A New Model for Dynamic Correlations under Skewness and Fat Tails

Xin Zhang [*] Drew Creal[†] Siem Jan Koopman[‡] André Lucas[§]

March 30, 2011

Abstract

We propose a new model for dynamic volatilities and correlations under skewed, heavy-tailed distributions. The model combines the class of Generalized Hyperbolic (GH) distributions with an observation driven model that is excited by the scaled density score. The key novelty in our approach is that the skewed and fat-tailed shape of the distribution directly affects the dynamic behavior of the time-varying parameters. This distinguishes our approach from familiar alternatives such as the DCC model with GH disturbances. Using simulated and empirical evidence, we show that the model outperforms its close competitors if skewness and kurtosis are relevant features of the data.

Keywords: Dynamic correlation models, Generalized Hyperbolic distributions, Observation driven models.

JEL Code: C10, C16, C22, C32.

[*]Department of Finance, VU University Amsterdam and Tinbergen Institute.
[†]Booth School of Business, University of Chicago.
[‡]Department of Econometrics, VU University Amsterdam and Tinbergen Institute.
[§]Department of Finance, VU University Amsterdam, Duisenberg school of finance and Tinbergen Institute.

1 Introduction

Many economic time series exhibit excess kurtosis, negative skewness, and time-varying volatilities and correlations, see for example McNeil et al. (2005), Eberlein and Keller (1995), Franses and Van Dijk (2000), and Engle (2002). In this paper, we present a new dynamic observation driven model for conditional correlations and volatilities based on multivariate Generalized Hyperbolic (GH) distributions and dynamics that are driven by scaled local density scores. The Generalized Hyperbolic (GH) distribution was introduced by Barndorff-Nielsen (1977) and further explored in Barndorff-Nielsen (1978) and Blæsild (1981). The distinguishing feature of our approach is that the fat-tailed and skewed nature of the error distribution not only affects the likelihood, but also the dynamics of correlations and volatilities. This differentiates our approach from other well-known models in the literature, such as the combination of the a Student's t distribution with the Dynamic Conditional Correlation (DCC) model of Engle (2002).

The literature on time-varying parameter models for volatilities and correlations is vast. It includes multivariate generalizations of the stochastic volatility model, see the overview in Shephard (2005), or the multivariate GARCH class of models, see for example the VEC model of Engle and Kroner (1995), the BEKK model by Engle and Kroner (1995), the Orthogonal GARCH by Alexander (1998) and Alexander (2001), and the Generalized Orthogonal GARCH by van der Weide (2002) and Boswijk and Van der Weide (2006). To reduce the number of parameters to be estimated, Bollerslev (1990) introduced the CCC model, which has GARCH volatility dynamics, but constant conditional correlations. Engle (2002) generalized the CCC model to a Dynamic Conditional Correlation (DCC) model, making the conditional correlation matrix time-varying while at the same time retaining parsimony.

Most of the above models were originally derived under the parametric assumption of a normal distribution, but later generalized using alternative distributions to accommodate the stylized facts of economic data. Distributions used include the Student's t, the skewed Student's t, and the Generalized Hyperbolic distribution, see for example Bauwens and Laurent (2005), Fiorentini et al. (2003), Hu (2005), Sentana (2004), Peters (2001) and Prause (1999).

The remaining puzzle emerging from this previous literature is that the form of the distribution has no impact on the specification of the volatility and correlation dynamics. If the

distribution is for example leptokurtic, one might expect that a large absolute innovation is not necessarily attributable to a recent increase in variance. Equally likely, the large innovation may be a mere reflection of the fat-tailed nature of the data generating process. Similarly, one would expect the variance dynamics to react differently to a large positive realization compared to a negative one if the data is drawn from a skewed distribution.

Our main contribution in this paper is to provide a model for time-varying variances and correlations in which the precise form of the error distribution directly governs the specification of volatility and correlation dynamics. To do so, we extend the framework of Creal et al. (2008) and Creal et al. (2011) to a multivariate setting with both skewness and kurtosis. Creal, Koopman, and Lucas (2011) use the same framework for the special case when the data is drawn from a multivariate Student's t distribution, see also the univariate models in Harvey and Chakravarty (2008) and Nelson and Foster (1994)). Similar to their approach, our current model provides an automatic mechanism which limits the impact of 'outlying' or large observations on future correlations and volatilities. At an intuitive level, the new model attributes part of the large observation to the fat-tailed nature of the data generating process rather than to recent increases in volatilities or correlations.

Our results provide a full generalization of Creal et al. (2011) to the multivariate skewed setting. We show that the volatility and correlation updating mechanism now includes a natural quasi-leverage effect to allow for a different impact of negative versus positive realizations on volatility and correlation dynamics. For example, if the distribution is left-skewed, large negative realizations are more likely and should not automatrically be attributed to local volatility increases. A large positive realization for a right-skewed distribution, on the other hand, is extremely unlikely unless volatilities and correlations have increased recently. The impact of positive versus negative observations on the volatility dynamics should, therefore, also be different if the true distribution is skewed.

Estimation of our new model is straightforward. Similar to multivariate GARCH models, the likelihood has an analytically tractable form and can easily be computed using a standard prediction error decomposition. However, there is a tendency in the literature to compute the maximum likelihood estimator of the GH distribution using Expectation-Maximization (EM) algorithm of Dempster, Laird, and Rubin (1977). Estimating the new dynamic model using EM is not straightforward due to the highly non-linear functions of the data that are used to drive

the volatility and correlation dynamics. We show how to modify the standard EM algorithm in this particular setting to make estimation by EM feasible. The key step is to replace the score as a driving mechanism by a conditional score that runs in parallel to the conditional expectations taken in the expectations step of the EM algorithm.

In a simulation experiment, we compare the performance of our new model with some direct competitors, including the popular DCC model. We do so in a simulation setting with different correlation and volatility dynamics and a variety of error distributions. In doing so, we also estimate the DCC model using GH distributions. Alhough this is not the primary focus of our paper, this has to the best of our knowledge not been done before in the literature. If the true error distribution is normal, differences in performance between the different statistical models are limited. For fat-tailed error distributions, however, our new model based on the GH distribution has superior performance. In addition, if the errors process is clearly skewed, our new model performs best.

We provide an empirical illustration of the new model to investigate the volatilities and correlations between four blue-chip stocks from different industries. The sample period includes the recent financial crisis. We find that the estimated correlation dynamics differ substantially between our new approach and a traditional DCC analysis. The new approach seems much less influenced by incidental influential observations. Accounting for the skewness and fat-tailed nature of the data, we show that volatilities for all series are smaller in magnitudes compared to the DCC estimates. Also, the estimated persistince of volatilities and correlations is higher in general for the new approach.

The paper is set up as follows. Section 2 introduces the model. Section 3 discusses various alternatives for model parameterization. Section 4 develops a model for the scale rather than the covariance matrix and proposes a modified EM algorithm for estimation. Section 5 provides Monte Carlo evidence on the performance of the new model compared to some of its competitors from the literature. Section 6 presents our empirical illustration. Section 7 concludes.

2 The dynamic GH model

We assume our data generating process is given by

$$y_t = L_t \varepsilon_t, \qquad \Sigma_t = L_t L_t', \tag{1}$$

where $y_t, \varepsilon_t \in \mathbb{R}^k$ for $t = 1, \ldots, n$, L_t is a $k \times k$ lower triangular matrix giving rise to a time-varying $k \times k$ covariance matrix Σ_t, and ε_t follows a Generalized Hyperbolic (GH) distribution with zero mean and unit covariance matrix. The specification (1) can easily be extended to include a conditional or unconditional non-zero and possibly time-varying mean for y_t. In line with Engle (2002), we also decompose the covariance matrix Σ_t as

$$\Sigma_t = L_t L_t' = D_t R_t D_t, \tag{2}$$

with D_t a diagonal matrix containing the standard deviations of the elements in y_t, and R_t the correlation matrix of y_t.

The Generalized Hyperbolic (GH) distribution as introduced by Barndorff-Nielsen (1977) is a flexible distribution that accommodates both thin and fat-tailed as well as positively and negatively-skewed distributions. This is most easily seen by constructing the GH class as a normal mean-variance mixture,

$$\varepsilon_t = \mu_\varepsilon + \zeta_t T\gamma + \sqrt{\zeta_t} T z_t, \tag{3}$$

with $z_t \sim \mathrm{N}(0, \mathrm{I}_k)$, and $\zeta_t \in \mathbb{R}^+$ a positively valued random scalar that is independent of z_t. The three parameters in this mixture construction have a natural interpretation as the location parameter $\mu_\varepsilon \in \mathbb{R}^k$, the $k \times k$ scaling matrix TT', and the skewness parameter $\gamma \in \mathbb{R}^k$. The GH class includes distributions such as the normal ($\gamma = 0$ and $\zeta_t \equiv 1$), the (skewed) multivariate Student's t (ζ_t has an inverse Gamma distribution, and $\gamma = 0$ for the symmetric case or $\gamma \neq 0$ for the asymmetric case), the (skewed) variance-gamma distribution (ζ_t has a Gamma distribution), the Generalized Hyperbolic distribution (ζ_t has a Generalized Inverse Gaussian (GIG) distribution with parameters λ, χ, and ψ), as well as the Gamma, inverse Gamma, and GIG distribution ($L_t = 0$).

As we assume ε_t to have zero mean and unit covariance matrix, we obtain from (3) that

$$0 = \mathrm{E}[\varepsilon_t] = \mu_\varepsilon + \mu_\zeta T\gamma \quad \Leftrightarrow \quad \mu_\varepsilon = -\mu_\zeta T\gamma, \tag{4}$$

and

$$\mathrm{I}_k = \mathrm{E}[\varepsilon_t \varepsilon_t'] = T\left(\mu_\zeta \mathrm{I} + \sigma_\zeta^2 \gamma\gamma'\right) T' \quad \Leftrightarrow \quad (T'T)^{-1} = \mu_\zeta \mathrm{I} + \sigma_\zeta^2 \gamma\gamma', \tag{5}$$

where μ_ζ and σ_ζ^2 denote the mean and variance of ζ_t. Note that the mean and variance, respectively, of ζ_t need to exist for the mean and variance of ε_t to exist. The density of y_t for this parameterization is spelled out in detail in the appendix.

