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 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 conditionally 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
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 \( v \) degrees of freedom. To simplify the notation, we set the location parameter, \( \mu_t \), 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

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

where \( \Sigma_t \) is the covariance matrix of \( y_t \). We assume that \( v > 2 \), such that the covariance matrix exists; however, it is straightforward to generalize the model to the case where \( 0 < v \leq 2 \) 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 \( \Sigma_t \) 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

\[
f_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1},
\]

where \( \omega \) 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 \( \theta \), 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,

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

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,

\[
S_t = \mathcal{I}_t^{-1}, \quad \mathcal{I}_t^{-1} = E_{t-1} [\nabla_t \nabla_t']
\]

such that Var(\( s_t \)) = \( \mathcal{I}_t^{-1} \). 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 \( \mathcal{I}_t^{-1/2} \) rather than on \( \mathcal{I}_t^{-1} \). Our choice to set \( S_t = \mathcal{I}_t^{-1/2} \) brings our GAS model closest to the familiar GARCH specification. Consider the case of a univariate series \( y_t \) that is normally distributed, \( y_t \sim N(0, \sigma_t^2) \). We let \( f_t = \sigma_t^2 \) and adopt the scaling (4) such that \( \nabla_t = -1/2f_t^{-1} + 1/2f_t^{-2} y_t \) and \( S_t = 2f_t^2 \). When we scale by the inverse information matrix, the updating equation (2) with \( p = q = 1 \) becomes

\[
f_{t+1} = \omega + A_1(y_t^2 - f_t) + B_1f_t,
\]

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

\[
f_{t+1} = \omega + A_1 \cdot (1 + 3v^{-1}) \cdot \left( \frac{(1 + v^{-1})}{(1 - 2v^{-1})(1 + v^{-1})y_t^2 / ((1 - 2v^{-1}) \cdot f_t)} y_t^2 - f_t \right)
+ B_1f_t,
\]

(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 $v$ 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 parameterized 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 $\alpha_1$ and $\beta_1$ 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 $A_1$ and $B_1$ 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 $A_1$ coefficients, whereas the $B_1$ 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 $A_\odot = A \otimes A$. The operator vec$(A)$ vectorizes matrix $A$ into a column vector, whereas vec$(A)$ 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 $[D_k \text{vec}(A) = \text{vec}(A)]$, elimination $[B_k \text{vec}(A) = \text{vec}(A)]$, and commutation $[C_k \text{vec}(A) = \text{vec}(A')]$ 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 $\Sigma_t$ 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 $\Sigma_t$ in detail in Sections 3 and 4. A simple example of such a link function is $f_t' = [\text{diag}(D_t r_t, \text{vec}(Q_t))]$ 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 $R_t = \text{diag}(Q_t)^{-1/2} \text{diag}(Q_t)^{-1}$. 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

$$\Psi_t = \Psi(f_t) = \frac{\partial \text{vec}(\Sigma_t)}{\partial f_t} \text{ for } \Sigma_t = \Sigma(f_t). \quad (7)$$

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

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

$$\nabla_t = \frac{\partial \log p_t(y_t | \Sigma_t; v)}{\partial f_t}$$

$$= \frac{1}{2} \Psi'_t D_t' \Sigma_t^{-1} [w_t y_{t\odot} - \text{vec}(\Sigma_t)]. \quad (8)$$

$$J_t = E[\nabla_t \nabla_t']$$

$$= \frac{1}{4} \Psi'_t D_t' J'_t G^{-1} g G - \text{vec}(I) \text{vec}(I)] J_t \odot D_t \Psi_t, \quad (9)$$

where $D_t$ is the duplication matrix and with scalar $w_t = (v + k)/(v - 2 + y_t \Sigma_t^{-1} y_t)$, matrix $J_t$ defined implicitly as $J_t^{-1} = J'_t I$, scalar $g = (v + k)/(v + 2 + k)$, 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 $w_t y_{t\odot} y_t'$ from the local covariance matrix $\Sigma_t$. 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 $\Sigma_t$ are accounted for by the matrix function $\Psi_t$, which gathers the derivatives of the full covariance matrix $\Sigma_t$ with respect to the 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. Thus the GAS framework can accommodate a wide
class of models for time-varying covariance matrices or, alter-
natively, for time-varying correlation matrices under a copula
specification with known variances. Some other examples of
decomposition of the covariance matrix are discussed in Sec-
tions 3 and 4.

Third, the weight \(w_t\) in (8) also appears in the univariate spec-
ification of (6) with \(k = 1\). If the density of the observations \(y_t\)
is heavy-tailed (\(v^{-1} > 0\)), then large values in \(y_t\) (in absolute
terms) do not automatically lead to dramatic changes in the ele-
ments of \(\Sigma_t\). 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) au-
tomatically accounts for extreme values because it decreases if
\(y_t^2 \Sigma_t^{-1} y_t\) is large.

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

\[
f_t = \omega + \sum_{i=1}^{\infty} B_i A_{i, t-1}.
\]

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 in-
side the unit circle. Generally, however, with inverse informa-
tion matrix scaling, that is \(S_t = T_{t-1}^{-1}\), the conditional variance
of \(s_t\) will not be constant, and formulating conditions for a sta-
 tionary process will be much harder. Creal, Koopman, and Lu-
cas (2010) addressed this issue by scaling with \(S_t = T_{t-1}^{-1/2}\) in-
stead. In the developments that follow, we want to stay close to
the multivariate GARCH framework, which requires that the
score be scaled by \(T_{t-1}^{-1}\), 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 im-
 pact of observations corresponding to large values of \(y_t^2 \Sigma_t^{-1} y_t\)
on the updating of correlations. Such a mechanism is absent in
the DCC model. Thus, our current specification differs substan-
tially from the DCC specification with Student \(t\)–distributed er-
ror terms. Moreover, when we consider the normal distribution
\((v^{-1} = 0\) 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, unex-
epected deviations \(w_t y_t y_t' - \Sigma_t\) drive the evolution of volatil-
ities 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 nor-
mal distribution \((v^{-1} = 0\)). We have \(\text{vec}(\Sigma_t) = (1, \rho_t, \rho_t, 1, 1)'\).

The key component in updating the time-varying correla-
tion is the score \(V_t\) in Theorem 1, for which it follows that
\(\Psi_t / D_t^k = (0, 1, 1, 0)\). This does not imply that only the sec-
ond and third elements of \(w_t y_t y_t' - \text{vec}(\Sigma_t)\) (the cross-prod-
tect terms), are taken into account. Because \(\Psi_t / D_t^k\) is first post-multi-
plied by \(\Sigma_t^{-1} \otimes \Sigma_t^{-1}\), we obtain

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

where \(y_{it}\) is the \(i\)th 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 cor-
relation when \(y_{1t} y_{2t}\) exceeds \(\rho_t\). 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 vec-
tors 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. Con-
sequently, even though the cross-products are the same, the sec-
ond 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 in-
creases 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 multivari-
ate Student \(t\) density with different decompositions of the time-
varying covariance matrix \(\Sigma_t\). 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 ma-
trix by

\[
\Sigma_t = D_t R_t D_t',
\]

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

\[
R_t = \Delta_t^{-1} Q_t \Delta_t^{-1},
\]

where \(Q_t\) is a symmetric positive definite matrix and \(\Delta_t\) 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

$$Q_{t+1} = \Omega_{\text{dcc}}(1 - A_{\text{dcc}} - B_{\text{dcc}}) + A_{\text{dcc}} \odot \hat{e}^t \hat{e}^t + B_{\text{dcc}} \odot Q_t,$$  

(14)

where $\hat{e}_t$ is a $k \times 1$ vector with elements $\hat{e}_t = y_t/\sigma_t$, $\odot$ is the Hadamard product (element by element multiplication) and $k \times k$ matrices $Q_0$, $\Omega_{\text{dcc}}$, $A_{\text{dcc}}$, and $B_{\text{dcc}}$ are fixed and unknown. The covariance matrix $Q_t$ is positive definite for all $t$ when $\Omega_{\text{dcc}}$ 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 $A_{\text{dcc}}$ and $B_{\text{dcc}}$, 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 $\Omega_{\text{dcc}}$ with the sample correlation matrix of the standardized residuals $y_{it}/\sigma_{it}$ and by restricting $A_{\text{dcc}} = a_{\text{dcc}} \cdot \mu^t$ and $B_{\text{dcc}} = b_{\text{dcc}} \cdot \mu^t$, where $a_{\text{dcc}}$ and $b_{\text{dcc}}$ are scalars and $\mu$ is a vector of 1s. Aielli (2008) showed that this approach of preestimating $\Omega_{\text{dcc}}$ 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 $\hat{e}_t$ are defined as $\hat{e}_t = (y_t/\sqrt{\hat{\sigma}_t})$ instead of $\hat{e}_t = y_t/\sigma_t$, where $\hat{\sigma}_t$ is the $i$th 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 $\Sigma_t$ given by (12) and (13), we specify our factor as

$$f_t = \left( \text{diag}(D_t^2) \right)^1.$$  

(15)

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

$$\Psi_t = \frac{\partial \text{vec}(\Sigma_t)}{\partial f_t} = B_k \frac{\partial \text{vec}(D_tR_tD_t)}{\partial f_t} = B_k \frac{\partial \text{vec}(\Delta_{\text{dcc}}^{-1}Q_t\Delta_{\text{dcc}}^{-1})}{\partial f_t},$$

$$= B_k(1 \otimes D_tR_t) \frac{\partial \text{vec}(D_t)}{\partial f_t} + B_kD_{\otimes} \frac{\partial \text{vec}(\Delta_{\text{dcc}}^{-1}Q_t\Delta_{\text{dcc}}^{-1})}{\partial f_t},$$

$$= B_k(1 \otimes D_tR_t) \frac{\partial \text{vec}(D_t)}{\partial f_t} + B_kD_{\otimes} \Delta_{\text{dcc}}^{-1}$$

$$\times \left( D_k \frac{\partial \text{vec}(Q_t)}{\partial f_t} - (\Delta_{\otimes} \otimes Q_t) \Delta_{\text{dcc}}^{-1} \frac{\partial \text{vec}(\Delta_t)}{\partial f_t} \right).$$

Define the matrices $W_{\Delta t}, W_{Dt}, S_{\Delta}, S_{Dt},$ and $S_Q$ 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 $S_{\Delta}, S_{Dt},$ and $S_Q$ are selection matrices containing only 1s and 0s such that $\text{diag}(\Delta_t^{-1}) = S_{\Delta} \text{vec}(\Delta_t)$, $\text{diag}(D_t^{-1}) = S_{Dt}$, and $\text{vec}(\Delta_t) = S_{Q}f_t$. We then obtain

$$\Psi_t = B_k(I \otimes D_tR_t) \cdot W_{Dt}S_{\Delta}$$

$$+ B_k D_{\otimes} \Delta_{\text{dcc}}^{-1} \left[ D_k - (\Delta_{\otimes} \otimes Q_t) \Delta_{\text{dcc}}^{-1} W_{\Delta t}S_{\Delta} \right] S_{Q},$$  

(16)

The updating of time-varying $\Sigma_t$ is implied by (2) and Theorem 1, with $\Psi_t$ given by (16). If only the correlations in $R_t$ are time-varying, then $S_{Dt} = 0$, 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 specific specification by setting $S_{Q} = 0$ and the second term in (16) vanishes.

The current specification where the vector of factors $f_t$ contains $\text{vec}(Q_t)$ 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 $\text{vec}(Q_t)$ 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 $\omega$ in the GAS model or $k$ elements in $\Omega_{\text{dcc}}$ of the DCC model to identify either model. In particular, the elements of $\omega$ (or $\Omega_{\text{dcc}}$) 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 $\omega$ 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 $D_1 = I$ and a time-varying correlation that is updated by (11). We then have

$$\Psi_t = \frac{1}{2\sqrt{q_{11,t}q_{22,t}}} \left( \begin{array}{cc} 2 & -q_{12,t}/q_{22,t} \\ -q_{12,t}/q_{22,t} & 2 \end{array} \right) \left( \begin{array}{c} 0 \\ 1 \end{array} \right),$$  

(17)

where $q_{ij,t}$ is the $(i, j)$ element of $Q_t$ for $i, j = 1, 2$. The updating mechanism for each element of $f_t = \text{vec}(Q_t)$ 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 $\hat{y}_{1,t}^2, \hat{y}_{1,t}y_{2,t},$ and $\hat{y}_{2,t}^2$, 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 $u_t$ denote the probability integral transform of $y_t$, which is the $i$th series in $y_t$. The Student $t$ density then operates on the $k \times 1$ vector $y_t = [P_v^{-1}(u_{1,t}), \ldots, P_v^{-1}(u_{k,t})]'$ via Sklar’s theorem (Sklar 1959), where $P_v^{-1}()$ is the univariate inverse Student $t$ distribution with $v$ degrees of freedom. In the copula specification, the decomposition (12) simplifies because $D_t \equiv I$. We obtain

$$\Psi_t = B_kD_{\otimes} \Delta_{\text{dcc}}^{-1} \left[ D_k - (\Delta_{\otimes} \otimes Q_t) \Delta_{\text{dcc}}^{-1} W_{\Delta t}S_{\Delta} \right] S_{Q},$$  

(18)
where \( S_Q = \mathbf{I} \) because \( f_t = \vech(Q_t) \). 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 \( \nu \), which must be estimated.

4. ALTERNATIVE DECOMPOSITIONS OF THE COVARIANCE MATRIX

In this section we consider alternative decompositions of 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 (12), we can, for example, define the vector of factors \( f_t \) as

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

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

\[
    \Psi_t = B_t(I \oplus D_t R_t)W_{D_t} D_t^2 S_D \\
    + B_t D_t \otimes \Delta_Q^{-1} [D_k - (\Delta_t \oplus Q_t) \Delta_Q^{-1} W_{D\Delta} S_{\Delta}] S_Q,
\]

which is the same as \( \Psi_t \) 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 \( \text{diag}(D_t^2) \) with respect to \( f_t \). In the setting where \( R_t = R, S_D \) is the identity matrix and \( S_Q = 0 \).

4.2 Time-Varying Correlations Based on Hyperspherical Coordinates

In this section we provide an alternative decomposition of the correlation matrix \( \Sigma_t \). The covariance matrix decomposition \( \Sigma_t = D_t R_t D_t^\top \) 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 \( R_t = X_t^\top X_t \), where \( X_t \) originates from the upper-triangular matrix, that is,

\[
    X_t(\phi_t) = \begin{pmatrix} 1 & c_{12t} & c_{13t} & \cdots & c_{1kt} \\ 0 & c_{23t} & c_{23t} & \cdots & c_{2kt} \end{pmatrix}.
\]

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 \( i, j = 1, \ldots, k \). 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 \( f_t = [\text{diag}(D_t^2)'] \phi_t \). 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 required to the previous specifications is the form of \( \Psi_t \). We use the derivatives

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

(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

\[
    Z_t = \frac{\partial \text{vec}(X_t)}{\partial \phi_t'}.
\]

In the GAS framework, define the selection matrix \( S_{\phi} \) such that \( \phi_t = S_{\phi} \phi_t \). Also note that

\[
    \frac{\partial \text{vec}(R_t)}{\partial \phi_t'} = \frac{\partial \text{vec}(X_t^\top X_t)}{\partial \phi_t'} S_{\phi} = [(I \otimes X_t') + (X_t' \otimes I) C_k] Z_t S_{\phi},
\]

(23) where \( C_k \) is the commutation matrix. Combining these results, we obtain

\[
    \Psi_t = B_t(I \oplus D_t R_t)W_{D_t} D_t^2 S_D \\
    + B_t D_t \otimes [(I \otimes X_t') + (X_t' \otimes I) C_k] Z_t S_{\phi}.
\]

(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 \( \nu \) are identified, and the scaling matrix \( S_t \) in (4) is nonsingular. When the dimension \( k \) of \( \gamma_t \) 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, \( f_t' = [\text{diag}(D_t')', \text{vech}(Q_t)'] \). With a slight abuse of notation, we have

\[
\text{diag}(D_t') = a + S_0 f_t, \quad \text{vech}(Q_t) = b + S_0 f_t, \tag{25}
\]

with vectors \( a \) and \( b \), and where \( S_D \) and \( S_Q \) 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 \( \Psi_t \) is now the same as in (16).

For the hyperspherical coordinate specification, the result remains (24), with \( S_0 \) 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 \( S_D \) and \( S_Q \) (or \( S_0 \)). 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) \( S_0 \). 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 \( \Sigma_t \) is determined by the GAS updating equations for \( D_t \) and \( R_t \). For given values of \( \nu, \omega, A_1, B_1 \), 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 \( \theta \), and its estimation is based on the maximization of the log-likelihood with respect to \( \theta \). The initial values \( f_j (j < 1) \) are set to their unconditional expectation, \( (I_k - B_1 - \cdots - B_q)\omega_0 \). 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 \( \rho_t \). We consider the following time-varying patterns for \( \rho_t \):

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

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

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

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 \( \rho_t \).

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 \( \nu, \omega, A_1, \) 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

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{\rho}_i - \rho_t|, \quad \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\rho}_i - \rho_t)^2. \tag{27}
\]

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 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

<table>
<thead>
<tr>
<th></th>
<th>Constant</th>
<th>Sine</th>
<th>Fast Sine</th>
<th>Step</th>
<th>Ramp</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-GAS</td>
<td>0.7661</td>
<td>0.9071</td>
<td>1.0252</td>
<td>0.9460</td>
<td>0.9615</td>
<td>0.9358</td>
</tr>
<tr>
<td>t-GAS-h</td>
<td>0.7150</td>
<td>0.8982</td>
<td>1.0354</td>
<td>0.9345</td>
<td>0.9677</td>
<td>0.9354</td>
</tr>
<tr>
<td>t-cDCC</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>EWMA</td>
<td>3.029</td>
<td>1.2416</td>
<td>1.1126</td>
<td>1.0648</td>
<td>1.1067</td>
<td>1.0949</td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-GAS</td>
<td>0.44649</td>
<td>0.8844</td>
<td>1.0342</td>
<td>0.9718</td>
<td>1.0031</td>
<td>0.8887</td>
</tr>
<tr>
<td>t-GAS-h</td>
<td>0.3811</td>
<td>0.8595</td>
<td>1.0455</td>
<td>0.9443</td>
<td>0.9981</td>
<td>0.8862</td>
</tr>
<tr>
<td>t-cDCC</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>EWMA</td>
<td>8.4414</td>
<td>1.4465</td>
<td>1.3049</td>
<td>1.1458</td>
<td>1.2557</td>
<td>1.2149</td>
</tr>
</tbody>
</table>

NOTE: The table presents the MAE and MSE for the estimated dynamic correlation patterns for four 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 Audri (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 Datasream. 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 heavy-tailed performance as well as in an out-of-sample forecasting comparison. We evaluate the models in terms of in-sample performance.

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 $\alpha_i = A_1$ and $\beta_i = B_1 - A_1$ 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 $A_{dcc} = a_5 \cdot u_i$ and $B_{dcc} = b_5 \cdot u_i$ in (14). The variance and correlation parameters of the GAS model are modeled simultaneously as follows. The first elements in vector $f_i$ represent the variances in, for example, $\text{diag}(D^2)$ of the decomposition (12). The remaining elements of $f_i$ represent the correlations in $Q_i$, as defined in (13) or the hyperspherical angles in $\phi_i$, as defined after (21). The dynamic specification of $f_i$ 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 empirical study with four series, the diagonal elements of $A_1$ and $B_1$ are given by $a_1, \ldots, a_5$ and $b_1, \ldots, b_5$, respectively. The parameters are estimated 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 $f_i$ 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 $f_i$. The model t-GAS-h refers to the t-GAS model with a hypersphere specification of the variance matrix $\Sigma = D_i X_i X_i D_i$, where $X_i$ 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...
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 \( w_t = (v + k)/(v - 2 + \gamma_t \Sigma_t^{-1} y_t) \), which is part of the score \( \theta_t \) 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 \( v^{-1} = 0 \). 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 \( w_t \), 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. This 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 \( 4 \times 1 \) weight

---

**Table 2. Parameter estimates, likelihoods, and information criteria**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>g-DCC</th>
<th>t-DCC</th>
<th>g-GAS</th>
<th>t-GAS</th>
<th>t-GAS-h</th>
<th>t-GAS-l</th>
<th>t-GAS-h-l</th>
<th>tg-GAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>0.0534</td>
<td>0.1068</td>
<td>0.0523</td>
<td>0.0670</td>
<td>0.0670</td>
<td>0.0684</td>
<td>0.0685</td>
<td>0.1072</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>(0.0043)</td>
<td>(0.0136)</td>
<td>(0.0043)</td>
<td>(0.0080)</td>
<td>(0.0080)</td>
<td>(0.0079)</td>
<td>(0.0079)</td>
<td>(0.0137)</td>
</tr>
<tr>
<td>( a_3 )</td>
<td>0.0297</td>
<td>0.0856</td>
<td>0.0360</td>
<td>0.0566</td>
<td>0.0565</td>
<td>0.0517</td>
<td>0.0517</td>
<td>0.0857</td>
</tr>
<tr>
<td>( a_4 )</td>
<td>(0.0022)</td>
<td>(0.0146)</td>
<td>(0.0023)</td>
<td>(0.0076)</td>
<td>(0.0075)</td>
<td>(0.0066)</td>
<td>(0.0066)</td>
<td>(0.0141)</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>0.0113</td>
<td>0.1030</td>
<td>0.1111</td>
<td>0.0639</td>
<td>0.0633</td>
<td>0.0589</td>
<td>0.0585</td>
<td>0.0963</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>0.1285</td>
<td>0.1307</td>
<td>0.1292</td>
<td>0.0652</td>
<td>0.0648</td>
<td>0.0602</td>
<td>0.0600</td>
<td>0.1345</td>
</tr>
<tr>
<td>( b_3 )</td>
<td>0.0036</td>
<td>0.0032</td>
<td>0.0054</td>
<td>0.0113</td>
<td>0.0111</td>
<td>0.0115</td>
<td>0.0112</td>
<td>0.0034</td>
</tr>
<tr>
<td>( b_4 )</td>
<td>(0.0006)</td>
<td>(0.0011)</td>
<td>(0.0012)</td>
<td>(0.0020)</td>
<td>(0.0020)</td>
<td>(0.0020)</td>
<td>(0.0020)</td>
<td>(0.0006)</td>
</tr>
</tbody>
</table>

Log-like: \(-25,784.4 \quad -23,700.7 \quad -25,768.0 \quad -23,680.5 \quad -23,678.3 \quad -23,678.8 \quad -23,705.0\)

AIC: \(51,727.5 \quad 47,568.0 \quad 51,694.7 \quad 47,527.6 \quad 47,523.3 \quad 47,524.1 \quad 47,576.7\)

BIC: \(51,727.5 \quad 47,568.0 \quad 51,694.7 \quad 47,527.6 \quad 47,523.3 \quad 47,524.1 \quad 47,576.7\)

**NOTE:** The table contains the estimated parameters and their standard errors for nine alternative models, including the GARCH/cDCC. The prefix “c” 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 “l,” whereas the logarithmic specification for the variances is denoted by the additional suffix “h.” 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 \((v^{-1} = 0)\) density. The first four parameters, \( a_1 \) and \( b_1 \), 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.
The forecasting experiment covers the last 2500 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 $y_{t+1}^1, \ldots, y_{t+5}^5$ 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 $p_{j,t+k}^s = g_j^s y_{j,t+k}^s$ for $j = 1, \ldots, 6$ and $k = 1, \ldots, 5$. We use the sample of 10,000 simulated paths for portfolio returns $p_{j,t+1}^1, \ldots, p_{j,t+5}^5$ 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 $j$th portfolio at time $t$ and coverage level $\alpha$ is denoted by $\text{VaR}_j(t, \alpha)$. To evaluate the forecasting performance of the models, we require the indicator function

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

VaR$_j(t, \alpha)$ at time $t$. Note where the VaR$_j(t, \alpha)$ is estimated by the quantile of the 10,000 simulated paths for $p_{j,t+1}^s$ and $p_{j,t+5}^s$, as described earlier. The overall number of exceedances for the $j$th portfolio in our forecasting sample is measured as $\gamma_j = \sum_{t=1}^{T} I_{jt}$, with the average number of VaR exceedances given by $\bar{\gamma}_j = \gamma_j/T$. The forecast coverage $\tilde{\alpha}_j$ should be equal to the unconditional coverage $\alpha$. 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 (1995) is given by

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

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 exceedances $\alpha_j$ 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-spectrum for the variances is important for

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.
out-of-sample analyses. The increase in $p$-values over those of the regular $t$-GAS is often substantial. As before, the hyper- 

sphere 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 $z \sim N(0, aI)$ for some scalar $a > 0$, then the fourth (cross-)moments are given by

$$E[z_i z_j z_m z_n] = a^2(\delta_{ij}\delta_{lm} + \delta_{il}\delta_{jm} + \delta_{im}\delta_{jl})$$

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

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

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

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

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

for $i, j, l, m = 1, \ldots, k$.

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

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

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

$$\frac{\partial \log p(y_l | \Sigma_t; \nu)}{\partial \text{vech}(\Sigma_t)} = \frac{1}{2} D_{\chi}^i(\Sigma_t^{-1} \otimes \Sigma_t^{-1})$$

$$\times \left[ \frac{(v + k)}{(v + y_i^\prime \Sigma_t^{-1} y_i - 2)} y_i \otimes y_i - \text{vec}(\Sigma_t) \right]$$

$$= \frac{1}{2} D_{\chi}^i(\chi_{\gamma} - \text{vec}(I)), \quad (A.3)$$

where $\gamma_i = J\gamma_i$ follows a Student $t$ distribution with mean 0, covariance matrix $I_k$, 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 \( \Psi_t \).

To obtain the expression for \( \mathbb{E}[\nabla_t \nabla_t'] \), we need to show that
\[
\mathbb{E}\left[(w_t \tilde{y}_t - \text{vec}(I))(w_t \tilde{y}_t - \text{vec}(I))' \right] = gG - \text{vec}(I) \text{vec}(I)'.
\]
Because we take expectations with respect to the Student \( t \) density \( p(y_t | \Sigma_t; \nu) \), we have
\[
\mathbb{E}[\nabla_t \tilde{y}_t \otimes \tilde{y}_t] = \left( \begin{array}{c} (\nu + k) \\ (\nu - 2) \end{array} \right) \mathbb{E}\left[ \frac{\Gamma((\nu + k)/2)}{\Gamma(\nu/2)(\nu - 2)^{\pi k/2}} \times \bar{y}_t \otimes \bar{y}_t \right] = \left( \begin{array}{c} (\nu + k) \\ (\nu - 2) \end{array} \right) \mathbb{E}[\tilde{y}_t \otimes \tilde{y}_t],
\]
where \( \tilde{y}_t \) has a Student \( t \) distribution with mean 0, covariance matrix \( (\nu - 2)/\nu \) (\( \nu + 2 \)), and \( \nu + 4 \) degrees of freedom. In addition, because \( \tilde{x}_t = x_t / \sqrt{\nu/4} \) with \( x_t \) a mean 0 normal with covariance matrix \( (\nu - 2)/\nu (\nu + 4) \), and \( u \) independent \( \chi^2 \) \( \nu + 4 \) random variable, we use Results 1 and 2 of this Appendix to express (A.4) as
\[
\left( \begin{array}{c} (\nu + k) \\ (\nu - 2) \end{array} \right) \mathbb{E}[\tilde{y}_t \otimes \tilde{y}_t] = \left( \begin{array}{c} (\nu + k) \\ (\nu - 2) \end{array} \right) \mathbb{E}\left[ \frac{\Gamma((\nu + k)/2)}{\Gamma(\nu/2)} \times \frac{\Gamma((\nu + 4)/2)}{(\nu + 4)^2} \cdot G \right] = \frac{\nu + k}{\nu + 2} G.
\]
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

Haryadi, A. C., and Charkravarty, T. (2008), "Beta--(E)GARCH," unpublished manuscript, Faculty of Economics, Cambridge University. [554]