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

Drew CREAL
Booth School of Business, University of Chicago, Chicago, IL 60637 (*dcreal@chicagobooth.edu*)

Siem Jan KOOPMAN
Department of Econometrics, VU University Amsterdam and Tinbergen Institute, Amsterdam, 1081 HV, The Netherlands (*s.j.koopman@vu.nl*)

André LUCAS
Department of Finance, VU University Amsterdam, Duisenberg School of Finance, and Tinbergen Institute, Amsterdam, 1081 HV, The Netherlands (*alucas@vu.nl*)

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 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. We provide an empirical illustration for a panel of daily equity returns.

KEY WORDS: Copula; Dynamic dependence; Multivariate Student t distribution.

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 t model with time-varying volatilities and correlations. The multivariate Gaussian model is treated as a special case. Our model accommodates alternative decompositions of the covariance matrix; for example, we can consider the square root of the correlation matrix in terms of hyperspherical coordinates. At the same time, the general model formulation also enables us to impose a factor structure on either the time-varying volatilities, the time-varying correlations, or both.

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 heteroscedasticity (GARCH) family of models introduced by Engle (1982) and Bollerslev (1986). The second class comprises parameter-driven models, including the multivariate stochastic volatility models of Chib, Nardari, and Shephard (2006) and Gourieroux, Jasiak, and Sufana (2009). In this article we focus on observation-driven models for time-varying correlations. Observation-driven GARCH models for time-varying correlations were originally developed by Ding and Engle (2001), Engle (2002), Engle and Sheppard (2001), and Tse and Tsui (2002). Bauwens, Laurent, and Rombouts (2006) presented a survey on multivariate GARCH models for time-varying correlation as well as time-varying covariances.

When modeling the time-varying covariance matrix of a multivariate time series, the number of static parameters is known to grow quickly as more series are added. Numerical problems may become too challenging when the parameter dimension grows large. Various trade-offs have been recognized between building models with a better statistical fit of the data versus models that have clear advantages in terms of practical implementability. Many alternative models and estimation procedures to address these challenges have been proposed; see the references mentioned earlier.

One route is to impose restrictions on the parameter space and to limit the number of parameters that control the unobserved factors driving the covariance matrix. A second strategy is to use time-varying multivariate copulas in which the variances are modeled separately from the correlations. The numerical optimization problem is then separated into more manageable pieces; see the discussions in the articles by Patton (2006) and Lee and Long (2009), who explored time-varying multivariate Gaussian copulas. Both of these strategies were adopted by Engle (2002) in developing the successful dynamic conditional correlation (DCC) model. The modeling approach of Engle is motivated by pragmatic considerations, because the DCC intends to scale well when the cross-sectional dimension of the time series. In related models, a factor structure is imposed on the volatilities and correlations (see, e.g., 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

Journal of Business & Economic Statistics
October 2011, Vol. 29, No. 4
DOI: 10.1198/jbes.2011.10070

allow us to explore 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 common dynamics on groups of assets. Ultimately, the appropriateness of a model and its associated estimation procedure depends on the application.

This article presents details on how time-varying volatilities and correlations can be incorporated into the multivariate Student 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 equity returns for a group of financials over the period January 2000 to September 2010. We show that our Student t GAS model accounts for outliers in a very natural way when updating the correlations and volatilities over time.

The remainder of the article is organized as follows. Sections 2 and 3 present the basic model specification and updating equation. Section 4 proposes alternative specifications and factor model structures. Section 5 discusses maximum likelihood estimation and reports a Monte Carlo study comparing our model's performance with that of other dynamic correlation models. Section 6 presents our illustration, and Section 7 concludes.

2. BASIC MODEL AND GENERAL RESULT

2.1 Multivariate Student t Density

Let observation vector $y_t \in \mathbb{R}^k$ follow a standardized Student t distribution with ν degrees of freedom. To simplify the notation, we set the location parameter, μ_t, of y_t to 0. If $\mu_t \neq 0$, then y_t is replaced by $y_t - \mu_t$. Thus the model can be easily extended to allow for regressors and dynamics in the mean. The observation density of y_t is given by

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

where Σ_t is the covariance matrix of y_t. We assume that $\nu > 2$, such that the covariance matrix exists; however, it is straightforward to generalize the model to the case where $0 < \nu \leq 2$ by taking the scaling matrix of the Student t distribution rather than the covariance matrix as the key parameter.

2.2 Generalized Autoregressive Score Model

The GAS model is an observation-driven model that allows parameters like the covariance matrix Σ_t to change over time using information from the score of the observation density. Although we focus on time-varying variances and correlations in this article, Creal, Koopman, and Lucas (2010) demonstrated that the GAS framework 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). For the multivariate Student t distribution, we collect the time-varying parameters of a model density in the vector f_t and specify the autoregressive updating function by

$$f_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1}, \quad (2)$$

where ω is a vector of constants, coefficient matrices A_i and B_j have appropriate dimensions for $i = 1,\ldots,p$ and $j = 1,\ldots,q$, and s_t is an appropriately scaled function of current and past data. The unknown coefficients in (2) are functions of the parameter vector θ, that is, $\omega = \omega(\theta)$, $A_i = A_i(\theta)$, and $B_j = B_j(\theta)$ for $i = 1,\ldots,p$ and $j = 1,\ldots,q$. The coefficient matrices $B_1,\ldots,B_q$ determine the persistence of the vector f_t over time. In the case of the GAS(1, 1) model, the matrix B_1 contains the autoregressive parameters.

The GAS framework was 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,

$$s_t = S_t\nabla_t, \qquad \nabla_t = \frac{\partial \log p(y_t|f_t,\mathcal{F}_{t-1};\theta)}{\partial f_t}, \quad (3)$$

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 $j = 1,\ldots,t$), and S_t is a scaling matrix of appropriate dimension. For a given density function, the equations (2) and (3) constitute our GAS model of orders p and q, which we designate 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 at time t. For example, we can let S_t equal the inverse of the Fisher information matrix, that is,

$$S_t = \mathcal{I}_{t|t-1}^{-1}, \qquad \mathcal{I}_{t|t-1} = \mathrm{E}_{t-1}[\nabla_t\nabla_t'], \quad (4)$$

such that $\mathrm{Var}(s_t) = \mathcal{I}_{t|t-1}^{-1}$. The resulting recursion in (2) can then be interpreted as a Gauss–Newton algorithm for estimating f_t through time.

Alternative choices for S_t are possible as well. For example, Nelson and Foster (1994) derived optimal filtering properties for the GARCH updating equation when the scaling is based on $\mathcal{I}_{t|t-1}^{-1/2}$ rather than on $\mathcal{I}_{t|t-1}^{-1}$. Our choice to set $S_t = \mathcal{I}_{t|t-1}^{-1}$ brings our GAS model closest to the familiar GARCH specification. 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 (4) 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 (2) with $p = q = 1$ becomes

$$f_{t+1} = \omega + A_1(y_t^2 - f_t) + B_1 f_t, \quad (5)$$

which is equivalent to the GARCH model of Bollerslev (1986). A similar argument holds in the multivariate context.

When we consider the univariate Student t density for y_t and let $f_t = \sigma_t^2$ with scaling (4), we obtain the updating equation

$$f_{t+1} = \omega + A_1 \cdot (1 + 3\nu^{-1}) \cdot \left(\frac{(1+\nu^{-1})}{(1-2\nu^{-1})(1+\nu^{-1}y_t^2 \,/\, ((1-2\nu^{-1})\cdot f_t))}y_t^2 - f_t\right) + B_1 f_t, \quad (6)$$

which is different from the Student t-GARCH(1, 1) model of Bollerslev (1987). The standard t-GARCH(1, 1) model also uses a Student t density but includes updating equation (5). The denominator of the second term on the right side of (6) causes a more moderate increase in the variance for a large realization of $|y_t|$ as long as ν is finite; thus 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, then a large value of y_t^2 is not necessarily due to an increase in variance; it might be due to the heavy-tailed nature of the distribution. The functional form in (6) 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, among others, Nelson and Foster (1994) and Harvey and Chakravarty (2008). The models presented in this article, however, are inherently multivariate and applicable to a range of different specifications. Other extensions, such as considering asymmetric multivariate Gaussian or Student t distributions for the modeling of leverage in financial return series, are also possible.

A discussion on the interpretation of the dynamic parameters A_1 and B_1 is in order. The GARCH(1, 1) model with $f_t = \sigma_t^2$ is usually presented by $f_{t+1} = \omega + \alpha_1 y_t^2 + \beta_1 f_t$, so that $\alpha_1 = A_1$ and $\beta_1 = B_1 - A_1$ relative to (5). The GARCH parameters α_1 and β_1 have natural constraints $\alpha_1 \geq 0$ and $\beta_1 \geq 0$ for the variance to remain positive at all times, and $\alpha_1 + \beta_1 < 1$ for the process to be covariance stationary. The same intuition holds for the GAS(1, 1) case, but with $1 > B_1 \geq A_1 \geq 0$. These constraints can be imposed directly when estimating the model. Moreover, as shown later, the constraint that the roots of $I - B_1 z - \cdots - B_q z^q$ lie outside the unit circle is a generic necessary condition for covariance stationarity of the GAS model for particular choices of the scaling matrix S_t. This holds even in cases of complex models and parameterizations such as the hypersphere parameterization used later in the paper. In such cases, it is generally impossible to formulate restrictions on the parameters A_i and B_i that ensure that the time-varying parameters remain in their appropriate domain (e.g., positive for variances, between -1 and $+1$ for correlations, and positive definite for a covariance matrix). We circumvent this problem by adopting parameterizations in which f_t is allowed to float freely. Because the driving mechanism s_t in the GAS model is based on the score function, it automatically adapts to such a parameterization by modifying the appropriate functions of the data (s_t) that drive the changes in f_t. Consequently, no restrictions are needed for the A_i coefficients, whereas the B_i coefficients need only satisfy the stationarity requirements.

2.3 Matrix Notation and Definitions

To develop our results that follow, we adopt the following matrix notation and definitions. The Kronecker product is denoted by $A \otimes B$ for two matrices A and B. In the case where $B = A$, we define $A_{\otimes} = A \otimes A$. The operator vec($A$) vectorizes matrix A into a column vector, whereas vech(A) vectorizes the lower-triangular part of matrix A into a column vector. We define the operator $\oplus$ for two matrices A and B as $A \oplus B = (A \otimes B) + (B \otimes A)$. We write the duplication [$\mathcal{D}_k \operatorname{vech}(A) = \operatorname{vec}(A)$], elimination [$\mathcal{B}_k \operatorname{vech}(A) = \operatorname{vec}(A)$], and commutation [$\mathcal{C}_k \operatorname{vec}(A) = \operatorname{vec}(A')$] matrices, respectively (see Abadir and Magnus 2005, ch. 11).

To introduce the GAS updating function for the multivariate Student t distribution, we specify the covariance matrix Σ_t as a function $\Sigma(f_t)$ of the time-varying factor f_t in (2). We define the factors f_t and the functional link between f_t and Σ_t in detail in Sections 3 and 4. A simple example of such a link function is $f_t' = [\operatorname{diag}(D_t)', \operatorname{vech}(Q_t)']$ with $\Sigma_t = \Sigma(f_t) = D_t R_t D_t$, where D_t is a diagonal matrix containing the standard deviations and R_t is a correlation matrix constructed as $R_t = \operatorname{diag}(Q_t)^{-1/2} Q_t \operatorname{diag}(Q_t)^{-1/2}$. Other, more complex links between f_t and $\Sigma_t = \Sigma(f_t)$ are also possible and are discussed later.

2.4 General Result

The theorem presented in this section forms the basis for the model specifications in Sections 3 and 4. Define

$$\Psi_t = \Psi(f_t) = \frac{\partial \operatorname{vech}(\Sigma_t)}{\partial f_t'} \quad \text{for } \Sigma_t = \Sigma(f_t). \tag{7}$$

The matrix Ψ_t depends explicitly on the chosen relationship between f_t and Σ_t. The elements of f_t may represent elements of Σ_t, but also may represent log-variances, partial correlations, or other variables that are used for the construction of Σ_t.

Theorem 1. For the Student t density (1) and time-varying factor f_t in (2), we have

$$\begin{aligned} \nabla_t &= \frac{\partial \log p_t(y_t|\Sigma_t; \nu)}{\partial f_t} \\ &= \frac{1}{2} \Psi_t' \mathcal{D}_k' \Sigma_{t\otimes}^{-1} [w_t y_{t\otimes} - \operatorname{vec}(\Sigma_t)], \qquad (8) \\ \mathcal{I}_{t|t-1} &= \mathrm{E}[\nabla_t \nabla_t'] \\ &= \frac{1}{4} \Psi_t' \mathcal{D}_k' J_{t\otimes}' [gG - \operatorname{vec}(\mathrm{I}) \operatorname{vec}(\mathrm{I})'] J_{t\otimes} \mathcal{D}_k \Psi_t, \qquad (9) \end{aligned}$$

where $\mathcal{D}_k$ is the duplication matrix and with 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 (A.1) of the Appendix. The square root matrix J_t can be obtained from any convenient matrix decomposition procedure.

For the proof see the Appendix.

Theorem 1 reveals a number of important features. First, irrespective of the model specification and the definition of the factors f_t, 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 Σ_t. For the normal distribution, the weighting term w_t collapses to 1, and we obtain the familiar driving mechanism of a multivariate GARCH model.

Second, Theorem 1 shows that potentially different decompositions and parameterizations of Σ_t are accounted for by the matrix function Ψ_t, which gathers the derivatives of the full covariance matrix Σ_t with respect to the 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 Ψ_t needs to change in

that case. Thus the GAS framework 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 other examples of decomposition of the covariance matrix are discussed in Sections 3 and 4.

Third, the weight w_t in (8) also appears in the univariate specification of (6) with $k = 1$. If the density of the observations y_t is heavy-tailed ($\nu^{-1} > 0$), then large values in $y_t y_t'$ (in absolute terms) do not automatically lead to dramatic changes in the elements of Σ_t. Such large values may be due to the heavy-tailed feature of the distribution of y_t rather than to a dramatic change of elements of the covariance matrix. The weight w_t in (8) automatically accounts for extreme values because it decreases if $y_t' \Sigma_t^{-1} y_t$ is large.

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

$$f_t = \omega + \sum_{i=1}^{\infty} B_1^{i-1} A_1 s_{t-i}. \tag{10}$$

If the conditional variance of s_t is constant over time, then the process f_t is covariance stationary when the roots of B_1 lie 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 will be much harder. Creal, Koopman, and Lucas (2010) addressed this issue by scaling with $S_t = \mathcal{I}_{t|t-1}^{-1/2}$ instead. In the developments that follow, we want to stay close to the multivariate GARCH framework, which requires that the score be scaled by $\mathcal{I}_{t|t-1}^{-1}$, as shown for the univariate case in Section 2.2.

Our recursion for the factors in the GAS model shares some features with the updating recursion in the DCC model based on (14). For example, the number of parameters in the model can be limited substantially by setting the coefficient matrices in (2) to be scaled identity matrices or diagonal matrices. The GAS model does not reduce to the DCC model, however, due mainly to the presence of the weighting term w_t in (8). 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. Thus, our current specification differs substantially from the DCC specification with Student t–distributed error terms. Moreover, when we consider the normal distribution ($\nu^{-1} = 0$ and $w_t \equiv 1$), our modeling framework does not reduce to the DCC specification.

2.5 An Illustration for Time-Varying Correlation

In the time-varying variance model implied by Theorem 1, unexpected deviations $w_t y_t y_t' - \Sigma_t$ drive the evolution of volatilities and correlations. Unexpected large or small cross-products are taken into account, as are unexpected large or small squared observations. To understand how correlations are updated through time, 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 $\mathrm{vec}(\Sigma_t) = (1, \rho_t, \rho_t, 1)'$. The key component in updating the time-varying correlation is the score ∇_t in Theorem 1, for which it follows that $\Psi_t' \mathcal{D}_k' = (0, 1, 1, 0)$. This does not imply that only the second and third elements of $w_t y_{t\otimes} - \mathrm{vec}(\Sigma_t)$ (the cross-product terms), are taken into account. Because $\Psi_t' \mathcal{D}_k'$ is first postmultiplied by $\Sigma_t^{-1} \otimes \Sigma_t^{-1}$, we obtain

$$s_t = \frac{2}{(1-\rho_t^2)^2}[(1+\rho_t^2)(y_{1t}y_{2t} - \rho_t) - \rho_t(y_{1t}^2 + y_{2t}^2 - 2)], \tag{11}$$

where y_{it} is the ith element of y_t for $i = 1, 2$. The first term in this equation, $y_{1t}y_{2t} - \rho_t$, enforces an increase in the correlation when $y_{1t}y_{2t}$ exceeds ρ_t. The second term in the score $(y_{1t}^2 + y_{2t}^2 - 2)$ is less intuitive. To understand the impact of this term, assume that $\rho_t = 0.5$ and consider two potential vectors of observations, $(y_{1t} = 1, y_{2t} = 1)$, or, alternatively, $(y_{1t} = 0.25, y_{2t} = 4)$. Both vectors of observations are chosen such that the value of $y_{1t}y_{2t}$ is the same, whereas $y_{1t}^2 + y_{2t}^2$ differs. For the first vector $(y_{1t} = 1, y_{2t} = 1)$, the second term in (11) equals 0, because $y_{1t}^2 + y_{2t}^2$ is equal to its expected value of 2. For the second vector $(y_{1t} = 0.25, y_{2t} = 4)$, the second term is negative and completely offsets the effect of the first term. Consequently, even though the cross-products are the same, the second vector carries a much stronger signal that the current value of the correlation $\rho_t = 0.5$ is too high and should be decreased. In contrast, if the observation vector is $(y_{1t} = 4, y_{2t} = 4)$, then the first term offsets the second term, and the correlation increases even though $y_{1t}^2 + y_{2t}^2$ is large. All of these effects are in line with what would be expected intuitively.