To make the variances and correlations time-varying, we assume that both D_t and R_t depend on a time-varying parameter f_t, such that $D_t = D(f_t)$ and $R_t = R(f_t)$. In this way, we allow the correlations and standard deviations to have their own dynamics or to be driven by a smaller set of time-varying factors such as in the factor GARCH literature.

We model the dynamics of f_t using the framework of Creal, Koopman, and Lucas (2008, 2011). The transition equation of the time-varying factor f_t is then given by

$$f_{t+1} = \sum_{i=0}^{p-1} A_i s_{t-i} + \sum_{i=0}^{q-1} B_i f_{t-j}, \tag{6}$$

where A_i and B_i are appropriately sized matrices that depend on a static parameter vector θ, $A_i = A_i(\theta)$ and $B_i = B_i(\theta)$. The key component in (6) is s_t. It is chosen as an appropriate function of current and past values of y_t and f_t. For example, in the univariate case with $f_t = D_t^2$ and normally distributed y_t, the model in (6) embeds the standard GARCH model by setting $s_t = y_t^2$. Though such an obvious choice for s_t is sometimes possible if the model is sufficiently simple, in general it is much harder to formulate a natural candidate for s_t in a complex model. For example, in our current setting we want (6) to account for the possibly fat-tailed and skewed nature of the GH distribution, as well as for the adopted parameterization $D(f_t)$ and $R(f_t)$.

Following Creal et al. (2011) we use the scaled density score as the driver s_t,

$$s_t = S_t \nabla_t, \tag{7}$$

$$\nabla_t = \partial \ln p(y_t | \mathcal{F}_{t-1}; f_t, \theta) / \partial f_t, \tag{8}$$

where S_t is an $\mathcal{F}_{t-1}$-adapted scaling matrix, and $\mathcal{F}_t = \{y_t, \ldots, y_1\}$. For our standard GH distributed ε_t, we obtain the following result.

Result 1. *If ε_t in (1) has a GH distribution with zero mean and unit covariance matrix, we have*

$$\nabla_t = \Psi_t' H_t' \left(w_t (y_t \otimes y_t) - \mathrm{vec}(\tilde{\Sigma}_t) - (1 - w_t \mu_\zeta)(y_t \otimes \tilde{L}_t \gamma) \right), \tag{9}$$

where $\otimes$ denotes the Kronecker product, $\Psi_t = \partial \mathrm{vech}(\Sigma_t) / \partial f_t'$, $\tilde{L}_t = L_t T$, $\tilde{\Sigma}_t = \tilde{L}_t \tilde{L}_t'$, w_t is a scalar weight, and H_t is an $k \times k$ matrix. Both w_t, H_t, and Ψ_t are defined in detail in the appendix,

Our current model generalizes some of the well-known univariate and multivariate GARCH models. This is easily seen by considering the special case of the Gaussian. For the Gaussian distribution, $\gamma = 0$, $T = \mathrm{I}_k$, and $w_t = \mu_\zeta = 1$. Equation (9) then collapses to $\Psi_t' H_t' \mathrm{vec}(y_t y_t' - \Sigma_t)$, which is the natural driver for a multivariate model for time-varying volatilities and correlations. The matrix H_t is an intermediate result capturing the relation between $\tilde{L}_t$ and Σ_t. The matrix Ψ_t is determined by the parameterization of $D(f_t)$ and $R(f_t)$ on the dynamics of the time-varying parameter vector f_t.

There are two interesting differences between standard multivariate GARCH models driven by squares and cross-products of y_{t-1}, and the model driven by the score (9) of the GH distribution. The first difference is the presence of the term w_t in (9) that weights the lagged squared observations. The second difference is the presence of the leverage term $\gamma \otimes y_t$. These two effects are due to the fat-tailedness and skewness of the distribution of y_t. For the case of a symmetric Student's t distribution, Creal et al. (2011) also obtained a weight effect, though not a leverage effect.

As can be seen in the appendix, the weight w_t is generally a decreasing function of

$$d_{x_t}^{\chi} = \chi + x_t' x_t, \qquad x_t = \tilde{L}_t^{-1} y_t + \mu_\zeta \gamma,$$

for fat-tailed distributions in the GH class, where x_t is the usual standardized version of y_t. As a result, the impact of lagged (cross)-products in $y_t \otimes y_t$ on future values of f_t (and thus on volatilities and correlations) is mitigated by w_t if y_t is large in the sense that $d_{x_t}^{\chi}$ is large. The intuition for this is as follows. If y_t is drawn from a fat-tailed distribution, observing large values of y_t need not be due to local volatility or correlation increases. Instead, large y_ts may be due to the mere fat-tailed nature of the distribution. As a result, the dynamics of f_t (and volatilities and correlations) should only partly reflect the large value of y_t. The remainder should be ascribed to the fat-tailed nature of the distribution and should therefore not affect the volatility and correlation dynamics.

The second difference in (9) is the leverage term. Note that this term takes a different role than the usual leverage effect in volatility models. The latter captures the effect that volatilities tend to increase if returns are negative. Such a standard leverage effect could be included in (9) as well in the usual way. The current leverage effect, however, is purely due to the skewness of the distribution. If, for example, y_t is univariate and right-skewed ($\gamma > 0$), a large positive

value of y_t is more likely and, therefore, not automatically attributable to a local volatility increase. Similarly, a large negative value of y_t should be taken as a very strong signal of a volatility increase, because large negative observations are extremely unlikely for a right-skewed distribution, unless the volatility has increased. This is precisely the role taken by the leverage term in (9). As a result, the impact of a standard leverage effect would be estimated more strongly if we account for the skewed nature of y_t in the volatility and correlation dynamics.

The fact that both the shape of the distribution and the parameterization affect the dynamics of f_t differentiate the current model from standard GARCH-type models with non-Gaussian observations. For the latter class of models, the non-Gaussian assumption only affects the likelihood, but not the dynamic behavior of f_t. In our setting, the distributional properties of y_t affect both the likelihood and the f_t-dynamics.

Because of their resemblance to GARCH type models in the sense that s_t is $\mathcal{F}_t$-adapted, estimation of our model in (7)–(9) is as easy (or difficult) as that of GARCH models. In particular, the likelihood can be expressed in closed-form via the recursion (6) and a standard prediction error decomposition. This allows for fast likelihood evaluation. The interpretation of the model is also intuitive. Depending on the choice of the scaling matrix S_t, the step s_t can for example be interpreted as a local Gauss-Newton or Steepest-Ascent improvement to the likelihood at time t. The score of the observation density at time t, evaluated at the current estimate f_t of the time-varying parameter, determines in what direction f_t is best updated to improve the fit of the model. The additional lags and dynamics in (6) add further flexibility to the size and speed of these adjustments as time progresses.

We gather all static parameters of the model, such as γ, μ_ζ, σ_ζ^2, $A_1, \ldots, A_p$, $B_1, \ldots, B_q$, etc., into the static parameter vector θ. We then estimate θ using standard Maximum Likelihood (ML). Inference on θ is carried out in the standard way using the negative inverse Hessian of the log likelihood as the covariance matrix of the ML estimator.

3 Model parameterizations

The GH distribution has a considerable number of parameters. Not all of these are identified simultaneously. In particular, χ and ψ are not separately identified. Rather, we can only identify $\chi\psi$. There are several ways to ensure identification. One is to set $|\Sigma_t|$ to a fixed

constant, such as one. Alternatively, one can fix χ or ψ and estimate the other parameter in an unrestricted way. In our current set-up, we estimate $\kappa = (\chi\psi)^{1/2}$ and extract χ and ψ separately through the identifying assumption $\mu_\zeta = 1$. The latter identifying restriction turns out to be particularly useful when estimating the GH model using the Expectation Maximization (EM) algorithm, see Section 4. From the definition of the GH distribution, we obtain for $\mu_\zeta = 1$ and a fixed value of κ that

$$1 = \mu_\zeta = \frac{\sqrt{\chi\psi} K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{\psi K_\lambda\left(\sqrt{\chi\psi}\right)} \quad \Leftrightarrow \psi = \frac{\kappa \cdot K_{\lambda+1}(\kappa)}{K_\lambda(\kappa)}, \quad \chi = \kappa^2/\psi. \tag{10}$$

Following Creal et al. (2011), there are at least two obvious choices for the parameterization of both the diagonal matrix of variances (D_t^2) and the correlation matrix (R_t) in (2). For the variances, one can either model the variances $\text{diag}(D_t^2)$ directly, or opt for modeling the log-variances $\ln(\text{diag}(D_t^2))$. The latter parameterization has the advantage that the variances themselves always remain positive. If one adopts the former parameterization, parameter restrictions have to be imposed on the model's parameters A_i and B_i to ensure positivity of the variances. This quickly becomes complicated for models in higher dimensional settings that include more lags in the transition equation (6). We therefore restrict our attention to the parameterization in terms of log-variances.

To parameterize the correlation matrix, we have to satisfy two constraints. First, R_t has to be positive definite. Second, the diagonal elements of R_t have to be equal to one for all t. The first possible parameterization of the correlation matrix R_t is inspired by the DCC model of Engle (2002). Let $Q_t = Q(f_t)$ be an auxiliary time varying parameter matrix, and set

$$R_t = \Delta_t^{-1} Q_t \Delta_t^{-1}, \tag{11}$$

where Δ_t^2 is a diagonal matrix holding the diagonal elements of Q_t. The matrix Q_t has k redundant elements compared to the correlation matrix R_t. As a result, only $k(k-1)/2$ independent signals in ∇_t are distributed over the k^2 elements of Q_t. The details on the parameterization and its resulting specification of Ψ_t are presented in Creal et al. (2011).

An alternative parameterization for the correlation matrix is given by the hypersphere parameterization as used for example by Jaeckel and Rebonato (2000), van der Weide (2002), and Creal et al. (2011). Here, we decompose the correlation matrix into its Choleski decomposition $R_t = X_t X_t'$ for a lower triangular $k \times k$ matrix X_t. The matrix X_t is parameterized by a set of

$k(k-1)/2$ time-varying angles ϕ_{ijt}, such that

$$X_t = \begin{pmatrix} 1 & c_{12t} & c_{13t} & \cdots & c_{1kt} \\ 0 & s_{12t} & c_{23t}s_{13t} & \cdots & c_{2kt}s_{1kt} \\ 0 & 0 & s_{23t}s_{13t} & \cdots & c_{3kt}s_{2kt}s_{1kt} \\ 0 & 0 & 0 & \cdots & c_{4kt}s_{3kt}s_{2kt}s_{1kt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & c_{k-1,kt}\prod_{\ell=1}^{k-2} s_{\ell kt} \\ 0 & 0 & 0 & \cdots & \prod_{\ell=1}^{k-1} s_{\ell kt} \end{pmatrix}, \tag{12}$$

where $c_{ijt} = \cos(\phi_{ijt})$, $s_{ijt} = \sin(\phi_{ijt})$, and ϕ_{ijt} a set of angles in $[0, \pi]$. For the 2-dimensional case, this reduces to

$$X_t = \begin{pmatrix} 1 & \cos(\phi_{12,t}) \\ 0 & \sin(\phi_{12,t}) \end{pmatrix}, \qquad R_t = X_t'X_t = \begin{pmatrix} 1 & \cos(\phi_{12,t}) \\ \cos(\phi_{12,t}) & 1 \end{pmatrix}, \tag{13}$$

such that the correlation ρ_t is given by $\cos(\phi_{12,t})$. The second column of X_t in (13) expresses a two-dimensional unit-length vector in terms of its polar rather than its Cartesian coordinates. The generalization of this to the k-dimensional setting is given by the kth column of X_t in (12).

The number of free parameters in X_t equals the number of correlations in the matrix R_t such that there are no redundancies as in parameterization (11). We gather all the angles ϕ_{ijt} in the vector ϕ_t which is part of f_t. Irrespective of the exact value of ϕ_t, the matrix $X_t'X_t$ now always satisfies the properties of a correlation matrix. Again, the resulting specification for the matrix Ψ_t are provided in Creal et al. (2011).

To complete the definition of the step s_t in (6), we need to choose a scaling matrix S_t. Creal et al. (2008) discuss a number of possible choices, all of which are based on the local curvature of the model density at time t via the (local) information matrix. Computing the information matrix for the general GH distribution, however, is analytically not tractable. Therefore, we consider a computationally feasible alternative and set the scaling matrix equal to the inverse information matrix for a special case of the GH class, namely the symmetric Student's t distribution. This information matrix is known analytically as derived in Creal et al. (2011). This choice accommodates both the possible fat-tailed nature of the distribution and the time-variation in the volatilities and correlations. This form of scaling can be implemented

efficiently and works well for both simulated and empirical data, as is shown in the next sections. Moreover, it makes our current model directly comparable to the familiar multivariate GARCH models if the distribution is Gaussian.

An interesting final feature of our model is that one can easily impose a factor structure on the volatilities and correlations. This can be done by picking the dimension of f_t to be lower than the number of elements in $\text{diag}(D_t)$ and Q_t or X_t. This approach can be used if the same factors drive more correlations, or if correlations and volatilities are driven by the same factors. The model easily allows for this and the dynamics adapt automatically via a respecification of Ψ_t in (9). Via the density score our framework naturally weights and combines the different sources of information in y_t to improve the current estimates of volatilities and correlations by changing f_t.

4 Time-varying scale matrix and an EM algorithm

To this point, we have modeled the time varying covariance matrix $\Sigma_t = L_t L_t'$ directly via the dynamics for f_t in (6) and (9). There are two issues with this approach. First, by concentrating on the covariance matrix, we need to assume that the variances exist. This puts a constraint on the fat-tailedness of the mixing variable ζ_t. For example, in the case of a skewed Student's t distribution, it requires the degrees of freedom parameter to be higher than 4 (rather than the usual 2 for the symmetric case). This constraint may be problematic for financial data with heavier tails. The problem can be solved if one is willing to model time variation in the scaling matrix $\tilde{\Sigma}_t$ rather than in the covariance matrix Σ_t. The moment restrictions can then be fully relaxed, as the scaling matrix $\tilde{\Sigma}_t$ always exists.

A second well-known issue for the GH distribution is that the distribution has many parameters, which is claimed to sometimes complicate estimation particularly for higher dimensional y_t, see Hu (2005). Most authors compute the ML estimator using the Expectation-Maximization (EM) algorithm of Dempster, Laird, and Rubin (1977). A basic introduction of the EM algorithm for the GH distribution with time-invariant covariance matrix is provided in McNeil et al. (2005). Extensions to the multivariate GARCH model with GH distributions are found in for example (Hu 2005). The key advantage is that the estimation of the parameters of the mixing distribution can then be split from the parameters of the scaling matrix (and its dynamics).

This advantage breaks down for the specification of our model based on the covariance matrix Σ_t. However, if we make the scale matrix $\tilde{\Sigma}_t$ rather than the covariance matrix Σ_t our object of interest, we can develop a new, modified EM algorithm that recovers the usual advantages of EM estimation for GH distributions.

For both of the above reasons, we formulate a related model in this section in terms of the scaling matrix $\tilde{\Sigma}_t = \tilde{L}_t \tilde{L}_t'$ and propose a modified EM algorithm as an alternative method of estimating the model's static parameters. Using the root of the scaling matrix $\tilde{L}_t$, the mean-variance normal mixture construction for the observations y_t becomes

$$y_t = \mu_{ty} + \zeta_t \tilde{L}_t \gamma + \sqrt{\zeta_t} \tilde{L}_t z_t, \tag{14}$$

with $\mu_{ty} = -\mu_\zeta \tilde{L}_t \gamma$, compare (3). As $\tilde{\Sigma}_t = \tilde{L}_t \tilde{L}_t'$ is a covariance matrix as well, but then for the normal in the mixture construction, we can use the same machinery as in previous sections and define $\tilde{\Sigma}_t = \tilde{D}_t \tilde{R}_t \tilde{D}_t$ and $\tilde{\Psi}_t = \partial \text{vech}(\tilde{\Sigma}_t)/\partial f_t'$, accordingly.

In the usual implementation of the EM algorithm for the GH distribution, one can split the estimation of the parameters of the mixing variable's distribution from that of the other parameters, see McNeil et al. (2005). The main difficulty with this standard implementation in our current context lies in the complex dynamics of f_t via the score element in s_t. As the score of the GH distribution also depends on the parameters of the mixing variable, it seems impossible at first sight to obtain a split in the parameter vector and reduce a high-dimensional likelihood optimization into two lower dimensional problems. Our modification of the EM algorithm, however, circumvents this problem and is based on the following observation.

Result 2.

$$\nabla_t = \frac{\partial \ln p(y_t | f_t, \mathcal{F}_{t-1}; \theta)}{\partial f_t} = \mathrm{E}\left[\left. \frac{\partial \ln p(y_t | \zeta_t, f_t, \mathcal{F}_{t-1}; \theta)}{\partial f_t} \right| \mathcal{F}_t \right]. \tag{15}$$

This result will be used below to split the high-dimensional optimization problem into smaller pieces.

Before we comment on the precise modifications and their interpretation, we first present the complete algorithm. We partition the vector θ of static parameters as $\theta' = (\theta_1', \theta_2')$, with θ_2 denoting the parameters of the distribution of the mixing variable ζ_t, i.e., λ, χ, and ψ, while the remaining parameters are collected in θ_1. We define the log joint likelihood of the observation

y_t and the unobserved mixing variable ζ_t as

$$\sum_{t=1}^{n} \ln p(y_t, \zeta_t | f_t, \mathcal{F}_{t-1}; \theta) = \mathcal{L}_{1n}(\theta) + \mathcal{L}_{2n}(\theta_2), \tag{16}$$

with

$$\mathcal{L}_{1n}(\theta) = \sum_{t=1}^{n} \ln p(y_t | \zeta_t, f_t, \mathcal{F}_{t-1}; \theta) \tag{17}$$

and

$$\mathcal{L}_{2n}(\theta_2) = \ln p(\zeta_t; \theta_2), \tag{18}$$

where $p(y_t|\zeta_t, \mathcal{F}_{t-1}; \theta_1)$ is (conditionally) Gaussian, and $p(\zeta_t; \theta_2)$ is Generalized Inverse Gaussian (GIG), also denoted as $\mathrm{N}^{-}(\lambda, \chi, \psi)$. For the E-step in the EM algorithm, we define

$$Q_1(\theta, \hat{\theta}) = \int \ldots \int \mathcal{L}_{1n}(\theta) \left(\prod_{t=1}^{n} p(\zeta_t | y_t, \mathcal{F}_{t-1}; \hat{\theta}) \right) \mathrm{d}\zeta_n \ldots \mathrm{d}\zeta_1 = \mathrm{E}_{\hat{\theta}} \left[\mathcal{L}_{1n}(\theta) | \, \mathcal{F}_n \right], \tag{19}$$

and similarly

$$Q_2(\theta_2, \hat{\theta}) = \mathrm{E}_{\hat{\theta}} \left[\mathcal{L}_{2n}(\theta_2) | \, \mathcal{F}_n \right]. \tag{20}$$

In the appendix we show that under the normalization constraint $\mu_\zeta = 1$, Q_1 only depends on θ_1 and not on θ_2. Consequently, we write $Q_1(\theta_1, \hat{\theta})$ with a slight abuse of notation. The algorithm proceeds as follows.

Modified EM algorithm for a the dynamic GH model for the scale matrix

1. Start with an initial guess of the parameters, $\hat{\theta}^{(0)}$.

2. Given a trial value of the parameters $\hat{\theta}^{(\ell)}$, define the modified transition equation for the scaling matrix as

$$f_{t+1} = \sum_{i=0}^{p-1} A_i \tilde{s}_{t-i} + \sum_{i=0}^{q-1} B_i f_{t-j}, \tag{21}$$

   where $\tilde{s}_t = S_t \tilde{\nabla}_t$, and

$$\tilde{\nabla}_t^{(\ell)} = \mathrm{E}_{\hat{\theta}^{(\ell)}} \left[\partial p(y_t | \zeta_t, f_t, \mathcal{F}_{t-1}; \theta) / \partial f_t \mid \mathcal{F}_t \right], \tag{22}$$

   with $\tilde{\nabla}_t^{(\ell)}$ fully specified in the appendix.

3. Given the modified dynamics, compute $Q_1(\theta_1, \hat{\theta}^{(\ell)})$ and maximize it numerically with respect to θ_1. The maximum is obtained at $\tilde{\theta}_1$.

4. Update $\hat{\theta}^{(\ell)}$ to $\tilde{\theta}^{(\ell)} = (\tilde{\theta}_1', (\hat{\theta}_2^{(\ell)})')$, compute $Q_2(\theta_2, \tilde{\theta}^{(\ell)})$ and maximize it numerically with respect to θ_2. The maximum is obtained at $\tilde{\theta}_2$.

5. Update $\hat{\theta}^{(\ell)}$ to $\hat{\theta}^{(\ell+1)} = (\tilde{\theta}_1', \tilde{\theta}_2')'$ and iterate steps 2–5 until convergence.

Steps 3–5 are standard for the GH-EM algorithm, see for example McNeil et al. (2005). The E-steps are worked out in detail in the appendix. An important feature is that the optimization can be split into two lower dimensional problems in steps 3 and 4, respectively. Step 3 boils down to estimating a standard multivariate Gaussian GARCH type model with a weighted transition equation via (21). The most important modification to the algorithm, however, is step 2. In this step, the transition equation that depends on θ_2 through the score ∇_t is replaced by a simpler equation that no longer depends on θ_2. The intuition follows from our earlier Result 2. As with the E-step in the EM algorithm, the score is replaced by a conditional expectation of a score based on the parameters from the previous iteration, $\hat{\theta}^{(\ell)}$. The score that enters this conditional expectation is much simpler than that of the full GH distribution, as it is the score of the conditional (on ζ_t) distribution of y_t, which is Gaussian. Clearly, as the EM iterations converge to the ML estimates, the score $\tilde{\nabla}_t^{(\ell)}$ in the EM algorithm converges to the score ∇_t of the full GH distribution via (15).

5 Simulation results

To study the behavior of the new model, we first use a controlled simulation environment. In the next section, we then investigate the model's performance for empirical data. In both settings, we benchmark the model's performance to the popular DCC model. The simulations test the accuracy of the different models in estimating volatility and correlation patterns, similar to the experiments in Engle (2002). We first describe the set-up in Subsection 5.1. Next, Subsection 5.2 presents the results.

5.1 Simulation set-up

The simulation set-up is divided into two parts. First, we use a univariate set-up to investigate the performance of different models for volatility estimation. Second, we run a bivariate simu-

lation experiment to test the adequacy of the models for correlation estimation while keeping volatilities constant.

In our first experiment, we use a univariate series with zero mean and time-varying volatility. The volatilities follow a deterministic function over time, much in line with the original experiments for correlations as described in Engle (2002). We take the same deterministic functions as in Engle's paper, namely

(1) Constant: $f_t = 0.9$,

(2) Sine: $f_t = 0.5 + 0.4\cos(2\pi t/200)$,

(3) Fast Sine: $f_t = 0.5 + 0.4\cos(2\pi t/20)$,

(4) Step: $f_t = 0.9 - 0.5(t > 500)$,

(5) Ramp: $f_t = \mod(t/200)$.

This allows us to study the properties of competing statistical models under a range of volatility dynamics, such as slow and fast oscillations, structural breaks, and so on.

To limit the number of parameters, we concentrate on a particular subclass of the GH family, namely the GH Skewed t (GHST) distribution. The GHST distribution sets $\lambda = -\nu/2$, $\chi = \nu - 2$ (due to the restriction $\mu_\zeta = 1$), and $\psi = 0$, such that four parameters remain: the degrees of freedom parameter ν, the location parameter μ, the variance Σ_t, and the skewness parameter γ. The DGP takes the form

$$y_t \sim GHST(0, D_t^2, \gamma, \nu), \quad D_t = f_t. \tag{23}$$

The second simulation experiment concentrates on recovering dynamic correlation patterns. In this experiment, we consider a bivariate series y_t with zero mean. The variances of the two elements in y_t are set to unity to fully concentrate on the correlations. In line with the experiments in Engle (2002), we use the same five deterministic patterns for the time-variation in correlations as described earlier. The DGP takes the form

$$y_t \sim GHST(0, D_t R_t D_t, \gamma, \nu), \quad D_t = \mathrm{I}_2, \quad R_t = \begin{pmatrix} 1 & f_t \\ f_t & 1 \end{pmatrix}. \tag{24}$$

Given the five different volatility or correlation patterns, we consider three different GHST distributions in our experiments. Note the GHST distribution contains the symmetric Student's t and the normal distribution as special cases. In particular, the GHST collapses to

the symmetric Student's t distribution if the skewness γ goes to zero. It further reduces to the normal distribution if the degree of freedom ν goes to infinity. As a benchmark, we start with the normal distribution. Then we introduce mild kurtosis by setting $\nu = 5$. Finally, we introduce moderate skewness by setting $\gamma = -0.03$ for the univariate, or $\gamma = (-0.03, -0.03)'$ for the bivariate simulations.

For each of the $5 \times 2 = 10$ DGPs in our experiments, we estimate two models. In the univariate simulations, we estimate a classical GARCH(1,1) model as a benchmark,

$$\sigma_t^2 = c + A_1 y_{t-1}^2 + B_1 \sigma_{t-1}^2, \tag{25}$$

with $A_1, B_1 > 0$ and $A_1 + B_1 < 1$. The estimated GARCH model uses the correct specification of the distribution, i.e., it assumes for example that y_t is Student's t distributed (with estimated degrees of freedom) if the DGP is the Student's $t(5)$ distribution. In this way, differences between benchmark models like the GARCH model and our own model are only attributable to the different dynamics for the latent dynamic variable f_t or σ_t^2. In the second experiment, we use the DCC model of Engle (2002) as our benchmark. Again, we assume that the correct distributional assumption is made for y_t. As an aside, to the best of our knowledge we have not seen any earlier applications of estimating the DCC model under the assumption of GHST or GH distributed error terms. The GARCH and DCC models are compared to the performance of our new model in the univariate and bivariate case, respectively. Our new model is estimated with diagonal 3×3 matrices A_1 and B_1 and always uses the GHST distribution for both likelihood evaluation and f_t-dynamics. For the bivariate setting, we use the hypersphere parameterization for the correlation parameter.

The performance of the different models can be measured in various ways. In our current experiments, we concentrate on a simple measure, namely the Mean Absolute Error (MAE) for the estimated volatility (or correlation). The MAE is measured as the total sum of absolute differences between the true and estimated variances (or correlations). We generate samples of size $T = 1,100$, discarding the first 100 observations to avoid dependence on initial conditions. The average MAE results are then computed over 100 Monte Carlo replications.

5.2 Results

Table 1 presents the results for the first experiment: the volatility simulations. For the Gaussian DGP, the GARCH model outperforms our new model in four out of five regimes. This is to be expected. Our new model is based on the GHST assumption and therefore has more parameters to estimate, which may cause some efficiency loss. For the t and GHST distributions, by contrast, our new model has smaller MAEs in four out of five regimes. This is in line with the intention of the new model. For fat-tailed and possibly skewed distributions, the dynamics of the GARCH model are more easily corrupted by incidental extreme observations that are due to the fat-tailedness of the error distribution rather than to the local increase of volatility. By exploiting the information in the shape of the error distribution, the new model automatically corrects for such effects, for example downweighting the effect of large values of y_t or of y_t values in the skewed tail area.

Looking at the different rows in Table 1, we see that if the volatility is constant, the new model performs slightly worse than the GARCH model. Again, this is to be expected as the new model uses more parameters, thus risking an efficiency loss. The comparison becomes more interesting if the DGP is the Student's t or GHST distribution and the volatility varies over time. In those cases, the MAEs of the new model is below that of the GARCH model, though the difference are modest. The average difference between the MAEs for the two models is around ten percent.

We now turn to the second experiment, which considers the adequacy of the different models in recovering dynamic correlation patterns. The results are presented in Table 2. As in the first experiment, the two models behave roughly similar if the true DGP has a constant correlation. The new models start to outperform the DCC models once we introduce non-normal distributions in the DGP.

If the data are generated using the Student's t or GHST distribution, the new model outperforms the DCC in terms of MAEs in four out of five cases. The differences are more pronounced than in the volatility case. As explained in the previous section, the main driver of this improved performance is the weighting function and leverage effect in the transition equation (6) and (9) for the factor f_t. As a result, incidentally large observations have less impact on distorting the correlation dynamics.

6 Empirical application

We now turn to an empirical illustration of the DGH-GHST model. We examine the correlation in a multivariate dataset with stock returns of four large companies in the US: Coca-Cola, IBM, Merck and J.P. Morgan. To estimate the volatilities and correlations, we use (i) the DCC model under Gaussian distribution, (ii) the DCC-GHST model, (iii) the three factor DGH(1,1)-GHST model and (iv) the three factor DGH(1,1)-GHST-G model with a GHST likelihood and a DGH recursion from Gaussian distributions. To retain comparability with the DCC models, we assume diagonal matrices A and B in the DGH model and correct for the ARCH in mean effect. In order to make the DGH model comparable, we also restrict correlations in the DGH model to load from one factor as in the DCC model.

We use daily log returns of the four stocks from January 1989 to December 2009 in the CRSP dataset. The dataset contains 5295 daily observations. Some descriptive statistics are provided in Table 3. It is clear that the series exhibit substantial excess kurtosis, warranting the use of a fat-tailed distribution. Moreover, both series exhibit non-zero skewness. The normality tests and skewness/kurtosis tests also provide strong statistical evidence for non-normality. Both features can be accommodated by the GHST distribution discussed in this paper.

The estimation results for the DGH(1,1) and DCC models are shown in Table 4. For the DGH model, almost all parameters are significant at the 5% level, except the skewness parameter for the Merck. The B_{ii} elements show that the correlation and volatility dynamics are highly persistent. The degrees of freedom estimate is around 6, which gives rise to the leptokurtic tails. The skewness parameter of the Merck is negative but statistically insignificant, while the skewness for the other three are positive but significant.

The estimated parameters for the DCC-G model are mostly significant. The parameters in B are more than 0.90, which shows the volatility and correlations are persistent over time. The estimated GARCH models in the DCC model all satisfy the stationarity conditions. The estimation result of the DCC-GHST model is quite interesting. In spite of many common conclusions with DGH-GHST model, the skewness parameters in the DCC model are mostly insignificant, except for the J.P. Morgan stock return. However, the signs of the skewness in the DCC model are the same as the DGH model.

We also include the DGH(1,1) model with GHST likelihood and the updating mechanism

from a Gaussian distribution. This DGH-GHST-G model is close to the DCC-GHST model. From the middle two columns in Table 3, we see that most of the parameters in this model agree with the DCC model estimates. In the following, we have several graphs to visualize the difference of the four models.

The estimated pairwise correlation dynamics are shown in Figure 1. Comparing the graphs, we see that the DGH-GHST estimates are the most stable over time. The correlations generally hover together for most of the period in all pairs. The DGH-GHST-G produces correlations which sometimes depart from the common pattern followed by the other three models. The results from the DCC-GHST and the DGH-GHST model are generally quite similar, apart in certain special periods.

In Figure 2 we plot a zoom in graph of the estimated correlations of different models for the sub-sample of 2002 to end of 2005. If we look at the figures, the DGH-GHST estimates are more persistent over time. The correlation from DCC-G model is the lowest among all models. The difference is more visible in the year 2005 when the DGH-GHST correlation is higher than the others. The DGH-GHST-G model even suggests negative correlation for the correlation between the other three and the Merck stock at early October 2004. It appears that the Merck company was experiencing a major worldwide withdrawal of its product. The Merck stock is shocked on the announcement publishing day. Surprisingly, the use of the GHST distribution in the DCC model does not help to remove these points attributed to this periods. Even after such point, the DCC models correlation are close to zero till the end of 2005. The DCC model with the GHST distribution cannot fully explore the information about skewness/kurtosis compared to the DGH model.

Further we plot the difference of the correlation estimates between the DCC-G, DGH-GHST and DGH-GHST-G models and DCC-GHST in Figure 3. The figure shows DCC-G and DCC-GHST models are quite similar to each other for all the correlation pairs. DCC-G correlations are always smaller than the DCC-GHST estimates. The bottom panel shows that DCC-GHST and DGH-GHST-G model are indeed close to each other. The correlation difference is hovering around zero, except for the correlation related to the Merck stock in the Merck withdrawal period. In such special period, the DGH-GHST-G produces extreme negative estimates of -0.2. We can see the difference of DGH-GHST and DCC-GHST is more interesting. The distance of the two estimates are generally larger than the other comparisons. If we look at the

correlation pairs including J.P. Morgan at the end of the sample for the Financial Crisis time, the correlation from DGH-GHST is larger than the DCC results. The DGH-GHST model can better recognize the correlation structure disturbed by the depression in financial industries.

We also plot the volatility estimates from the DCC and DGH models in Figure 4. All models indicate that the J.P. Morgan stock return is more volatile than the others. However, the variations in magnitude over time differ substantially among the models. The DCC-G volatilities prohibit similar characteristics as the DCC-GHST volatilities. The DGH-GHST volatility is smaller in magnitudes, although the time varying pattern is similar as the DCC volatility. The DGH-GHST-G estimates are less volatile than the other models, except the extreme large observation in the Merck series on year 2004.

7 Conclusion

In this paper we proposed a new time varying conditional correlation model that accounts for skewness and fat tails through the use of Generalized Hyperbolic (GH) distributions. The distinguishing feature of the model is that the shape of the observation distribution directly affects the mechanism by which volatilities and correlations are updated. The key mechanism by which the choice of distribution affects the updating equations for volatilities and correlations is the local density score. As a result, large observations are reweighted before they enter the updating equation. Because of this, the model is much less sensitive to outliers and incidental influential observations.

We showed that the model is easy to estimate by standard maximum likelihood procedures. In a simulation experiment, we demonstrated that the model does a better job at estimating the unknown correlation dynamics than competing models if the error distribution is fat-tailed and skewed. Compared to the DCC model, the new model is more stable under different parametric assumptions. When applied to real data, we showed that the model can result in a different assessment of volatility as well as correlation dynamics. Because of the automatic redistribution of influential observations to a part due to the fat-tailedness and a part due to increased volatility (or correlation) levels, the new model produces a more stable view on the level of volatility and correlation at any moment in time.

There are several directions in which the current work can be further pursued. The current

illustrations might be extended to applications in real economic decision scenarios. For example, the interesting link to exploit is the relation between the current empirical models, and the derivatives pricing literature based on Lévy processes. The new model provides a natural bridge between the two and might be a first step in specifying feasible empirical models that optimally approximate a Lévy stochastic process with stochastic volatility, similar to the work of Nelson and Foster (1994) for GARCH diffusions.

A Appendix

A.1 Skewness of the GH distribution

The skewness and kurtosis of y_t are most easily demonstrated by computing its higher order moments based on the mixture construction (14). Define $\bar{y}_t = \tilde{L}_t^{-1} y_t$ and $m_{i\zeta} = \mathrm{E}[(\zeta_t - \mu_\zeta)^i]$ for integer i. Also let e_i denote the ith column of I_k. We obtain

$$\mathrm{E}[\bar{y}_t] = 0, \tag{A1}$$

$$\mathrm{E}[\bar{y}_t \bar{y}_t'] = \mu_\zeta \mathrm{I} + m_{2\zeta} \gamma\gamma' = (T'T)^{-1}, \tag{A2}$$

$$\mathrm{E}[\bar{y}_t \otimes \bar{y}_t \bar{y}_t'] = m_{3\zeta} \gamma \otimes \gamma\gamma' + m_{2\zeta} \begin{pmatrix} \gamma_1 \mathrm{I}_k + \gamma e_1' + e_1 \gamma' \\ \vdots \\ \gamma_k \mathrm{I}_k + \gamma e_k' + e_k \gamma' \end{pmatrix}, \tag{A3}$$

such that the skewness only depends on γ and on the variance and skewness of the mixing variable ζ_t.

A.2 The Score of the GH distribution

Define the matrix $\mathrm{vec}(L) = \mathcal{D}_k^0 \mathrm{vech}(L)$ for a lower triangular matrix L. Note that $\mathcal{D}_k^0$ is different from the standard duplication matrix $\mathcal{D}_k$ for a symmetric matrix S, i.e., $\mathrm{vec}(S) = \mathcal{D}_k \mathrm{vech}(S)$ with $\mathcal{B}_k = (\mathcal{D}_k' \mathcal{D}_k)^{-1} \mathcal{D}_k'$. Also note that $\mathcal{D}_k^{0\prime} \mathcal{D}_k^0 = \mathrm{I}_k$, such that $\mathcal{B}_k^0 = \mathcal{D}_k^{0\prime}$. Finally, let $\mathcal{C}_k$ be the commutation matrix, $\mathrm{vec}(S') = \mathcal{C}_k \mathrm{vec}(S)$ for an arbitrary matrix S. For completeness, we mention that $\tilde{L}_t = L_t T$, and $\tilde{\Sigma}_t = \tilde{L}_t \tilde{L}_t'$.

An intermediate result is

$$\begin{aligned}
\mathrm{d}\Sigma_t &= \mathrm{d}(L_t L_t') \quad \Leftrightarrow \\
\mathrm{vec}(\mathrm{d}\Sigma_t) &= (\mathrm{I}_{k^2} + \mathcal{C}_k)(L_t \otimes \mathrm{I}_k) \mathrm{vec}(\mathrm{d}L_t) \quad \Leftrightarrow \\
\mathcal{D}_k \mathrm{vech}(\mathrm{d}\Sigma_t) &= (\mathrm{I}_{k^2} + \mathcal{C}_k)(L_t \otimes \mathrm{I}_k) \mathcal{D}_k^0 \mathrm{vech}(\mathrm{d}L_t) \quad \Leftrightarrow \\
\mathrm{vech}(\mathrm{d}L_t) &= \left(\mathcal{B}_k (\mathrm{I}_{k^2} + \mathcal{C}_k)(L_t \otimes \mathrm{I}_k) \mathcal{D}_k^0\right)^{-1} \mathrm{vech}(\mathrm{d}\Sigma_t).
\end{aligned} \tag{A4}$$

First define the standardized y_t as $x_t = \tilde{L}_t^{-1} y_t + \mu_\zeta \gamma$. The random variable x_t then has a GH distribution with location 0 and scaling matrix I_k. Let $d_z^\nu = \nu + z'z$ for a scalar ν and a vector

z. With this notation, the density of the GH distribution of y_t is given by

$$p^{GH}(y_t|f_t;\lambda,\chi,\psi,\mu_\zeta,\sigma_\zeta^2,\gamma,\Sigma_t) = \frac{e^{\gamma' x_t}}{|2\pi\tilde{L}_t\tilde{L}_t'|^{1/2}} \cdot \frac{\left(\sqrt{d_{x_t}^\chi/d_\gamma^\psi}\right)^{\lambda-k/2} \cdot K_{\lambda-k/2}\left(\sqrt{d_{x_t}^\chi d_\gamma^\psi}\right)}{(\sqrt{\chi/\psi})^\lambda \cdot K_\lambda\left(\sqrt{\chi\psi}\right)}, \quad \text{(A5)}$$

Let $k_\lambda(\cdot) = \ln K_\lambda(\cdot)$ with first derivative $k_\lambda'(\cdot)$. Define the scalar weight

$$w_t = -\frac{\lambda - k/2}{d_{x_t}^\chi} - \frac{k_{\lambda-k/2}'\left(\sqrt{d_{x_t}^\chi d_\gamma^\psi}\right)}{\sqrt{d_{x_t}^\chi/d_\gamma^\psi}}. \quad \text{(A6)}$$

We obtain

$$\nabla_t = \frac{\partial \text{vech}(\Sigma_t)'}{\partial f_t}\frac{\partial \text{vech}(L_t)'}{\partial \text{vech}(\Sigma_t)}\frac{\partial \text{vec}(\tilde{L}_t)'}{\partial \text{vech}(L_t)}\frac{\partial \ln p^{GH}(y_t|f_t)}{\partial \text{vec}(\tilde{L}_t)} = \Psi_t'\bar{H}_t'\frac{\partial \ln p^{GH}(y_t|f_t)}{\partial \text{vec}(\tilde{L}_t)},$$

with $\Psi_t = \partial \text{vech}(\Sigma_t)/\partial f_t'$ and

$$\bar{H}_t = \left((T' \otimes \mathrm{I}_k)\mathcal{D}_k^0\right)\left(\mathcal{B}_k\left(\mathrm{I}_{k^2} + \mathcal{C}_k\right)\left(L_t \otimes \mathrm{I}_k\right)\mathcal{D}_k^0\right)^{-1} \quad \text{(A7)}$$

using the intermediate result (A4).

Taking the derivative of the log-density with respect to $\text{vec}(\tilde{L}_t)$ and then via the chain rule with respect to f_t, we get

$$\begin{aligned}
\frac{\partial \ln p^{GH}(y_t|f_t)}{\partial \text{vec}(\tilde{L}_t)} &= \frac{\partial x_t'}{\partial \text{vec}(\tilde{L}_t)}\left(-0.5 w_t \frac{\partial d_{x_t}^\chi}{\partial x_t} + \gamma\right) - \text{vec}((\tilde{L}_t')^{-1}) \\
&= (\tilde{L}_t^{-1}y_t \otimes (\tilde{L}_t')^{-1})(w_t x_t - \gamma) - \text{vec}((\tilde{L}_t')^{-1}) \\
&= (\tilde{L}_t^{-1} \otimes (\tilde{L}_t')^{-1})(y_t \otimes \mathrm{I})(w_t\tilde{L}_t^{-1}y_t + w_t\mu_\zeta\gamma - \gamma) - \text{vec}((\tilde{L}_t')^{-1}) \\
&= (\tilde{L}_t' \otimes \mathrm{I})(\tilde{\Sigma}_t^{-1} \otimes \tilde{\Sigma}_t^{-1})\Big(w_t y_t \otimes y_t - \text{vec}(\tilde{\Sigma}_t) - (1 - w_t\mu_\zeta)(y_t \otimes \tilde{L}_t\gamma)\Big) \quad \text{(A8)}
\end{aligned}$$

The main result is now obtained by defining

$$H_t' = \bar{H}_t'(\tilde{L}_t' \otimes \mathrm{I})(\tilde{\Sigma}_t^{-1} \otimes \tilde{\Sigma}_t^{-1}).$$

A.3 EM algorithm for time-varying scale matrix $\tilde{\Sigma}_t$

We first prove Result 2. It is easy to check that

$$\begin{aligned}\nabla_t = \frac{\partial \ln p(y_t|f_t,\mathcal{F}_{t-1};\theta)}{\partial f_t} &= \frac{1}{p(y_t|f_t,\mathcal{F}_{t-1};\theta)}\int \frac{\partial p(y_t,\zeta_t|f_t,\mathcal{F}_{t-1};\theta)}{\partial f_t}\mathrm{d}\zeta_t \\ &= \int \frac{\partial p(y_t|\zeta_t,f_t,\mathcal{F}_{t-1};\theta)}{\partial f_t}\,\frac{p(\zeta_t;\theta_2)}{p(y_t|f_t,\mathcal{F}_{t-1};\theta)}\mathrm{d}\zeta_t \\ &= \int \frac{\partial \ln p(y_t|\zeta_t,f_t,\mathcal{F}_{t-1};\theta_1)}{\partial f_t}\,\frac{p(y_t,\zeta_t|f_t,\mathcal{F}_{t-1};\theta)}{p(y_t|f_t,\mathcal{F}_{t-1};\theta)}\mathrm{d}\zeta_t \\ &= \mathrm{E}\left[\frac{\partial \ln p(y_t|\zeta_t,f_t,\mathcal{F}_{t-1};\theta)}{\partial f_t}\middle|\,\mathcal{F}_t\right] \quad = \quad \tilde{\nabla}_t. \end{aligned} \tag{A9}$$

Throughout, we impose the normalization constraint $\mu_\zeta = 1$. We note that

$$\ln p(y_t|\zeta_t,f_t,\mathcal{F}_{t-1};\theta) = -\frac{1}{2}\ln|\tilde{\Sigma}_t| - \frac{k}{2}\ln(\zeta_t) - \frac{k}{2}\ln(2\pi) - \frac{1}{2\zeta_t}(y_t-(\zeta_t-\mu_\zeta)\tilde{L}_t\gamma)'\tilde{\Sigma}_t^{-1}(y_t-(\zeta_t-\mu_\zeta)\tilde{L}_t\gamma), \tag{A10}$$

and

$$\ln p(\zeta_t;\theta_2) = -\frac{\lambda}{2}\ln(\chi/\psi) - \ln(2) - \ln K_\lambda\left(\sqrt{\chi\psi}\right) + (\lambda-1)\ln(\zeta_t) - \frac{1}{2}(\chi\zeta_t^{-1}+\psi\zeta_t), \tag{A11}$$

where $\tilde{L}_t = \tilde{L}(f_t)$ and $\tilde{\Sigma}_t = \tilde{\Sigma}(f_t)$, and where the mapping from f_t to $\tilde{\Sigma}_t$ does not depend on θ_2.

We define $\tilde{x}_t = \tilde{L}_t^{-1}y_t + \gamma$. From (A10) and (A11) and the properties of the Generalized Inverse Gaussian distribution (see Appendix A.2 of McNeil et al. (2005)), we get

$$\delta_{1t}^{(\ell)} = \mathrm{E}_{\hat{\theta}^{(\ell)}}\left[\zeta^{-1}\middle|\,\mathcal{F}_n\right] = \left(\frac{d_{\tilde{x}_t}^\chi}{d_\gamma^\psi}\right)^{-1/2}\frac{K_{\lambda-1-k/2}\left(\sqrt{d_{\tilde{x}_t}^\chi d_\gamma^\psi}\right)}{K_{\lambda-k/2}\left(\sqrt{d_{\tilde{x}_t}^\chi d_\gamma^\psi}\right)}, \tag{A12}$$

$$\delta_{2t}^{(\ell)} = \mathrm{E}_{\hat{\theta}^{(\ell)}}\left[\zeta\middle|\,\mathcal{F}_n\right] = \left(\frac{d_{\tilde{x}_t}^\chi}{d_\gamma^\psi}\right)^{1/2}\frac{K_{\lambda+1-k/2}\left(\sqrt{d_{\tilde{x}_t}^\chi d_\gamma^\psi}\right)}{K_{\lambda-k/2}\left(\sqrt{d_{\tilde{x}_t}^\chi d_\gamma^\psi}\right)}, \tag{A13}$$

$$\delta_{3t}^{(\ell)} = \mathrm{E}_{\hat{\theta}^{(\ell)}}\left[\ln(\zeta)\middle|\,\mathcal{F}_n\right] = \frac{\partial}{\partial\xi}\left(\frac{d_{\tilde{x}_t}^\chi}{d_\gamma^\psi}\right)^{\xi/2}\frac{K_{\lambda+\xi-k/2}\left(\sqrt{d_{\tilde{x}_t}^\chi d_\gamma^\psi}\right)}{K_{\lambda-k/2}\left(\sqrt{d_{\tilde{x}_t}^\chi d_\gamma^\psi}\right)}\Bigg|_{\xi=0}, \tag{A14}$$

where $d_\cdot^\cdot$ is defined below (A4).

From (A10) and using $\mu_\zeta = 1$, we obtain

$$\frac{\partial \ln p(y_t|\zeta_t, f_t, \mathcal{F}_{t-1}; \theta)}{\partial f_t} = \tilde{\Psi}_t' \tilde{H}_t' \text{vec}\left(\zeta_t^{-1} y_t (y_t + \tilde{L}_t \gamma)' - \tilde{\Sigma}_t\right) \tag{A15}$$

with $\tilde{\Psi}_t = \partial \text{vech}(\tilde{\Sigma}_t)/\partial f_t'$ and

$$\tilde{H}_t = (\tilde{\Sigma}_t^{-1} \otimes \tilde{\Sigma}_t^{-1})(\tilde{L}_t \otimes \mathrm{I})\mathcal{D}_k^0(\mathcal{B}_k(\mathrm{I}_{k^2} + \mathcal{C}_k)(\tilde{L}_t \otimes \mathrm{I}_k)\mathcal{D}_k^0)^{-1},$$

and with $\tilde{L}_t$ a lower triangular matrix. Taking conditional expectations, we obtain

$$\tilde{\nabla}_t^{(\ell)} = \tilde{\Psi}_t' \tilde{H}_t' \left(\delta_{1t}^{(\ell)} y_t \otimes y_t - \text{vec}(\tilde{\Sigma}_t) - (1 - \delta_{1t}^{(\ell)})(y_t \otimes \tilde{L}_t \gamma)\right), \tag{A16}$$

which only depends on $\hat{\theta}^{(\ell)}$, γ, and $\tilde{\Sigma}_t$, and therefore not on θ_2. As a result, the modified model for y_t conditional on ζ_t depends on θ_1 only.

Using these results, it is clear that $Q_1(\cdot)$ only depends on θ_1 and $\hat{\theta}^{(\ell)}$ as well. We have

$$Q_1(\theta_1, \hat{\theta}^{(\ell)}) = -\frac{1}{2}\ln|\tilde{\Sigma}_t| - \frac{k}{2}\delta_{3t}^{(\ell)} - \frac{k}{2}\ln(2\pi) - \frac{1}{2}\delta_{1t}^{(\ell)}\tilde{x}_t'\tilde{x}_t + \tilde{x}_t'\gamma - \frac{1}{2}\delta_{2t}^{(\ell)}\gamma'\gamma. \tag{A17}$$

For expositional purposes, we restrict our attention to the model with order (1,1) dynamics,

$$f_{t+1} = A_1 S_t \tilde{\nabla}_t^{(\ell)} + B_1 f_t. \tag{A18}$$

Optimizing (A17) using the dynamics in (A18) now becomes similar to estimating a Gaussian multivariate GARCH in Mean model. The transition equation uses weighted (by $\delta_{1t}^{(\ell)}$) rather than unweighted squared observations to drive volatilities and correlations, see (A16). Similarly, there is weighting by $\delta_{1t}^{(\ell)}$ in the objective (A17). When optimizing over θ_1, however, these weights are fixed. Numerical optimization should therefore be faster than direct ML estimation of the full θ vector due to the less complicated likelihood and the lower dimensional parameter space.

Using the new estimate of θ_1 obtained by maximizing (A17), we update the parameter estimate to $\tilde{\theta}^{(\ell)}$, and use this new estimate to update the weights $\delta_{it}^{(\ell)}$. The second part of the EM maximization step then follows from

$$Q_2(\theta_2, \tilde{\theta}^{(\ell)}) = -\frac{\lambda}{2}\ln(\chi/\psi) - \ln(2) - \ln K_\lambda\left(\sqrt{\chi\psi}\right) + (\lambda - 1)\delta_{3t}^{(\ell)} - \frac{1}{2}(\chi\delta_{1t}^{(\ell)} + \psi\delta_{2t}^{(\ell)}), \tag{A19}$$

which can be optimized numerically with respect to θ_2 under the constraint $\mu_\zeta = 1$. This can be achieved by optimizing over $\kappa = \chi\psi > 0$ and λ, and using (10).

The similarity of (A16) and (A8) can be taken a step further by noting that $w_t = \delta_{1t}^{(\infty)}$, where $\delta_{1t}^{(\infty)}$ is evaluated using the true parameters. This follows from the fact that

$$w_t - \delta_{1t}^{(\infty)} = -\frac{\lambda - k/2}{d_{\tilde{x}_t}^{\chi}} + 0.5\frac{K_{\lambda-k/2+1}\left(\sqrt{d_{\tilde{x}_t}^{\chi} d_{\gamma}^{\psi}}\right) - K_{\lambda-k/2-1}\left(\sqrt{d_{\tilde{x}_t}^{\chi} d_{\gamma}^{\psi}}\right)}{K_{\lambda-k/2}\left(\sqrt{d_{\tilde{x}_t}^{\chi} d_{\gamma}^{\psi}}\right)\ \sqrt{d_{\tilde{x}_t}^{\chi}/d_{\gamma}^{\psi}}} \tag{A20}$$

and the properties of the modified Bessel function of the second kind,

$$K_{\lambda+1}(\kappa) = 2\lambda \cdot \kappa^{-1} \cdot K_{\lambda}(\kappa) + K_{\lambda-1}(\kappa),$$

and

$$\frac{\partial \ln K_{\lambda}(\kappa)}{\partial \kappa} = \frac{K_{\lambda+1}(\kappa) + K_{\lambda-1}(\kappa)}{2 \cdot K_{\lambda}(\kappa)},$$

such that from (A20) it follows that $w_t - \delta_{1t}^{(\infty)} = 0$.

References

Alexander, C. (1998). Volatility and correlation: Methods, models and applications. *Risk Manangement and Analysis: Measuring and Modelling Financial Risk*.

Alexander, C. (2001). Orthogonal garch. *Mastering risk 2*, 21–38.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 353*(1674), 401–419.

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of statistics 5*(3), 151–157.

Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics 23*(3), 346–355.

Blæsild, P. (1981). The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data. *Biometrika 68*(1), 251.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics 72*(3), 498–505.

Boswijk, H. and R. Van der Weide (2006). Wake me up before you GO-GARCH. *Discussion Papers*, 06–079.

Creal, D., S. Koopman, and A. Lucas (2008). A general framework for observation driven time-varying parameter models.

Creal, D., S. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, (forthcoming).

D'Agostino, R., A. Belanger, and R. D'Agostino Jr (1990). A suggestion for using powerful and informative tests of normality. *American Statistician 44*(4), 316–321.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B 39*(1), 1–38.

Eberlein, E. and U. Keller (1995). Hyperbolic distributions in finance. *Bernoulli 1*(3), 281–299.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. and K. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*(1), 122–150.

Fiorentini, G., E. Sentana, and G. Calzolari (2003). Maximum likelihood rstimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student's T innovations. *Journal of Business & Economic Statistics 21*(4), 532–547.

Franses, P. and D. Van Dijk (2000). *Nonlinear time series models in empirical finance*. Cambridge Univ Pr.

Harvey, A. and T. Chakravarty (2008). Beta-t-(E) GARCH. *Cambridge Working Papers in Economics*.

Hu, W. (2005). *Calibration Of Multivariate Generalized Hyperbolic Distributions Using The EM Algorithm, With Applications In Risk Management, Portfolio Optimization And Portfolio Credit Risk*. Ph. D. thesis.

Jaeckel, P. and R. Rebonato (1999/2000, Winter). The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk 2*(2), 17–28.

McNeil, A., R. Frey, P. Embrechts, et al. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton university press Princeton, NJ.

Nelson, D. B. and D. P. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica 62*(1), 1–41.

Peters, J. (2001). Estimating and forecasting volatility of stock indices using asymmetric GARCH models and (Skewed) Student-t densities. *Preprint, University of Liege, Belgium, 20 pp. pdf*.

Prause, K. (1999). The generalized hyperbolic model: estimation, financial derivatives, and risk measures. *University of Freiburg, Doctoral Thesis*.

Sentana, E. (2004). Estimation and testing of dynamic models with generalised hyperbolic innovations. *FMG Discussion Papers*.

Shephard, N. (2005). *Stochastic Volatility: selected readings*. Oxford University Press.

van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics 17*(5), 549–564.

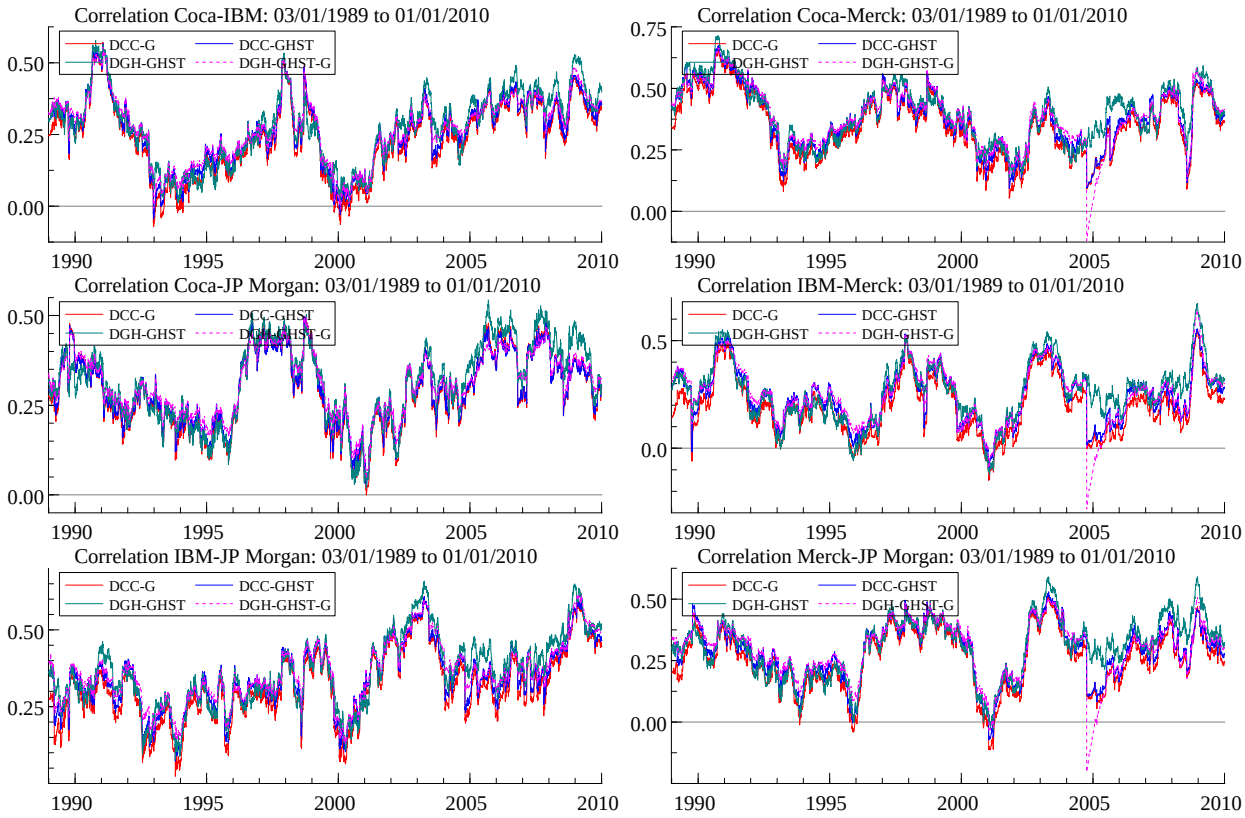

Figure 1: Estimated correlation paths for the DCC, and DGH models.

The correlation estimates from the DCC-G, DCC-GHST, DGH-GHST-G and DGH-GHST models with stock return data. We can see that the DGH(1,1) is less volatile than the estimates from DCC. The estimates from DGH-GHST is the most smooth estimate. The DCC correlation estimate does not change too much from the Gaussian distribution to the GH distribution. And the DGH-GHST-G correlation is quite close to the result from DCC-GHST. There are a few differences in several special periods. We will show it in a zoom-in picture in the next page.

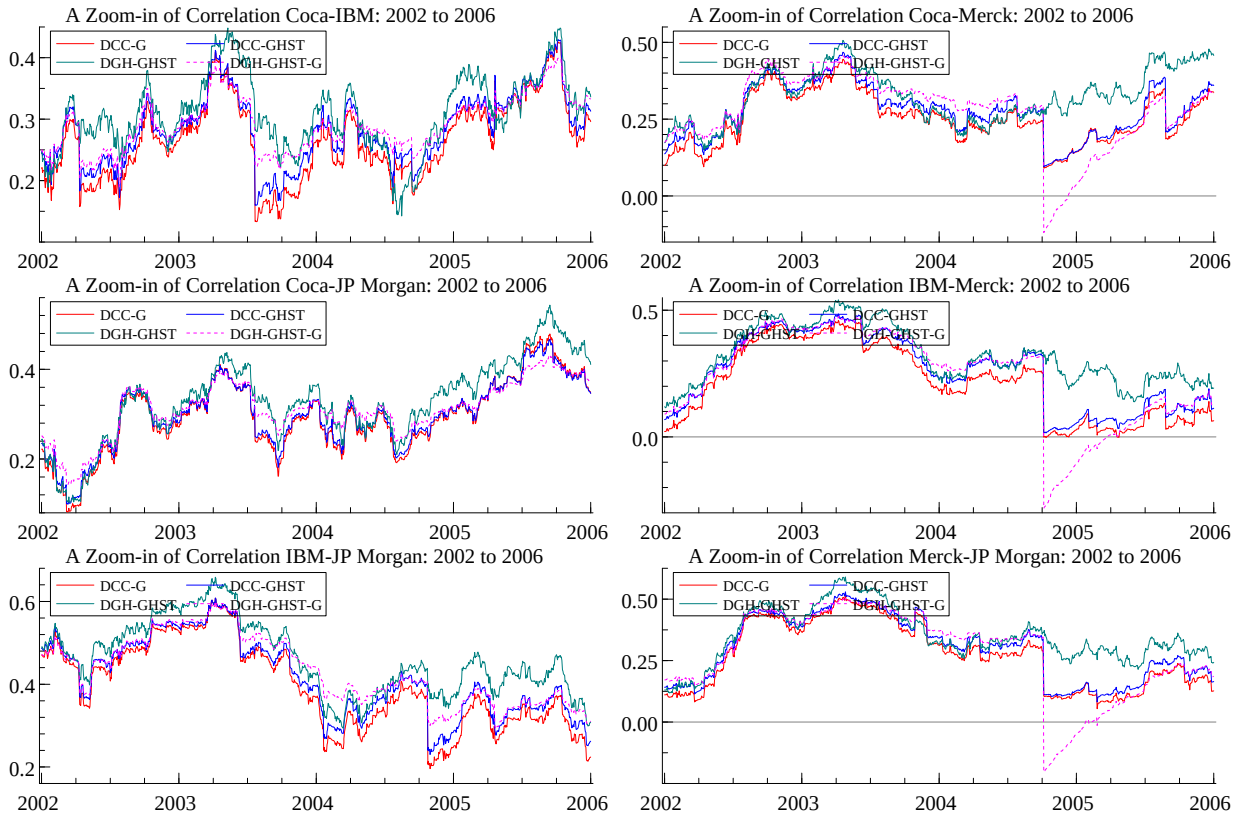

Figure 2: Estimated correlation subsample for the DCC, and DGH models.

To look into the difference of the DGH(1,1) and the DCC(1,1) model, we have zoom-in plot of the estimation under different parametric assumptions.

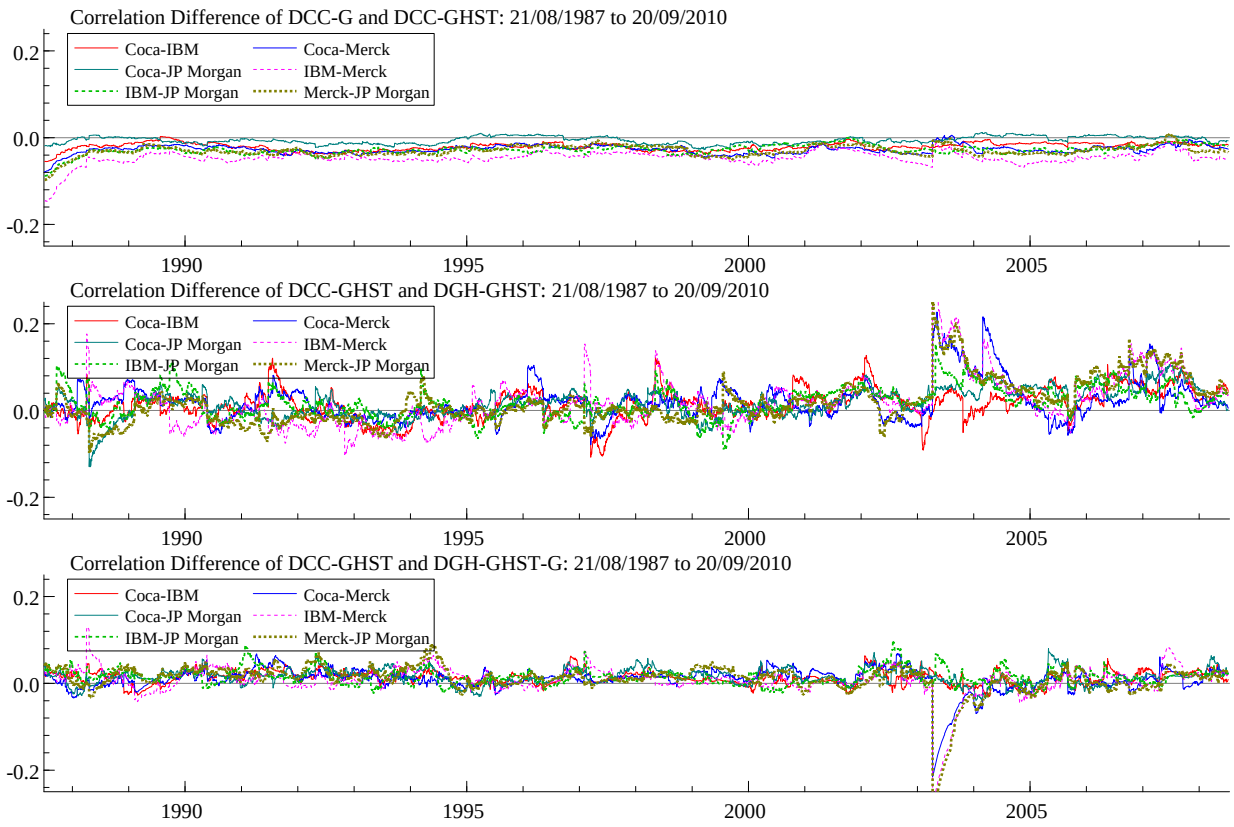

Figure 3: Estimated correlation difference for the DCC, and DGH models.

To look into the difference of the DGH(1,1) and the DCC(1,1) model, we have the plot of the difference in estimation under different models. It appears that the DCC-G is not so much difference from DCC-GHST. And DGH-GHST-G behaves closely to the DCC-GHST model.

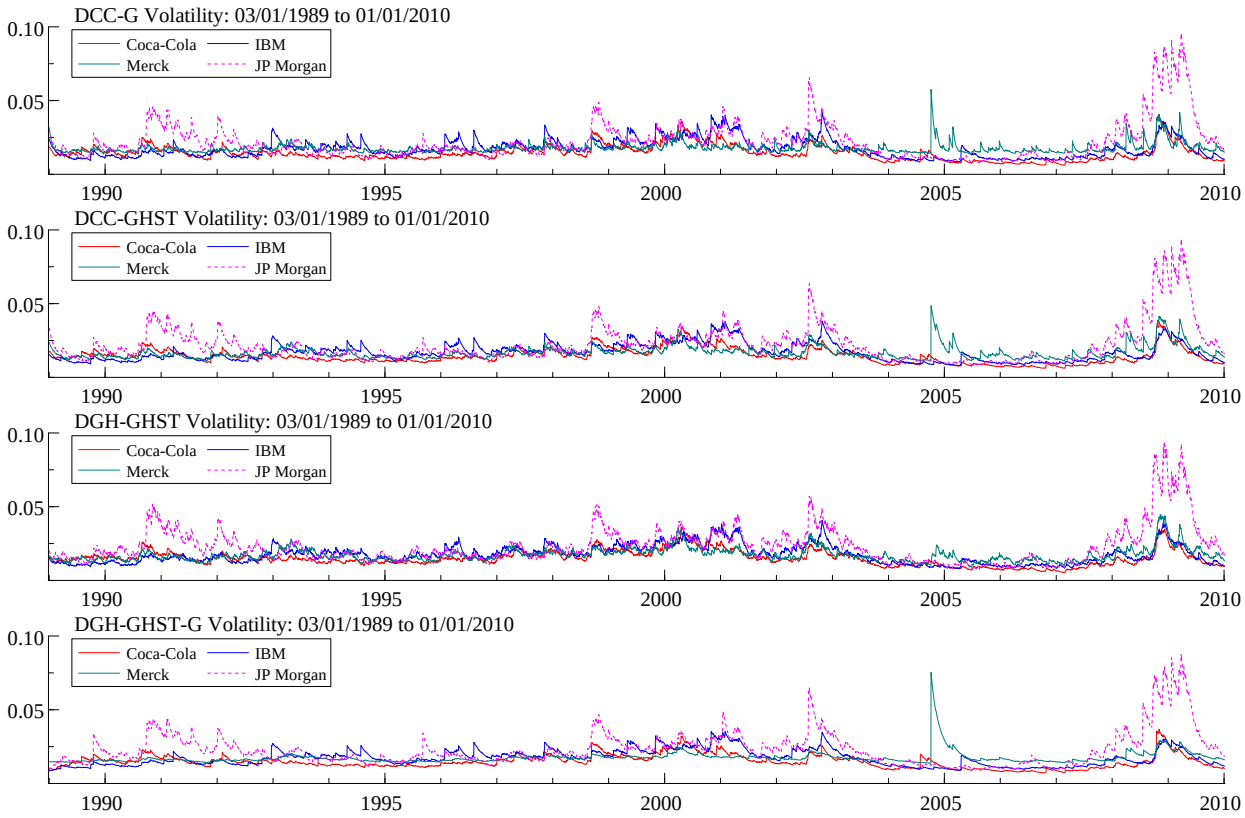

Figure 4: Volatility estimation from DCC and DGH stock returns

The volatility estimates from the DCC under Gaussian and GH distributions, DGH(1,1)-GHST and DGH(1,1)-GHST-G with stock return data. From the graphs, we can see that the volatility from DGH-GHST is smoother than the DCC estimates, even with GH distributions.

Table 1: Mean Absolute Value of Volatility Estimation with GARCH and DGH models.

The table presents the average Mean Absolute Error (MAE) over 100 Monte Carlo replications for the variance estimates for three different distributions (in pairs of columns) and five different time-varying volatility patterns. The distributions used are the normal, Student's $t(5)$, and GHST$(0, \Sigma_t, -0.03, \nu)$. The GARCH models use the correct specification of the distribution. For example, the GARCH model for the Student's $t(5)$ DGP is estimated using a Student's t distribution (with estimated degrees of freedom) for the error terms. The boldface numbers show the models with the smallest MAE for a given DGP. Monte Carlo standard errors of the MAEs are provided in parentheses.

	normal		$t(5)$		GHST	
	GARCH	model (6)–(9)	GARCH	model (6)–(9)	GARCH	model (6)–(9)
Constant	0.0375	**0.0353**	**0.0531**	0.0545	**0.0505**	0.0545
	(0.0244)	(0.0222)	(0.0362)	(0.0382)	(0.0327)	(0.0427)
Sine	**0.1209**	0.1487	0.1544	**0.1456**	0.1557	**0.1533**
	(0.0104)	(0.0554)	(0.0202)	(0.0237)	(0.0222)	(0.0242)
Fast Sine	**0.2405**	0.2593	0.3005	**0.2220**	0.3037	**0.2806**
	(0.0188)	(0.0161)	(0.0248)	(0.0119)	(0.0282)	(0.1783)
Step	**0.0810**	0.0851	0.1074	**0.0891**	0.1062	**0.0941**
	(0.0126)	(0.0146)	(0.0278)	(0.0364)	(0.0181)	(0.0272)
Ramp	**0.1455**	0.1633	0.1985	**0.1853**	0.2022	**0.1972**
	(0.0094)	(0.0199)	(0.0225)	(0.0226)	(0.0260)	(0.0389)

Table 2: Mean Absolute Value of Correlation Estimation with DCC and DGH models.

The table presents the average Mean Absolute Error (MAE) over 100 Monte Carlo replications for the correlation estimates for three different distributions (in pairs of columns) and five different time-varying volatility patterns. The distributions used are the normal, