.0000	1.0000	<b>1.0000</b>	1.0000	1.0000	1.0000
EWMA	3.2364	1.2473	1.1243	1.0243	1.1138	1.0750
adj-EWMA	2.3545	1.2608	1.0973	0.8966	1.1396	0.99041
MSE						
$t$ -GAS	0.54545	0.88406	1.0391	0.88652	0.98020	0.85446
$t$ -GAS-h	<b>0.5000</b>	<b>0.8579</b>	1.0560	<b>0.8582</b>	<b>0.9743</b>	<b>0.8499</b>
$t$ -DCC	1.0000	1.0000	<b>1.0000</b>	1.0000	1.0000	1.0000
EWMA	11.909	1.4348	1.3229	1.0638	1.2020	1.1690
adj-EWMA	5.1818	1.4087	1.2227	0.8794	1.2693	0.9765

are typically in the range of 10% to 15% improvement in forecasting accuracy, except for the ramp and fast sine DGP. The results for the GAS(1, 1) model based on (17) are second best. The DCC model only performs best in terms of minimizing MAE and MSE when the true DGP is the Fast Sine.

The results from Table 1 can be regarded as indicative of in-sample performance of the different models. To verify the different approaches in terms of short-term forecasting, we elaborate upon the Monte Carlo study as follows. We extend the simulation of a bivariate series to 1005 observations from which we use the first  $n = 1000$  for parameter estimation and the last 5 for out-of-sample forecasting. After estimation, the out-of-sample forecasts for  $\rho_{n+j}$  from each model are computed for  $j = 1, \dots, 5$ . In case of the GAS model, the forecasts are obtained from (4) with  $t = n + 1, n + 2, \dots$  and with  $s_{n+j} = 0$  for  $j = 1, 2, \dots$ . The forecast performance measures are based on the absolute and squared errors as given by

$$AE_f(h) = |\hat{\rho}_{n+h} - \rho_{n+h}|, \quad SE_f(h) = (\hat{\rho}_{n+h} - \rho_{n+h})^2,$$

for forecast horizon  $h = 1, \dots, 5$ . The results are presented in Table 2 for forecast horizons 1 and 5 days. The reported  $\text{MAE}_f(h)$  and  $\text{MSE}_f(h)$  values are the means of  $\text{AE}_f(h)$  and  $\text{SE}_f(h)$  over the 500 Monte Carlo repetitions, respectively.

The forecast results in Table 2 show that the  $t$ -GAS models remain convincingly superior for the Constant, Sine and Model correlation patterns. The performance of the  $t$ -GAS model for constant correlations over time is convincing. It implies that the  $t$ -GAS models is robust to model misspecification when time-varying correlations are imposed while the underlying correlations are constant. The forecasts of Fast Sine correlation patterns is most successful for the  $t$ -DCC model for  $h = 1$  although the  $t$ -GAS performance is close and best for  $\text{MSE}_f(5)$ . The results for Step and Ramp patterns are close for all models with the adjusted EWMA method being most successful.

## 6 Empirical illustration: global equity returns

In this section, we adopt the models developed in Sections 2 to 4 for a panel of daily global equity returns from January 1, 1989 to March 18, 2009 obtained from Datastream. The data set consists of 5,272 time series observations for the six countries Australia, France, Germany, Hong Kong, the United Kingdom and the United States. The data are in U.S. Dollars and have been adjusted for dividends and splits.

In the empirical study we consider GARCH/DCC and GAS classes of models. For the GARCH/DCC models, we specify GARCH models for the marginal variances of each series and a DCC model for the correlation process between series. The long-run variances and correlations within each of the GARCH and DCC models are pre-estimated using the sample variances and sample correlations for the entire series. All remaining parameters of the model are estimated jointly in one-step. Each individual GARCH model has two parameters governing its dynamics (these are traditionally labeled as  $\alpha$  and  $\beta$  as in the discussion below equation (7).) The DCC model was estimated with two common parameters as well; one parameter for each matrix  $A_{\text{dcc}}$  and  $B_{\text{dcc}}$  in (18). The variance and correlation parameters of the GAS model are estimated simultaneously as follows. The first factors in vector  $f_t$  represent the variances; for example,  $\text{diag}(D_t^2)$  in the decomposition (16). The remaining factors in  $f_t$  represent the correlations, either by means of elements in  $Q_t$  as defined in (17) or by means of hyperspherical angles in  $\phi_t$

Table 2: *Mean absolute error and mean squared error results: out-of-sample*

The table contains the mean absolute error ( $\text{MAE}_f(h)$ ) and mean squared error ( $\text{MSE}_f(h)$ ) for the correlation forecasts from six different models and six different correlation processes; see Table 1 for further details. The results are presented for the forecast horizons of 1 and 5 days. The measures are reported relative to the measure of  $t$ -DCC.

	Constant	Sine	Fast Sine	Step	Ramp	Model
$\text{MAE}_f(1)$						
$t$ -GAS	0.6948	<b>0.6597</b>	1.0147	1.0548	1.0050	0.9245
$t$ -GAS-h	<b>0.6392</b>	0.6781	1.0400	1.0352	1.0173	0.9251
$t$ -DCC	1	1	<b>1</b>	1	1	1
EWMA	3.0218	1.2751	1.0762	1.1567	0.9752	0.9901
adj-EWMA	2.1865	1.3202	1.0452	<b>0.9901</b>	<b>0.9733</b>	<b>0.9244</b>
$\text{MSE}_f(1)$						
$t$ -GAS	0.4302	<b>0.3963</b>	1.0075	1.1151	1.0057	<b>0.8628</b>
$t$ -GAS-h	<b>0.3681</b>	0.4338	1.0436	1.0832	1.0308	0.8642
$t$ -DCC	1	1	<b>1</b>	1	1	1
EWMA	11.1348	1.4087	1.2333	1.3424	0.9507	1.0294
adj-EWMA	4.6311	1.3721	1.1206	<b>0.9935</b>	<b>0.9446</b>	0.8759
$\text{MAE}_f(5)$						
$t$ -GAS	0.7030	0.6862	0.9067	1.0547	<b>0.9832</b>	0.9303
$t$ -GAS-h	<b>0.6559</b>	<b>0.6824</b>	<b>0.9094</b>	1.0318	1.0037	<b>0.9287</b>
$t$ -DCC	1	1	1	1	1	1
EWMA	3.1351	1.0705	1.1705	1.1620	0.9991	1.0003
adj-EWMA	2.2685	1.1083	1.0050	<b>0.9946</b>	0.9971	0.9409
$\text{MSE}_f(5)$						
$t$ -GAS	0.4389	<b>0.4227</b>	0.7958	1.1101	<b>0.9603</b>	0.8703
$t$ -GAS-h	<b>0.3878</b>	0.4387	<b>0.7906</b>	1.069	1.0011	<b>0.8699</b>
$t$ -DCC	1	1	1	<b>1</b>	1	1
EWMA	12.1572	1.0889	1.9560	1.3551	0.9957	1.0327
adj-EWMA	5.0565	1.0583	1.3752	1.0029	0.9890	0.8934

as defined below (26). We specify  $f_t$  by the GAS(1, 1) model with diagonal coefficient matrices  $A_1$  and  $B_1$  where their first diagonal elements have different values for different variances while the remaining diagonal elements associated with the correlations have the same value. The long-run variances and correlations for all the GAS models are pre-estimated using the sample variances and sample correlations for the entire series, just as in the GARCH/DCC models described above. In our example with six series, the diagonal elements of  $A_1$  and  $B_1$  are then given by  $a_1, \dots, a_6, a_7, a_7, \dots, a_7$  and  $b_1, \dots, b_6, b_7, b_7, \dots, b_7$ , respectively.

Table 3 reports the estimated parameters as well as the (maximised) log-likelihood values, the Akaike (AIC) and Schwartz's Bayesian (BIC) information criteria for eight different models.



In the DCC case, the parameters  $a_1, \dots, a_6$  and  $b_1, \dots, b_6$  relate to the univariate GARCH models while  $a_7 = a_{\text{dcc}}$  and  $b_7 = b_{\text{dcc}}$  as discussed below (18). The first two columns in Table 3 present the estimates for the DCC model with Gaussian and the Student's  $t$  likelihood specifications but with the same DCC updating (18). As expected for stock return data, the Student's  $t$  specification produces a huge improvement in the log-likelihood value at its maximum.

The remaining columns of Table 3 consider the GAS(1, 1) model with different densities and different specifications of the variance matrix. The columns with headings  $g$ -GAS and  $t$ -GAS are based on Gaussian and Student's  $t$  densities, respectively, with variance specification (16) and (17), and with vector  $f_t$  given by (19). The number of parameters are the same as for the DCC model. The  $g$ -GAS model performs worse when compared to its DCC counterpart while the  $t$ -GAS model performs best when compared to the DCC and  $g$ -GAS model. Hence we conclude that the assumption of normality is clearly rejected by the data. The  $t$ -GAS specification outperforms the  $t$ -DCC model by more than 60 points with the same number of parameters. It illustrates the flexibility of the  $t$ -GAS model to adapt itself to changing volatilities and correlations. The estimated autoregressive parameter values  $b_i$ ,  $i = 1, \dots, 7$ , for the  $t$ -GAS model are all very close to unity. It indicates that the estimated factors from  $t$ -GAS are highly persistent. These results also indicate that a more parsimonious model may be obtained by restricting the coefficients  $b_1, \dots, b_6$  to be equal.

The fifth, sixth and seventh columns in Table 3 present the estimation results for GAS models with three alternative specifications for the vector  $f_t$  and/or the variance matrix. The model  $t$ -GAS-h refers to the  $t$ -GAS model with a hypersphere specification of the variance matrix  $\Sigma_t = D_t X_t' X_t D_t$  where  $X_t$  is given by (26) in Section 4.2. The  $t$ -GAS-l model is the  $t$ -GAS model as described above but with the log-volatility factors of Section 4.1. The  $t$ -GAS-hl model is the combination of the two. On the basis of the log-likelihood values, the combination  $t$ -GAS-hl model appears to work best. The log-likelihood improvement in comparison with  $t$ -GAS is small. We therefore conclude that for our data set of daily returns, the variance matrix specification is less important for the formulation of volatility and correlation dynamics.

The estimated correlations for a selection of three return series Germany, Hong Kong, and United States are presented in the left-hand side plots in Figure 1. The right-hand side plots present the three observed return series. The two series of estimated correlations in each left-hand side plot are based on  $g$ -GAS-h and  $t$ -GAS-h models with correlation matrix  $R_t = X_t' X_t$

Table 3: *Parameter estimates, Likelihoods, and Information Criteria*

The table contains the estimated parameters and their standard errors for nine alternative models including the DCC. The prefix “*t*” on *t*-DCC and *t*-GAS is for a Student’s *t* density while “*g*” is for the Gaussian density. The hypersphere specification of  $R_t$  is denoted by the suffix ‘h’ while the logarithmic specification for the variances is denoted by the additional suffix ‘l’. The prefix “*tg*” on *tg*-GAS denotes a GAS model where the likelihood is a Student’s *t* density, while the factor recursion in (4) is based on the Gaussian ( $\nu^{-1} = 0$ ) density. The first six parameters  $a_i$  and  $b_i$  relate to the volatility dynamics. The parameters  $a_7$  and  $b_7$  relate to the correlation dynamics.

	<i>g</i> -DCC	<i>t</i> -DCC	<i>g</i> -GAS	<i>t</i> -GAS	<i>t</i> -GAS-h	<i>t</i> -GAS-l	<i>t</i> -GAS-h-l	<i>tg</i> -GAS
$a_1$	0.0678 (0.0078)	0.0437 (0.0055)	0.0632 (0.0068)	0.0405 (0.0046)	0.0405 (0.0046)	0.0365 (0.0038)	0.0366 (0.0039)	0.0401 (0.0049)
$a_2$	0.0581 (0.0046)	0.0411 (0.0040)	0.0522 (0.0042)	0.0374 (0.0032)	0.0376 (0.0032)	0.0341 (0.0027)	0.0343 (0.0027)	0.0366 (0.0035)
$a_3$	0.0612 (0.0046)	0.0444 (0.0041)	0.0560 (0.0042)	0.0410 (0.0033)	0.0412 (0.0033)	0.0371 (0.0028)	0.0372 (0.0028)	0.04 (0.0037)
$a_4$	0.0930 (0.0070)	0.0663 (0.0061)	0.0831 (0.0058)	0.0503 (0.0044)	0.0504 (0.0045)	0.0484 (0.0041)	0.0484 (0.0041)	0.0582 (0.0052)
$a_5$	0.0651 (0.0058)	0.0439 (0.0048)	0.0594 (0.0052)	0.0392 (0.0038)	0.0394 (0.0038)	0.0365 (0.0034)	0.0366 (0.0034)	0.0396 (0.0043)
$a_6$	0.0643 (0.0068)	0.0498 (0.0051)	0.0595 (0.0059)	0.0459 (0.0040)	0.0459 (0.0040)	0.0437 (0.0037)	0.0437 (0.0037)	0.0465 (0.0046)
$a_7$	0.0087 (0.0008)	0.0077 (0.0007)	0.0054 (0.0003)	0.0079 (0.0007)	0.0080 (0.0007)	0.0081 (0.0007)	0.0081 (0.0007)	0.0055 (0.0003)
$b_1$	0.9214 (0.0091)	0.9475 (0.0068)	0.9837 (0.0032)	0.9920 (0.0020)	0.9920 (0.0020)	0.9924 (0.0018)	0.9924 (0.0018)	0.9892 (0.0023)
$b_2$	0.9351 (0.0054)	0.9520 (0.0048)	0.9894 (0.0018)	0.9934 (0.0013)	0.9932 (0.0013)	0.9920 (0.0014)	0.9918 (0.0014)	0.9918 (0.0015)
$b_3$	0.9322 (0.0052)	0.9498 (0.0047)	0.9891 (0.0017)	0.9936 (0.0012)	0.9935 (0.0012)	0.9919 (0.0014)	0.9917 (0.0014)	0.9923 (0.0014)
$b_4$	0.9026 (0.0072)	0.9294 (0.0064)	0.9856 (0.0021)	0.9945 (0.0012)	0.9944 (0.0012)	0.9915 (0.0017)	0.9915 (0.0017)	0.9911 (0.0016)
$b_5$	0.9287 (0.0063)	0.9499 (0.0056)	0.9891 (0.0018)	0.9935 (0.0013)	0.9934 (0.0013)	0.9923 (0.0015)	0.9921 (0.0015)	0.9921 (0.0015)
$b_6$	0.9329 (0.0070)	0.9482 (0.0053)	0.9930 (0.0013)	0.9960 (0.0009)	0.9960 (0.0009)	0.9945 (0.0013)	0.9944 (0.0013)	0.9955 (0.0010)
$b_7$	0.9897 (0.0010)	0.9912 (0.0009)	0.9988 (0.0002)	0.9988 (0.0003)	0.9987 (0.0003)	0.9987 (0.0003)	0.9986 (0.0003)	0.9991 (0.0002)
$\nu$	-	8.2062 (0.3062)	-	8.9238 (0.3714)	8.9261 (0.3716)	8.8269 (0.3650)	8.8243 (0.3648)	8.1122 (0.2979)
Log-likelihood	-40781.6	-39841.2	-40824.1	-39779.5	-39776.3	-39777.8	-39774.9	-39869.0
AIC	81591.2	79712.5	81676.2	79589.0	79582.6	79585.6	79579.9	79767.9
BIC	81683.2	79811.0	81768.2	79687.5	79681.2	79684.1	79678.4	79866.5

of Section 4.2. They illustrate the differences between GAS models with Gaussian and with Student’s *t* densities for the errors. Several large outliers are present in the return series during the 1990’s. In this period the correlation estimates from the two GAS models are substantially different. For both GAS models, the estimated correlations appear to be volatile. More smooth

estimates can be obtained by increasing  $q$  and by setting the coefficient matrices  $A_1, \dots, A_q$  equal to each other in the GAS model (4).

In Figure 1, the outliers appear to have a strong effect on the estimated correlations when based on the Gaussian density whereas correlations based on the Student's  $t$  density appear to be robust against the outliers. The major difference in the factor recursions between the two models is the weighting factor  $w_t = (\nu + k) / (\nu - 2 + y_t' \Sigma_t^{-1} y_t)$ , which is part of the score in the Student's  $t$  model. At the end of the sample, we observe fewer outliers in the individual time series. As a result, the differences between the estimated correlations from the different model specifications become smaller. However, differences remain visible for the Germany-US correlations during the years 2002, 2003, and 2008.

An important question relates to the contribution of the new factor recursion under the Student  $t$  distribution to the fit of the model. It can be argued that the improved fit compared to the Gaussian case is only due to the switch to the Student's  $t$  distribution rather than to the new form of the GAS factor recursion based on Theorem 1. To verify this, we consider a final model which is referred to as the  $tg$ -GAS model. The likelihood function is based on the Student's  $t$  density while the recursion for the factor is based on the Gaussian density. In effect, we adopt the specification of  $s_t$  from the expressions in Theorem 1 for  $\nu^{-1} = 0$  rather than for the estimated (finite) value of  $\nu$  from the likelihood maximization. The estimation results are presented in the last column of Figure 3. Comparing the results of  $t$ -GAS with those of  $tg$ -GAS, we find that the new form of the Student's  $t$ -based GAS recursion substantially contributes to the model fit. The log-likelihood value is increased approximately by 90 points with the same number of parameters. The log-likelihood value of the  $tg$ -GAS model is closer to the one of the  $t$ -DCC model. We conclude that it is not only important to account for fat-tails of the error distribution, but also for the effect of the fat-tails on volatility and correlation dynamics.

Finally, the estimated coefficients  $b_1, \dots, b_7$  have lower values for the  $tg$ -GAS specification compared to those of the  $t$ -GAS model. This finding is probably due to large errors that have an impact on the Gaussian factor recursions. Since incidental large squared errors enter the Gaussian factor recursion without the weight  $w_t$  in Theorem 1, the maximum likelihood procedure downplays their impact on future volatilities and correlations by reducing the persistence parameters  $b_1, \dots, b_7$ . By contrast, the weight  $w_t$  plays a role in the Student's  $t$ -based factor recursion and, hence, it takes care of these observations in a natural way. As a consequence,

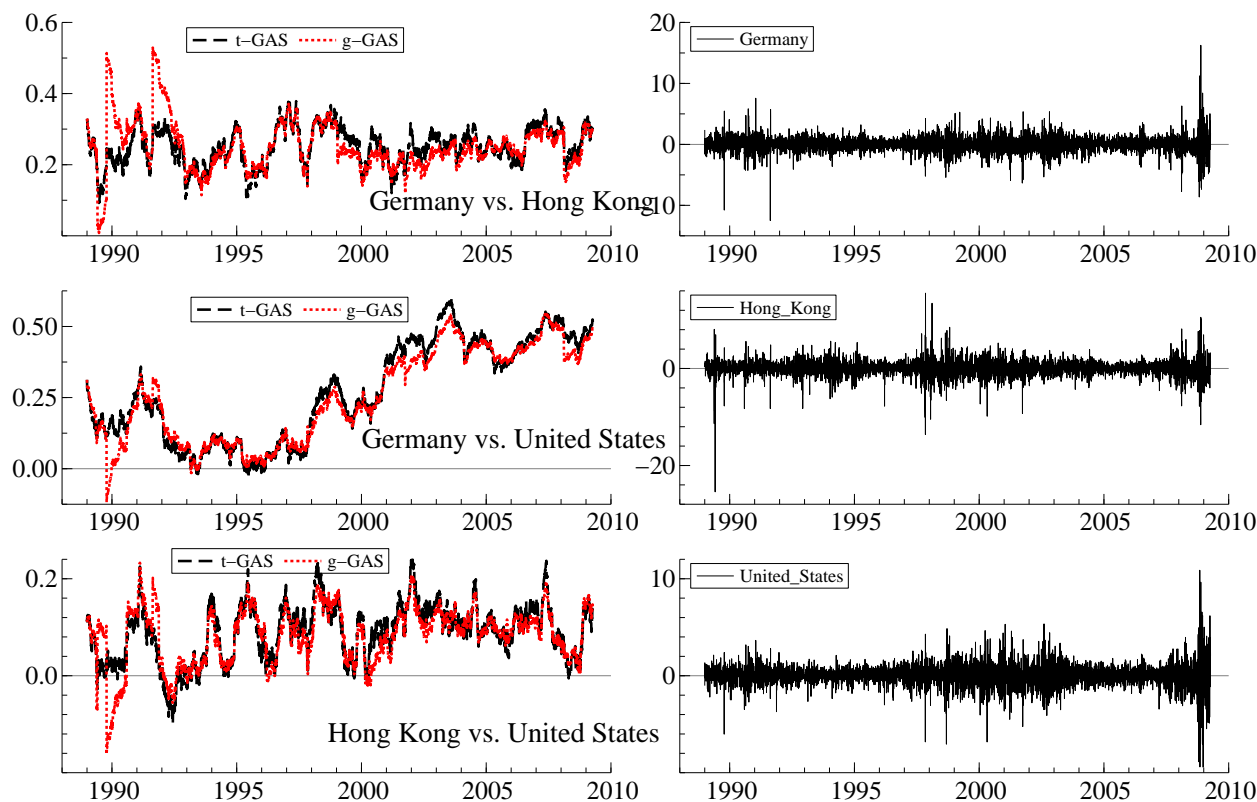


Figure 1: *Estimated correlations*

The left hand panels contain a comparison of the estimated correlations from the  $t$ -GAS(1,1) copula model versus the  $g$ -GAS(1,1) copula model both with the hypersphere specification of  $R_t$ . The three rows contain the correlations between the different possible pairs of Germany, Hong Kong, and the United States. The right-hand panels display the return data.

there is no further need to reduce the persistence parameters to increase the fit of the model. It underlines the importance of a robust form of the factor updating equation as illustrated by the comparison between the  $t$ -GAS and the  $tg$ -GAS models.

## 7 Conclusion

We have introduced the multivariate Student's  $t$  GAS model for volatilities and correlations, where the multivariate normal distribution is included as a special case. The models include two levels of flexibility. First, we propose a copula version of the model such that the marginal

distribution of the individual time series can be treated separately from the correlations. It enables the optimization problem to be broken into more manageable pieces. Secondly, we show how the GAS framework can accommodate alternative specifications of the covariance matrix. The model formulation is sufficiently general to impose a factor structure on either the time-varying volatilities or correlations, or both. We have focused on the decomposition of the conditional covariance matrix into a volatility matrix and a correlation matrix. In future we can explore alternative decompositions of the covariance matrix. For example, Cholesky or spectral decompositions will lead to new GAS formulations for dynamic volatilities and correlations. Another extension is to consider the multivariate skewed Student's  $t$  distribution that is recently proposed by Bauwens and Laurent (2005). By incorporating asymmetry in the updating recursion for the correlations, we account for the notion of leverage in financial returns. Such and other extensions provide interesting avenues for further research.

## Acknowledgments

Financial support to the first author from the University of Chicago Booth School of Business is gratefully acknowledged. All the computations reported in this paper were carried out using the OxMetrix 6.0 programming environment of Doornik (2009).

## References

- Abadir, K. M. and J. R. Magnus (2005). *Matrix Algebra*. New York, NY: Cambridge University Press.
- Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 23(3), 346–354.
- Bauwens, L., S. Laurent, and J. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21(1), 79–109.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* 69(3), 542–547.

- Chib, S., F. Nardari, and N. Shephard (2006). Analysis of high-dimensional multivariate stochastic volatility models. *Journal of Econometrics* 134(2), 341–371.
- Creal, D. D., S. J. Koopman, and A. Lucas (2010). Generalized Autoregressive Score Models with Applications. Working paper.
- Ding, Z. and R. F. Engle (2001). Large scale conditional covariance matrix modeling, estimation and testing. *Academia Economic Papers* 29, 157–184.
- Doornik, J. A. (2009). *Object-Oriented Matrix Programming using Ox 6.0*. London: Timberlake Consultants Press.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.
- Engle, R. F. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20(3), 339–350.
- Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131, 3–27.
- Engle, R. F. and B. Kelly (2009). Dynamic equicorrelation. Unpublished manuscript, Stern School of Business, New York University.
- Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66(5), 1127–1162.
- Engle, R. F. and K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Unpublished manuscript, University of California San Diego.
- Fan, J., M. Wang, and Q. Yao (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society, Series B* 70(4), 679–702.
- Gourieroux, C., J. Jasiak, and R. Sufana (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics* 150(2), 167–181.
- Harvey, A. C. and T. Chakravarty (2008). Beta-t-(E)GARCH. Unpublished manuscript, Faculty of Economics, Cambridge University.
- Jaeckel, P. and R. Rebonato (1999/2000, Winter). The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk* 2(2), 17–28.
- Nelson, D. B. and D. P. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica* 62(1), 1–41.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* 8, 229–231.
- Tsay, R. S. (2005). *Analysis of Financial Time Series* (Second ed.). New York: John Wiley & Sons Press.
- Tse, Y. K. and A. K. C. Tsui (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 20, 351–362.

# Appendix

The following two results are straightforward to prove.

**Result 1:** Let  $z \in \mathbb{R}^k$  and  $z \sim N(0, a\mathbf{I})$  for some scalar  $a > 0$ , then the fourth (cross-)moments are given by

$$\mathbb{E}[z_i z_j z_\ell z_m] = a^2 (\delta_{ij}\delta_{\ell m} + \delta_{i\ell}\delta_{jm} + \delta_{im}\delta_{j\ell}), \quad \text{for } i, j, \ell, m = 1, \dots, k,$$

where  $z_i$  denotes the  $i$ th element of  $z$  and where Kronecker delta  $\delta_{ij}$  is unity when  $i = j$  and zero otherwise. The  $k^2 \times k^2$  matrix  $G$  is implicitly defined by

$$\mathbb{E}[(zz')_{\otimes}] = a^2 G,$$

where the element  $G[\cdot, \cdot]$  of matrix  $G$  is given by

$$G[(i-1) \cdot k + \ell, (j-1) \cdot k + m] = \delta_{ij}\delta_{\ell m} + \delta_{i\ell}\delta_{jm} + \delta_{im}\delta_{j\ell}, \quad (35)$$

for  $i, j, \ell, m = 1, \dots, k$ .

**Result 2:** Let the random variable  $u \sim \chi_\nu^2$  where  $\chi_\nu^2$  is the Chi-square distribution with  $\nu$  degrees of freedom. For any scalar  $a < \nu$ , we have

$$\mathbb{E}\left[\left(\frac{\nu}{u}\right)^{a/2}\right] = \frac{\Gamma\left(\frac{\nu-a}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2}\right)^{a/2}. \quad (36)$$

**Proof of Theorem 1:** From the basic matrix calculus results in Abadir and Magnus (2005) and based on the notation and definitions in section 2.4, it is straightforward to derive (12) as

$$\begin{aligned} \frac{\partial \log p(y_t | \Sigma_t; \nu)}{\partial \text{vech}(\Sigma_t)} &= \frac{1}{2} \mathcal{D}'_k (\Sigma_t^{-1} \otimes \Sigma_t^{-1}) \left[ \frac{(\nu + k)}{(\nu + y_t' \Sigma_t^{-1} y_t - 2)} y_t \otimes y_t - \text{vec}(\Sigma_t) \right] \\ &= \frac{1}{2} \mathcal{D}'_k J'_{t \otimes} [w_t \bar{y}_t \otimes - \text{vec}(\mathbf{I})], \quad w_t = \frac{(1 + \nu^{-1}k)}{(1 + \nu^{-1}(\bar{y}_t' \bar{y}_t - 2))}, \end{aligned} \quad (37)$$

where  $\bar{y}_t = J_t y_t$  follows a Student's  $t$  distribution with zero mean, covariance matrix  $\mathbf{I}_k$ , and  $\nu$  degrees of freedom. The derivative of  $\text{vech}(\Sigma_t)$  with respect to  $f_t$  is denoted by  $\Psi_t$  and completes the result for  $\nabla_t$ .

To obtain the expression for  $\mathbb{E}[\nabla_t \nabla_t']$ , we need to show that

$$\mathbb{E}[(w_t \bar{y}_t \otimes - \text{vec}(\mathbf{I}))(w_t \bar{y}_t \otimes - \text{vec}(\mathbf{I}))'] = gG - \text{vec}(\mathbf{I})\text{vec}(\mathbf{I})'.$$

Since we take expectations with respect to the density  $p(y_t|\Sigma_t; \nu)$  in (1), we have

$$\begin{aligned} \mathbb{E} (w_t^2 \bar{y}_t \bar{y}_t' \otimes \bar{y}_t \bar{y}_t') &= \frac{(\nu+k)^2}{(\nu-2)^2} \int \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2}) [(\nu-2)\pi]^{k/2}} \frac{\bar{y}_t \bar{y}_t' \otimes \bar{y}_t \bar{y}_t'}{(1 + \bar{y}_t' \bar{y}_t / (\nu-2))^{(\nu+4+k)/2}} d\bar{y}_t \\ &= \frac{(\nu+k)^2}{(\nu-2)^2} \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})} \frac{\Gamma(\frac{\nu+4}{2})}{\Gamma(\frac{\nu+4+k}{2})} \mathbb{E} [\bar{x}_t \bar{x}_t' \otimes \bar{x}_t \bar{x}_t'], \end{aligned} \quad (38)$$

where  $\bar{x}_t$  has a Student's  $t$  distribution with zero mean, covariance matrix  $(\nu-2)\mathbf{I}/(\nu+2)$ , and  $\nu+4$  degrees of freedom. Also, since  $\bar{x}_t = x_t / \sqrt{u/(\nu+4)}$  with  $x_t$  a zero mean normal with covariance matrix  $(\nu-2)\mathbf{I}/(\nu+4)$ , and  $u$  an independent  $\chi_{\nu+4}^2$  random variable, we use Result 1 and 2 as stated at the beginning of this appendix to express (38) as

$$\frac{(\nu+k)^2}{(\nu-2)^2} \frac{(\nu+2)\nu}{(\nu+2+k)(\nu+k)} \cdot \frac{(\nu-2)^2}{(\nu+4)^2} \cdot \frac{(\nu+4)^2}{(\nu+2)\nu} \cdot G = \frac{\nu+k}{\nu+2+k} G. \quad (39)$$

This completes the proof. Theorem 1 also applies to the multivariate normal density by setting  $\nu^{-1} = 0$ . ■