3. UPDATE EQUATIONS FOR VOLATILITIES AND CORRELATIONS

We develop the GAS updating equations for the multivariate Student t density with different decompositions of the time-varying covariance matrix Σ_t. In all cases, Theorem 1 applies because a different specification only affects how $\Sigma_t = \Sigma(f_t)$ depends on f_t. In this section we decompose the covariance matrix by

$$\Sigma_t = D_t R_t D_t, \tag{12}$$

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 (12) closely follows the DCC model of Engle (2002), although other decompositions might be followed as well; see Section 4.

3.1 Time-Varying Volatilities and Correlations

Here we consider the multivariate Student t density with time-varying D_t, as well as R_t. We decompose the correlation matrix R_t as

$$R_t = \Delta_t^{-1} Q_t \Delta_t^{-1}, \tag{13}$$

where Q_t is a symmetric positive definite matrix and Δ_t is a diagonal matrix whose nonzero elements equal the square root of the diagonal elements of Q_t. The transformation (13) ensures that the correlation matrix R_t is positive definite and symmetric

with off-diagonal elements between $(-1, 1)$. The specification (13) 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

$$Q_{t+1} = \Omega_{\text{dcc}}(\iota - A_{\text{dcc}} - B_{\text{dcc}}) + A_{\text{dcc}} \odot \hat{\varepsilon}_t \hat{\varepsilon}_t' + B_{\text{dcc}} \odot Q_t, \quad (14)$$

where $\hat{\varepsilon}_t$ is a $k \times 1$ vector with elements $\hat{\varepsilon}_{it} = y_{it}/\sigma_{it}$, $\odot$ is the Hadamard product (element by element multiplication) and $k \times k$ matrices Q_0, Ω_{dcc}, A_{dcc}, and B_{dcc} are fixed and unknown. The covariance matrix Q_t is positive definite for all t when Ω_{dcc} and Q_0 are positive definite and the conditions $a_{ij} \geq 0$, $b_{ij} \geq 0$, and $a_{ij} + b_{ij} < 1$ apply for all $i, j = 1, \ldots, k$, with a_{ij} and b_{ij} as the (i, j) elements of A_{dcc} and B_{dcc}, respectively. [For a more general discussion on positive definite conditions for the DCC model, see Engle and Sheppard (2001, proposition 2).] It is common practice to reduce the parameter space by replacing Ω_{dcc} with the sample correlation matrix of the standardized residuals $y_{i,t}/\sigma_{i,t}$ and by restricting $A_{\text{dcc}} = a_{\text{dcc}} \cdot \iota\iota'$ and $B_{\text{dcc}} = b_{\text{dcc}} \cdot \iota\iota'$, where a_{dcc} and b_{dcc} are scalars and ι is a vector of 1s. Aielli (2008) showed that this approach of preestimating Ω_{dcc} leads to inconsistent estimates and provides an alternative specification known as the cDCC model that resolves this issue. In the cDCC model, the elements of the vector $\hat{\varepsilon}_t$ are defined as $\hat{\varepsilon}_{it} = (y_{it}\sqrt{q_{it}})/\sigma_{it}$ instead of $\hat{\varepsilon}_{it} = y_{it}/\sigma_{it}$, where q_{it} is the ith diagonal element of Q_t. This specification leads to consistent and asymptotically normal parameter estimates. We include the cDCC in our comparisons, but refrain from preestimation. Instead, we estimate all of the model's static parameters jointly by maximizing the full likelihood.

For the decomposition of Σ_t given by (12) and (13), we specify our factor as

$$f_t = \begin{pmatrix} \text{diag}(D_t^2) \\ \text{vech}(Q_t) \end{pmatrix}. \quad (15)$$

Our proposed updating equation (2) is given in Theorem 1, with

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

Define the matrices $W_{\Delta t}$, W_{Dt}, $\mathcal{S}_\Delta$, $\mathcal{S}_D$, 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 0s. Similarly, the matrix W_{Dt} 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 0s. The matrices $\mathcal{S}_\Delta$, $\mathcal{S}_D$, and $\mathcal{S}_Q$ are selection matrices containing only 1s and 0s 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 $\text{vech}(Q_t) = \mathcal{S}_Q f_t$. We then obtain

$$\begin{aligned}
\Psi_t &= \mathcal{B}_k (\iota \oplus D_t R_t) \cdot W_{Dt} \mathcal{S}_D \\
&\quad + \mathcal{B}_k D_{t\otimes} \Delta_{t\otimes}^{-1} [\mathcal{D}_k - (\Delta_t \oplus Q_t) \Delta_{t\otimes}^{-1} W_{\Delta t} \mathcal{S}_\Delta] \mathcal{S}_Q. \quad (16)
\end{aligned}$$

The updating of time-varying Σ_t is implied by (2) and Theorem 1, with Ψ_t given by (16). If only the correlations in R_t are time-varying, then $\mathcal{S}_D = 0$, and the first term in (16) drops out. Similarly, if only the volatilities are time-varying and the correlations are constant, then (16) gives the appropriate specification by setting $\mathcal{S}_Q = 0$ and the second term in (16) vanishes.

The current specification where the vector of factors f_t contains $\text{vech}(Q_t)$ has obvious drawbacks that it shares with the DCC model. The number of time-varying factors that drive the correlations is greater than the number of free correlations in R_t. For example, vector $\text{vech}(Q_t)$ contains $k(k+1)/2$ parameters, whereas R_t has $k(k-1)/2$ free parameters. The information matrix is singular as a result, and we use its Moore–Penrose pseudoinverse to scale the score steps in (4). Moreover, when using the decomposition of the correlation matrix (13), we must restrict k elements of ω in the GAS model or k elements in Ω_{dcc} of the DCC model to identify either model. In particular, the elements of ω (or Ω_{dcc}) corresponding to the diagonal elements of Q_t can be multiplied by any arbitrary positive number without changing the decomposition. In our empirical work, we resolve this by setting the elements in ω that correspond to the diagonal elements of Q_t equal to 1.

3.2 An Illustration for Time-Varying Correlation (Continued)

To illustrate our driving mechanism and compare it with the DCC model, we return to the bivariate illustration, $k = 2$, of Section 2.5 with constant unit variances $D_t = \iota$ and a time-varying correlation that is updated by (11). We then have

$$\begin{aligned}
\Psi_t' &= \frac{1}{2\sqrt{q_{11,t} q_{22,t}}} \\
&\quad \times \begin{pmatrix} -q_{12,t}/q_{11,t} & 2 & -q_{12,t}/q_{22,t} \end{pmatrix}' \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \quad (17)
\end{aligned}$$

where $q_{ij,t}$ is the (i, j) element of Q_t for $i, j = 1, 2$. The updating mechanism for each element of $f_t = \text{vech}(Q_t)$ is driven by the single driver (11) that is multiplied by the three elements in the first vector in (17). In contrast, the driving mechanism in the DCC model for the elements $q_{11,t}$, $q_{12,t}$, and $q_{22,t}$ are given by y_{1t}^2, $y_{1t}y_{2t}$, and y_{2t}^2, respectively. We thus have three different driving mechanisms for the DCC model, compared with only one driving mechanism for the GAS model. The numerical intuition is provided in Section 2.5. The single driver of the GAS model contains all of the relevant information from the data (up to first order) on changes in the correlation parameter.

3.3 A Time-Varying Copula Specification

It is straightforward to transform the general framework of Section 3.1 to a copula specification similar to those reviewed by Patton (2009). 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 ith series in y_t. The Student t density then operates on the $k \times 1$ vector $\tilde{y}_t' = [P_\nu^{-1}(u_{1t}), \ldots, P_\nu^{-1}(u_{kt})]'$ via Sklar's theorem (Sklar 1959), where $P_\nu^{-1}(\cdot)$ is the univariate inverse Student t distribution with ν degrees of freedom. In the copula specification, the decomposition (12) simplifies because $D_t \equiv \iota$. We obtain

$$\Psi_t = \mathcal{B}_k D_{t\otimes} \Delta_{t\otimes}^{-1} [\mathcal{D}_k - (\Delta_t \oplus Q_t) \Delta_{t\otimes}^{-1} W_{\Delta t} \mathcal{S}_\Delta] \mathcal{S}_Q, \quad (18)$$

where $\mathcal{S}_Q = \mathrm{I}$ because $f_t = \mathrm{vech}(Q_t)$. Note that the marginal densities drop out of the expression for the GAS step s_t in Theorem 1, because they do not depend on R_t. Thus they vanish when taking derivatives of the log-density with respect to f_t. But to evaluate the likelihood and estimate the parameters, the marginal densities must be considered as well, mainly because they depend on the unknown parameter ν, which must be estimated.

4. ALTERNATIVE DECOMPOSITIONS OF THE COVARIANCE MATRIX

In this section we consider alternative decompositions of the time-varying covariance matrix Σ_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 Ψ_t.

4.1 Time-Varying Log-Volatilities

To impose positive standard deviations in the decomposition (12), we can, for example, define the vector of factors f_t as

$$f_t = \begin{pmatrix} \log(\mathrm{diag}(D_t^2)) \\ \mathrm{vech}(Q_t) \end{pmatrix}. \tag{19}$$

In the log-variance specification with correlation matrix given by (13), we obtain

$$\begin{aligned}\Psi_t = {} & \mathcal{B}_k(\mathrm{I} \oplus D_t R_t) W_{Dt} D_t^2 \mathcal{S}_D \\ & + \mathcal{B}_k D_{t\otimes} \Delta_{t\otimes}^{-1} [\mathcal{D}_k - (\Delta_t \oplus Q_t)\Delta_{t\otimes}^{-1} W_{\Delta t} \mathcal{S}_\Delta] \mathcal{S}_Q,\end{aligned} \tag{20}$$

which is the same as Ψ_t in (16), except the first term is multiplied by D_t^2. The multiplication by D_t^2 accounts for the log transformation in the derivative of $\mathrm{diag}(D_t^2)$ 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 decomposition of the correlation matrix R_t. The covariance matrix decomposition $\Sigma_t = D_t R_t D_t$ remains as in (12). The difficulty with specifying a correlation matrix is that necessary conditions are needed: (a) The matrix R_t has to be positive (semi) definite, and (b) the diagonal elements of R_t are equal to one for all values of t. To satisfy (a), we can adopt the Cholesky decomposition of R_t rather than R_t itself. However, the Cholesky decomposition by itself does not automatically satisfy (b). An alternative decomposition that satisfies both conditions simultaneously is based on hyperspherical coordinates and is given by $R_t = X_t' X_t$, where $X_t = X_t(\phi_t)$ is an upper-triangular matrix, that is,

$$\begin{aligned} R_t &= X_t(\phi_t)' X_t(\phi_t), \\ X_t(\phi_t) &= \begin{pmatrix} 1 & c_{12t} & c_{13t} & \cdots & c_{1kt} \\ 0 & s_{12t} & c_{23t}s_{13t} & \cdots & c_{2kt}s_{1kt} \\ 0 & 0 & s_{23t}s_{13t} & \cdots & c_{3kt}s_{2kt}s_{1kt} \\ 0 & 0 & 0 & \cdots & c_{4kt}s_{3kt}s_{2kt}s_{1kt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & c_{k-1,kt}\prod_{\ell=1}^{k-2} s_{\ell kt} \\ 0 & 0 & 0 & \cdots & \prod_{\ell=1}^{k-1} s_{\ell kt} \end{pmatrix}, \end{aligned} \tag{21}$$

with scalar $c_{ijt} = \cos(\phi_{ijt})$, scalar $s_{ijt} = \sin(\phi_{ijt})$, and scalar ϕ_{ijt} as the time-varying angle measured in radians. The vector ϕ_t contains the $k(k-1)/2$ angles, ϕ_{ijt}, for $i, j = 1, \ldots, k$. The columns of X_t consist of hyperspherical coordinates with unit length. This decomposition was used by Jaeckel and Rebonato (2000) with constant angles over time as a way to model term structure data. A similar decomposition for parameterizing a set of orthonomal matrices has been adopted in the GO-GARCH model of van der Weide (2002). The GAS framework is ideally suited to allow these angles, and thus the correlations, to be time-varying. In this case, we define the vector of factors as $f_t' = [\mathrm{diag}(D_t^2)', \phi_t']$. The general structure of Theorem 1 automatically provides the functions of the data to drive the angles ϕ_t through the GAS updating variable s_t.

As before, the only change required to the previous specifications is the form of Ψ_t. We use the derivatives

$$\frac{\partial x_{ijt}}{\partial \phi_{\ell mt}} = \begin{cases} 0 & \text{if } i > j,\ \ell \geq m,\ \ell \geq i, \text{ or } j \neq m \\ -x_{ijt} \cdot \tan(\phi_{ijt}) & \text{if } i = \ell \text{ and } i \neq j \\ x_{ijt} / \tan(\phi_{\ell jt}) & \text{otherwise} \end{cases} \tag{22}$$

for $i, j, \ell, m = 1, \ldots, k$, 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

$$Z_t = \frac{\partial \,\mathrm{vec}(X_t)}{\partial \phi_t'}.$$

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

$$\begin{aligned} \frac{\partial \,\mathrm{vec}(R_t)}{\partial f_t'} &= \frac{\partial \,\mathrm{vec}(X_t' X_t)}{\partial \phi_t'} \mathcal{S}_\phi \\ &= [(\mathrm{I} \otimes X_t') + (X_t' \otimes \mathrm{I})\mathcal{C}_k] Z_t \mathcal{S}_\phi, \end{aligned} \tag{23}$$

where $\mathcal{C}_k$ is the commutation matrix. Combining these results, we obtain

$$\begin{aligned}\Psi_t = {} & \mathcal{B}_k(\mathrm{I} \oplus D_t R_t) W_{Dt} \mathcal{S}_D \\ & + \mathcal{B}_k D_{t\otimes} [(\mathrm{I} \otimes X_t') + (X_t' \otimes \mathrm{I})\mathcal{C}_k] Z_t \mathcal{S}_\phi.\end{aligned} \tag{24}$$

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

The hyperspherical specification has an advantage over the decomposition of R_t in (13) in 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 ω are identified, and the scaling matrix S_t in (4) is nonsingular. When the dimension k of y_t increases, the number of elements in f_t for the hyperspherical decomposition is also smaller than that for the decomposition (13). A possible disadvantage of this decomposition relative to (13) is the interpretation of the factors as angles. Each angle is identified only within the region $[-\pi, \pi]$. This constraint can be imposed on the factors via a transformation. In practical cases, however, our experience is that numerical problems do not occur when such constraints are omitted.

4.3 Common Dynamic Factors

To this point, we have emphasized the setting where the dimension of f_t is at least as large as the number of free elements in D_t and R_t. However, time-varying features may be shared between different volatilities and different correlations (even between volatilities and correlations). To reduce the dimension of f_t, we can impose a common factor structure on the volatilities and correlations. Consider partitioning f_t as in (15), that is, $f_t' = [\text{diag}(D_t^2)', \text{vech}(Q_t)']$. With a slight abuse of notation, we have

$$\text{diag}(D_t^2) = a + \mathcal{S}_D f_t, \qquad \text{vech}(Q_t) = b + \mathcal{S}_Q f_t, \tag{25}$$

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 nonlinear factor structures are straightforward. The form of Ψ_t is now the same as in (16). For the hyperspherical coordinate specification, the result remains (24), with $\mathcal{S}_\phi$ specified as a real-valued matrix rather than a selection matrix consisting of 0s and 1s. For multivariate GARCH models, factor structures have been imposed by Tsay (2005) and Bauwens, Laurent, and Rombouts (2006).

A low-dimensional f_t reduces the number of time-varying factors but typically increases the number of unknown static parameters in the factor loading matrices $\mathcal{S}_D$ and $\mathcal{S}_Q$ (or $\mathcal{S}_\phi$). 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 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 of the correlations and having unit values in the matrix (here, vector) $\mathcal{S}_Q$. This structure imposed on R_t may have both statistical and computational advantages when modeling very high-dimensional systems.

5. ESTIMATION AND MONTE CARLO EVIDENCE

In this section we report a Monte Carlo study carried out 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 present the Monte Carlo design, and finally discuss our results.

5.1 Parameter Estimation

Given n realizations for the observation vector y_t with mean 0, the log-likelihood function for the multivariate Student t model is $\mathcal{L} = \sum_{t=1}^{n} \log p(y_t|\Sigma_t; \nu)$, where $p(y_t|\Sigma_t; \nu)$ is given in (1) and the time variation of Σ_t is determined by the GAS updating equations for D_t and R_t. For given values of ν, ω, A_i, B_i, and initial conditions f_j $(j < 1)$ in (2), the log-likelihood function can be evaluated in a straightforward way. The unknown coefficients are collected in the parameter vector θ, and its estimation is based on the maximization of the log-likelihood with respect to θ. The initial values f_j $(j < 1)$ are set to their unconditional expectation, $(\mathrm{I}_k - B_1 - \cdots - B_q)\omega$. Maximization can be done via a standard quasi-Newton numerical optimization procedure.

5.2 Monte Carlo Study

The design of our Monte Carlo study is similar to that of the study conducted by Engle (2002). We simulate a series of $n = 1000$ observations from a bivariate Student t distribution with unit variance and time-varying correlation ρ_t. We consider the following time-varying patterns for ρ_t:

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 $h_t = -0.4(1 - 0.99) + 0.99h_{t-1} + 0.14\eta_t$, $\eta_t \sim \mathcal{N}(0, 1)$.

Graphs depicting each of these patterns were provided by Engle (2002). Given one of these patterns, the data-generation process is given by

$$y_t \sim p(y_t|\Sigma_t; \nu), \qquad \Sigma_t = R_t = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}, \nu = 5, \tag{26}$$

where $p(y_t|\Sigma_t; \nu)$ is the bivariate Student t density with $k = 2$. In the Monte Carlo study, we simulate 1000 bivariate series for each of the six time-varying patterns ρ_t.

For a realized bivariate time series, we consider the t-GAS(1, 1) model with recursion (2) for f_t. We consider both the specification of R_t given by (13) and that given by the hyperspherical coordinates in (21). For both models, we estimate the unknown parameters ν, ω, A_1, and B_1 by maximum likelihood. Our two measures of accuracy are the mean absolute error (MAE) and the mean squared error (MSE) as given by

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

Once the two GAS specifications are treated for a simulated series y_t, we consider alternative approaches of estimating ρ_t for comparison purposes. In particular, we benchmark the two GAS models against the cDCC model of Aielli (2008), which was recently used by Engle and Kelly (2009) and Brownlees and Engle (2010). We also consider as a benchmark the exponentially weighted moving average (EWMA) recursion for Q_t in (13) given by

$$Q_{t+1} = \lambda Q_t + (1 - \lambda) y_t y_t', \tag{27}$$

where the smoothing parameter is set to $\lambda = 0.96$.

The results from this Monte Carlo study are presented in Table 1, where we take the cDCC 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 (21) appears to be preferred to the other models for five out of the six different correlation paths. The improvements in MSE or MAE are typically in the range of 5%–15%, except for the ramp and fast sine data-generating processes. The results for the GAS(1, 1) model based on the decomposition of R_t given by (13) are second best. The cDCC model performs best in terms of minimizing MAE and MSE only when the true data-generating process is the fast sine.

Table 1. MAE and MSE results: in-sample

	Constant	Sine	Fast Sine	Step	Ramp	Model
MAE						
t-GAS	0.7661	0.9071	1.0252	0.9460	**0.9615**	0.9358
t-GAS-h	**0.7150**	**0.8982**	1.0354	**0.9345**	0.9677	**0.9354**
t-cDCC	1.000	1.0000	**1.0000**	1.0000	1.0000	1.0000
EWMA	3.029	1.2416	1.1126	1.0648	1.1067	1.0949
MSE						
t-GAS	0.44649	0.8844	1.0342	0.9718	1.0031	0.8887
t-GAS-h	**0.3811**	**0.8595**	1.0455	**0.9443**	**0.9981**	**0.8862**
t-cDCC	1.0000	1.0000	**1.0000**	1.0000	1.0000	1.0000
EWMA	8.8414	1.4465	1.3049	1.1458	1.2557	1.2149

NOTE: The table presents the MAE and MSE for the estimated dynamic correlation patterns for six different models and six different correlation processes. The details of the *t*-GAS method are provided in Section 3.1. The *t*-GAS-h method is based on the hyperspherical coordinates specification of Section 4.2. The results for *t*-cDCC are obtained from the specification of Aielli (2008) using a Student *t* likelihood function. The EWMA method refers to equation (27) with $\lambda = 0.96$. The MAE and MSE measures are presented relative to the *t*-cDCC model.

6. EMPIRICAL ILLUSTRATION: A PANEL OF EQUITY RETURNS

In this section, we adopt the models developed in Sections 2 and 4 for a panel of daily equity (total) returns between January 1, 2000, to September 14, 2010, obtained from Datastream. We evaluate the models in terms of in-sample performance as well as in an out-of-sample forecasting comparison that includes the financial crisis of 2008. The data set consists of 2792 time series observations for four U.S. financial equities with ticker symbols FRE (Federal Home Loan Mortgage Corporation, or Freddie Mac), STT (State Street Corporation), AET (Aetna), and CVH (Coventry Health Care). The data have been adjusted for dividends and stock splits. The histograms of the returns for these four equities appeared to produce the heaviest tails among a large group of selected equities, such that the effect of fat tails is clearly visible.

6.1 In-Sample Performance

Our empirical study considers GARCH/DCC and GAS classes of models. For GARCH/DCC models, we follow Brownlees and Engle (2010) in specifying GARCH models for the marginal variances of each series and a cDCC model for the correlation process between series. We estimate the long-run variances and correlations within each of the GARCH and cDCC models simultaneously, together with the remaining parameters. For the *i*th individual GARCH model, with $i = 1, \ldots, 4$, the parameters a_i and b_i represent $\alpha_1 = A_1$ and $\beta_1 = B_1 - A_1$ in (5), respectively. These are the usual GARCH coefficients as defined after (5). The DCC model has two common parameters, a_5 and b_5, which specify the matrix coefficients $A_{\text{dcc}} = a_5 \cdot \iota\iota'$ and $B_{\text{dcc}} = b_5 \cdot \iota\iota'$ in (14). The variance and correlation parameters of the GAS model are modeled simultaneously as follows. The first elements in vector f_t represent the variances in, for example, $\text{diag}(D_t^2)$ of the decomposition (12). The remaining elements of f_t represent the correlations in Q_t as defined in (13) or the hyperspherical angles in ϕ_t as defined after (21). The dynamic specification of f_t is given 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, whereas the remaining diagonal elements associated with the correlations have the same value. In our emprical study with four series, the diagonal elements of A_1 and B_1 are given by $a_1, \ldots, a_4, a_5, \ldots, a_5$ and $b_1, \ldots, b_4, b_5, \ldots, b_5$, respectively. The parameters are estimation as described in Section 5.1.

Table 2 reports the estimated parameters as well as the (maximized) log-likelihood values, the Akaike information criterion (AIC) and the Schwartz Bayes information criterion (BIC) for eight different models. The first two columns in Table 2 present the estimates for the GARCH/cDCC model with Gaussian and Student *t* likelihood specifications. As is typical for stock return data, the Student *t* specification produces a much higher log-likelihood value at its maximum. The remaining columns of Table 2 consider the GAS(1, 1) model with different densities and different specifications of the covariance matrix. The columns "*g*-GAS" and "*t*-GAS" are based on Gaussian and Student *t* densities, respectively, with variance specifications (12) and (13), and with vector f_t given by (15). The number of parameters in these specifications is the same as for their GARCH/cDCC counterparts. The *g*-GAS model performs similar to the *g*-GARCH/*g*-cDCC model, but the *t*-GAS model performs best compared with both the *t*-GARCH/*t*-cDCC and *g*-GAS models. We conclude that the assumption of normality is clearly rejected for this dataset. Moreover, the *t*-GAS specification outperforms the *t*-cDCC model by more than 20 points with the same number of parameters. Most of the estimated autoregressive parameter values b_i, $i = 1, \ldots, 5$, for the *t*-GAS model are close to unity such that the estimated factors are highly persistent. These results also indicate that a more parsimonious model can be obtained by restricting the coefficients $b_1, \ldots, b_4$ to be the same.

The fifth, sixth, and seventh columns in Table 2 present the estimation results for GAS models with three alternative specifications for the vector f_t. 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 (21) in Section 4.2. The *t*-GAS-l model is the *t*-GAS model with the variances replaced by the log-volatility factors of Section 4.1. The *t*-GAS-hl model is the combination of the two previous models. For our data set of daily returns, the differences in the log-likelihood values for different *t*-GAS models are small. Thus, the covariance matrix parameterization appears to be less important for the volatility and correlation dynamics in our dataset.

The estimated correlations for a selection of series are presented in the left panels of Figure 1. The right panels show the estimated volatilities for two of the series during the financial crisis. The two time series of estimated correlations in each left-side plot are based on the *t*-cDCC and *t*-GAS models. The figures illustrate the differences between GAS and cDCC models with Student *t* densities for the errors. The correlation estimates from the two models are substantially different for this dataset. The differences in the estimated correlation patterns for the cDCC and *t*-GAS models are also reflected in differences in the estimated coefficients. For example, the coefficients a_5 reported in Table 2 are significantly higher for the GAS models than for the cDCC models. This leads to a more responsive estimated correlation pattern for the GAS vis-à-vis the cDCC models, as shown in Figure 1. From the right-side panels of

Table 2. Parameter estimates, likelihoods, and information criteria

	g-cDCC	t-cDCC	g-GAS	t-GAS	t-GAS-h	t-GAS-l	t-GAS-h-l	tg-GAS
a_1	0.0534	0.1068	0.0523	0.0670	0.0670	0.0684	0.0685	0.1072
	(0.0043)	(0.0136)	(0.0043)	(0.0080)	(0.0080)	(0.0079)	(0.0079)	(0.0137)
a_2	0.0297	0.0856	0.0300	0.0566	0.0565	0.0517	0.0517	0.0857
	(0.0022)	(0.0146)	(0.0023)	(0.0076)	(0.0075)	(0.0066)	(0.0066)	(0.0141)
a_3	0.0113	0.1030	0.0111	0.0639	0.0633	0.0589	0.0585	0.0963
	(0.0007)	(0.0175)	(0.0007)	(0.0105)	(0.0105)	(0.0096)	(0.0095)	(0.0170)
a_4	0.1285	0.1307	0.1292	0.0652	0.0648	0.0602	0.0600	0.1345
	(0.0118)	(0.0222)	(0.0118)	(0.0100)	(0.0100)	(0.0088)	(0.0088)	(0.0228)
a_5	0.0036	0.0032	0.0054	0.0113	0.0111	0.0115	0.0112	0.0034
	(0.0006)	(0.0011)	(0.0012)	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0006)
b_1	0.9458	0.8893	0.9992	0.9977	0.9978	0.9939	0.9938	0.9959
	(0.0044)	(0.0138)	(0.0003)	(0.0012)	(0.0012)	(0.0018)	(0.0018)	(0.0022)
b_2	0.9693	0.9082	0.9990	0.9962	0.9962	0.9928	0.9928	0.9947
	(0.0023)	(0.0151)	(0.0004)	(0.0024)	(0.0024)	(0.0023)	(0.0023)	(0.0027)
b_3	0.9886	0.8308	0.9998	0.9803	0.9807	0.9784	0.9788	0.9377
	(0.0007)	(0.0254)	(0.0001)	(0.0075)	(0.0074)	(0.0073)	(0.0072)	(0.0136)
b_4	0.8670	0.8386	0.9966	0.9970	0.9971	0.9875	0.9877	0.9830
	(0.0118)	(0.0251)	(0.0018)	(0.0027)	(0.0027)	(0.0043)	(0.0042)	(0.0139)
b_5	0.9908	0.9961	0.9911	0.9941	0.9944	0.9942	0.9946	0.9937
	(0.0018)	(0.0002)	(0.0034)	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0014)
ν	–	3.9158	–	4.1892	4.1883	4.1999	4.1997	3.9083
		(0.1484)		(0.1885)	(0.1885)	(0.1894)	(0.1894)	(0.1531)
Log-like	−25,784.4	−23,700.7	−25,768.0	−23,680.5	−23,680.5	**−23,678.3**	−23,678.8	−23,705.0
AIC	51,608.8	47,443.4	51,576.0	47,403.0	47,403.0	**47,398.7**	47,399.5	47,452.0
BIC	51,727.5	47,568.0	51,694.7	47,527.6	47,527.6	**47,523.3**	47,524.1	47,576.7

NOTE: The table contains the estimated parameters and their standard errors for nine alternative models, including the GARCH/cDCC. The prefix "t" on t-cDCC and t-GAS represents the Student t density, whereas "g" represents the Gaussian density. The hypersphere specification of R_t is denoted by the suffix "h," whereas the logarithmic specification for the variances is denoted by the additional suffix "l." The prefix "tg" on tg-GAS denotes a GAS model in which the likelihood is a Student t density, whereas the factor recursion in (2) is based on the Gaussian ($\nu^{-1} = 0$) density. The first four parameters, a_i and b_i, relate to the volatility dynamics. The parameters a_5 and b_5 relate to the correlation dynamics. The best log-likelihood, AIC, and BIC values across models are indicated in bold type.

Figure 1, we see that outliers appear to have a strong effect on estimated volatilities for the GARCH model, whereas those for the t-GAS model appear to be robust against outliers. The major difference in the factor recursions between the two models is the weight $w_t = (\nu + k)/(\nu - 2 + y_t' \Sigma_t^{-1} y_t)$, which is part of the score ∇_t in the t-GAS model.

We next investigate whether the improved fit of the t-GAS model compared with the g-GAS model is due to the Student t distribution or to the different factor recursion based on Theorem 1. For this purpose, we consider the tg-GAS model, which has the likelihood function based on the Student t density and its factor recursion based on the Gaussian density. In other words, we adopt the t-GAS model where s_t is evaluated as in Theorem 1 but with restriction $\nu^{-1} = 0$. The estimation results for the tg-GAS model are presented in the last column of Table 2. Comparing the t-GAS and tg-GAS models shows that the t-GAS factor recursion substantially contributes to the model fit. The t-GAS log-likelihood value is increased by approximately 30 points over the tg-GAS log-likelihood value, whereas both models have the same number of parameters. The log-likelihood value of the tg-GAS model is close to that for the t-GARCH/t-cDCC model because both models have much in common. Thus, we conclude that it is important that the likelihood function accounts for fat tails in the error distribution, and it is equally important to adjust the factor recursion for these fat tails.

Finally, the estimated coefficients $b_1, \ldots, b_4$ have lower values for the tg-GAS specification compared with those of the t-GAS model. Because incidental large squared residuals 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 $b_1, \ldots, b_4$. The weight w_t plays an important role in the Student t-based factor recursion and, thus takes care of these observations in a natural way. Therefore, the persistence parameters do not need to be reduced to improve the fit of the model. Thus underscores the importance of a robust form of the factor updating equation, as illustrated by the comparison of t-GAS and the tg-GAS models.

6.2 Out-of-Sample Performance

The results in Table 2 describe the in-sample performance of the different models. To verify the short-term forecasting capability of the different approaches, we compare the models in terms of correctly forecasting the 1% and 5% value at risk (VaR) at 1-day and 5-day horizons for different portfolios that can be constructed from the four equities. We define six different arbitrary portfolios, $p_{jt} = g_j y_t$, for given 4×1 weight

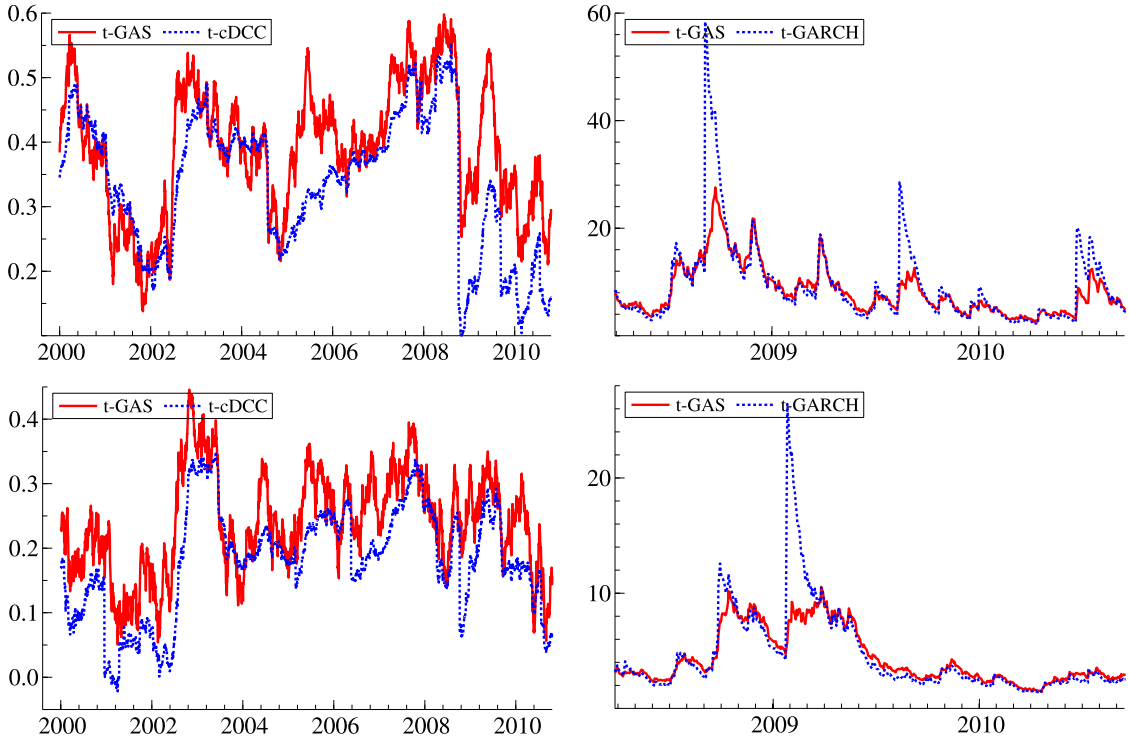

Figure 1. Comparison of the estimated correlation and volatilities. The left-side panels contain a comparison of the estimated correlations from the t-GAS$(1,1)$-l model versus the t-cDCC$(1,1)$ model. Top left, FRE/STT; bottom left, FRE/AET. The right-side panels compare the estimated volatilities from the t-GAS$(1,1)$-l model and t-GARCH$(1,1)$ model during the financial crisis. Top right, FRE; bottom right, STT. The online version of this figure is in color.

vectors g_j and for $j = 1, \ldots, 6$. By ordering the stocks as FRE, STT, AET, and CVH, we construct the following long-only and long-short portfolios: $g_1 = (0.25, 0.25, 0.25, 0.25)$, $g_2 = (0.4, 0.2, 0.3, 0.1)$, $g_3 = (0.5, 0.5, 0.5, -0.5)$, $g_4 = (0.5, -0.5, 0.5, 0.5)$, $g_5 = (0.5, 0.5, -0.5, -0.5)$, and $g_6 = (0.5, -0.5, 0.5, -0.5)$. The long-short positions reflect relative value bets within or between the banking and insurance sectors, respectively. The forecasting experiment covers the last $T^* = 500$ days in our dataset and thus includes the financial crisis. We use a rolling window of length 1250 days for estimating the parameters of the model. Given the parameter estimates at time t, we simulate 10,000 sample paths for $y_{t+1}, \ldots, y_{t+5}$, denoted by $y^s_{t+1}, \ldots, y^s_{t+5}$ for $s = 1, \ldots, 10{,}000$, using the multivariate Student t distribution with time-varying variances and correlations (as implied by the different model specifications). We then construct the simulated portfolio returns $p^s_{j,t+k} = g_j' y^s_{t+k}$ for $j = 1, \ldots, 6$ and $k = 1, \ldots, 5$. We use the sample of 10,000 simulated paths for portfolio returns $p^s_{j,t+1}, \ldots, p^s_{j,t+5}$ to estimate the quantiles of the forecasting distribution at the 1-day and 5-day horizons. We repeat these computations for each t, which includes estimation of parameters and simulation of 10,000 new sample paths for computing the portfolio forecasting quantiles.

The VaR estimate for the jth portfolio at time t with coverage level α is denoted by $\mathrm{VaR}_{jt}(\alpha)$. To evaluate the forecasting performance of the models, we require the indicator function

$$\bar{I}_{jt} = \begin{cases} 1 & \text{if } p_{j,t} < \mathrm{VaR}_{j,t}(\alpha) \\ 0 & \text{if } p_{j,t} \geq \mathrm{VaR}_{j,t}(\alpha), \end{cases}$$

$\mathrm{VaR}_{j,t}(\alpha)$ at time t. Note where the $\mathrm{VaR}_{j,t}(\alpha)$ is estimated by the quantile of the 10,000 simulated paths for $p^s_{j,t+1}$ and $p^s_{j,t+5}$, as described earlier. The overall number of exceedances for the jth portfolio in our forecasting sample is measured as $\gamma_j = \sum_{t=1}^{T^*} \bar{I}_{jt}$, with the average number of VaR exceedences given by $\tilde{\alpha}_j = \gamma_j / T^*$. The forecast coverage $\tilde{\alpha}_j$ should be equal to the unconditional coverage α. We adopt the test statistic of Kupiec (1995) to test the null hypothesis $H_0 : \tilde{\alpha}_j = \alpha$ of a correct coverage against the alternative $H_1 : \tilde{\alpha}_j \neq \alpha$ for each portfolio $j = 1, \ldots, 6$. The likelihood ratio test statistic of Kupiec is given by

$$LR_K = 2\left\{\log\left[\hat{\alpha}^{\gamma}(1-\hat{\alpha})^{T^*-\gamma}\right] - \log\left[\alpha^{\gamma}(1-\alpha)^{T^*-\gamma}\right]\right\},$$

and is $\chi^2(1)$ distributed asymptotically. The same test for evaluating forecast performance was used in a similar context by Bauwens and Laurent (2005) and Chib, Nardari, and Shephard (2006).

Table 3 presents the average number of VaR exceedences α_j and the p-values of the likelihood ratio test of Kupiec (1995). Two main findings emerge from these forecasting results. First, for almost all portfolios in the forecasting experiment, the t-GAS models perform better in general. The p-values for rejecting the hypothesis of correct coverage are substantially higher overall, indicating that the t-GAS models obtain better coverage. The exception is the portfolio with weight vector g_6 at the 5% coverage level, for which both the t-cDCC and the t-GAS models perform poorly. At the 1% level for portfolio 6, the t-GAS outperforms the t-cDCC in out-of-sample coverage. Second, the log-specification for the variances is important for

Table 3. Average number of VaR exceedences and p-values for the Kupiec test of correct coverage

	1%, $h=1$	1%, $h=5$	5%, $h=1$	5%, $h=5$	1%, $h=1$	1%, $h=5$	5%, $h=1$	5%, $h=5$
	portfolio g_1				portfolio g_2			
t-cDCC	2.00 (0.05)	3.00 (0.00)	8.20 (0.00)	8.80 (0.00)	1.80 (0.11)	2.40 (0.01)	7.40 (0.02)	8.60 (0.00)
t-GAS	2.00 (0.05)	1.60 (0.21)	6.20 (0.23)	7.00 (0.05)	1.60 (0.21)	2.00 (0.05)	5.80 (0.42)	5.60 (0.55)
t-GAS-l	1.80 (0.11)	1.40 (0.40)	5.80 (0.42)	6.00 (0.32)	1.60 (0.21)	1.60 (0.21)	5.20 (0.84)	4.60 (0.68)
t-GAS-h	2.00 (0.05)	1.80 (0.11)	6.60 (0.12)	7.20 (0.03)	1.60 (0.21)	1.80 (0.11)	5.80 (0.42)	5.60 (0.55)
t-GAS-hl	1.80 (0.11)	1.40 (0.40)	5.80 (0.42)	5.80 (0.42)	1.60 (0.21)	1.60 (0.21)	5.20 (0.84)	4.80 (0.84)
	portfolio g_3				portfolio g_4			
t-cDCC	1.40 (0.40)	2.20 (0.02)	4.00 (0.29)	6.00 (0.32)	2.00 (0.05)	1.80 (0.11)	5.40 (0.69)	6.40 (0.17)
t-GAS	1.60 (0.21)	1.60 (0.21)	4.20 (0.40)	4.80 (0.84)	1.80 (0.11)	1.40 (0.40)	6.00 (0.32)	5.60 (0.55)
t-GAS-l	1.40 (0.40)	1.60 (0.21)	3.60 (0.13)	3.80 (0.20)	1.80 (0.11)	1.40 (0.40)	5.20 (0.84)	4.40 (0.53)
t-GAS-h	1.60 (0.21)	1.80 (0.11)	4.00 (0.29)	4.80 (0.84)	2.00 (0.05)	1.40 (0.40)	6.20 (0.23)	5.60 (0.55)
t-GAS-hl	1.20 (0.66)	1.60 (0.21)	3.60 (0.13)	4.20 (0.40)	1.80 (0.11)	1.40 (0.40)	5.00 (1.00)	4.20 (0.40)
	portfolio g_5				portfolio g_6			
t-cDCC	1.40 (0.40)	2.00 (0.05)	4.00 (0.29)	5.00 (1.00)	1.60 (0.21)	3.00 (0.00)	3.80 (0.20)	5.00 (1.00)
t-GAS	1.60 (0.21)	1.80 (0.11)	5.20 (0.84)	5.80 (0.42)	1.80 (0.11)	3.00 (0.00)	5.00 (1.00)	5.20 (0.84)
t-GAS-l	1.40 (0.40)	1.80 (0.11)	4.60 (0.68)	5.20 (0.84)	1.80 (0.11)	2.60 (0.00)	5.00 (1.00)	4.80 (0.84)
t-GAS-h	1.40 (0.40)	1.80 (0.11)	5.20 (0.84)	5.80 (0.42)	2.00 (0.05)	3.00 (0.00)	5.00 (1.00)	5.20 (0.84)
t-GAS-hl	1.60 (0.21)	1.80 (0.11)	4.80 (0.84)	5.60 (0.55)	1.80 (0.11)	2.20 (0.02)	5.00 (1.00)	4.60 (0.68)

NOTE: The average numbers of VaR exceedences $\tilde{\alpha}_j$ are reported in percentages, with the p-values for the Kupiec test of correct coverage in parentheses. Forecasts are computed for six different portfolios at the 1% and 5% VaR levels and for horizons of $h=1$ and $h=5$ days.

out-of-sample analyses. The increase in p-values over those of the regular t-GAS is often substantial. As before, the hypersphere parameterization for the correlations or the standard parameterization have little effect on the forecasting accuracy. If anything, it appears that the hypersphere parameterization performs slightly worse than the regular parameterization for correlations. Overall, we conclude that the t-GAS-l model has the best out-of-sample forecasting behavior.

7. CONCLUSION

We have introduced the multivariate Student t GAS model for volatilities and correlations, where the multivariate normal distribution is included as a special case. The model includes several levels of flexibility. We have shown how the GAS framework can accommodate alternative specifications of the covariance matrix. The model formulation is sufficiently general to impose a factor structure on time-varying volatilities, correlations, or both. We have focused on the decomposition of the conditional covariance matrix into a volatility matrix and a correlation matrix. Future work 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 possible extension is to consider the multivariate skewed Student t distribution 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. These and other extensions provide interesting avenues for further research.

APPENDIX

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

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

where z_i denotes the ith element of z and where Kronecker delta δ_{ij} is unity when $i=j$ and 0 otherwise. The $k^2 \times k^2$ matrix G is implicitly defined by

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

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

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

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

Result 2. Let the random variable $u \sim \chi^2_\nu$ where χ^2_ν is the chi-squared distribution with ν degrees of freedom. For any scalar $a < \nu$, we have

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

Proof of Theorem 1. From the basic matrix calculus results of Abadir and Magnus (2005), and based on the notation and definitions in Section 2.4, it is straightforward to derive (8) as

$$\begin{aligned}
&\frac{\partial \log p(y_t|\Sigma_t;\nu)}{\partial \operatorname{vech}(\Sigma_t)} \\
&= \frac{1}{2}\mathcal{D}_k'(\Sigma_t^{-1}\otimes\Sigma_t^{-1}) \\
&\quad \times \left[\frac{(\nu+k)}{(\nu + y_t'\Sigma_t^{-1}y_t - 2)} y_t \otimes y_t - \operatorname{vec}(\Sigma_t)\right] \\
&= \frac{1}{2}\mathcal{D}_k' J_{t\otimes}'[w_t \bar{y}_{t\otimes} - \operatorname{vec}(\mathrm{I})], \\
w_t &= \frac{(1+\nu^{-1}k)}{(1+\nu^{-1}(\bar{y}_t'\bar{y}_t - 2))}, \quad \text{(A.3)}
\end{aligned}$$

where $\bar{y}_t = J_t y_t$ follows a Student t distribution with mean 0, covariance matrix I_k, and ν degrees of freedom. The derivative

of $\text{vech}(\Sigma_t)$ with respect to f_t is denoted by Ψ_t and completes the result for ∇_t.

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

$$\text{E}\big[(w_t \bar{y}_{t\otimes} - \text{vec}(\text{I}))(w_t \bar{y}_{t\otimes} - \text{vec}(\text{I}))'\big] = gG - \text{vec}(\text{I})\,\text{vec}(\text{I})'.$$

Because we take expectations with respect to the Student t density $p(y_t|\Sigma_t;\nu)$, we have

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

where $\bar{x}_t$ has a Student t distribution with mean 0, covariance matrix $(\nu-2)\text{I}/(\nu+2)$, and $\nu+4$ degrees of freedom. In addition, because $\bar{x}_t = x_t/\sqrt{u/(\nu+4)}$ with x_t a mean 0 normal with covariance matrix $(\nu-2)\text{I}/(\nu+4)$, and u an independent $\chi^2_{\nu+4}$ random variable, we use Results 1 and 2 of this Appendix to express (A.4) as

$$\begin{aligned}
&\frac{(\nu+k)^2}{(\nu-2)^2} \frac{(\nu+2)\nu}{(\nu+2+k)(\nu+k)} \cdot \frac{(\nu-2)^2}{(\nu+4)^2} \cdot \frac{(\nu+4)^2}{(\nu+2)\nu} \cdot G \\
&\qquad = \frac{\nu+k}{\nu+2+k} G. \qquad \text{(A.5)}
\end{aligned}$$

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

ACKNOWLEDGMENTS

We would like to thank Frank Diebold, Ruey Tsay, and seminar participants at the University of Pennsylvania, and Tinbergen Institute Amsterdam for comments on earlier versions of this paper. We thank the associate editor and two referees for constructive comments. Financial support to the first author from the Booth School of Business is gratefully acknowledged. All computations reported in this article were carried out using the OxMetrics 6.0 programming environment of Doornik (2009).

[Received March 2010. Revised January 2011.]

REFERENCES

Abadir, K. M., and Magnus, J. R. (2005), *Matrix Algebra*, New York: Cambridge University Press. [554,562]

Aielli, G. P. (2008), "Consistent Estimation of Large Scale Dynamic Conditional Correlations," technical report, Dept. of Statistics, University of Florence. [556,558,559]

Bauwens, L., and Laurent, S. (2005), "A New Class of Multivariate Skew Densities, With Application to Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business & Economic Statistics*, 23 (3), 346–354. [561,562]

Bauwens, L., Laurent, S., and Rombouts, J. (2006), "Multivariate GARCH Models: A Survey," *Journal of Applied Econometrics*, 21 (1), 79–109. [552,558]

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. [552,553]

——— (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *The Review of Economics and Statistics*, 69 (3), 542–547. [554]

Brownlees, C. T., and Engle, R. (2010), "Volatility, Correlation, and Tails for Systematic Risk Measurement," technical report, Dept. of Finance, Stern School of Business, New York University. [558,559]

Chib, S., Nardari, F., and Shephard, N. (2006), "Analysis of High-Dimensional Multivariate Stochastic Volatility Models," *Journal of Econometrics*, 134 (2), 341–371. [552,561]

Creal, D. D., Koopman, S. J., and Lucas, A. (2010), "Generalized Autoregressive Score Models With Applications," working paper, University of Chicago Booth School of Business. [553,555]

Ding, Z., and Engle, R. F. (2001), "Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing," *Academia Economic Papers*, 29, 157–184. [552]

Doornik, J. A. (2009), *Object-Oriented Matrix Programming Using Ox 6.0*, London: Timberlake Consultants Press. [563]

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50 (4), 987–1007. [552]

——— (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business & Economic Statistics*, 20 (3), 339–350. [552,555,556,558]

Engle, R. F., and Gallo, G. M. (2006), "A Multiple Indicators Model for Volatility Using Intra-Daily Data," *Journal of Econometrics*, 131, 3–27. [553]

Engle, R. F., and Kelly, B. (2009), "Dynamic Equicorrelation," unpublished manuscript, Stern School of Business, New York University. [558]

Engle, R. F., and Russell, J. R. (1998), "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66 (5), 1127–1162. [553]

Engle, R. F., and Sheppard, K. (2001), "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," unpublished manuscript, University of California San Diego. [552,556]

Fan, J., Wang, M., and Yao, Q. (2008), "Modelling Multivariate Volatilities via Conditionally Uncorrelated Components," *Journal of the Royal Statistical Society, Ser. B*, 70 (4), 679–702. [552]

Gourieroux, C., Jasiak, J., and Sufana, R. (2009), "The Wishart Autoregressive Process of Multivariate Stochastic Volatility," *Journal of Econometrics*, 150 (2), 167–181. [552]

Harvey, A. C., and Chakravarty, T. (2008), "Beta–t–(E)GARCH," unpublished manuscript, Faculty of Economics, Cambridge University. [554]

Jaeckel, P., and Rebonato, R. (2000), "The Most General Methodology for Creating a Valid Correlation Matrix for Risk Management and Option Pricing Purposes," *Journal of Risk*, 2 (2), 17–28. [557]

Kupiec, P. (1995), "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 2, 173–184. [561]

Lee, T.-H., and Long, X. (2009), "Copula-Based Multivariate GARCH Model With Uncorrelated Dependent Errors," *Journal of Econometrics*, 150 (2), 207–218. [552]

Nelson, D. B., and Foster, D. P. (1994), "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica*, 62 (1), 1–41. [553,554]

Patton, A. J. (2006), "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47 (2), 527–556. [552]

——— (2009), "Copula-Based Models for Financial Time Series," in *Handbook of Financial Time Series*, eds. J. P. K. T. G. Anderson, R. A. Davis, and T. Mikosch, New York: Springer-Verlag, pp. 767–786. [556]

Sklar, A. (1959), "Fonctions de répartition á n dimensions et leurs marges," *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229–231. [556]

Tsay, R. S. (2005), *Analysis of Financial Time Series* (2nd ed.), New York: Wiley. [552,558]

Tse, Y. K., and Tsui, A. K. C. (2002), "A Multivariate GARCH Model With Time-Varying Correlations," *Journal of Business & Economic Statistics*, 20, 351–362. [552,556]

van der Weide, R. (2002), "GO–GARCH: A Multivariate Generalized Orthogonal GARCH Model," *Journal of Applied Econometrics*, 17, 549–564. [557]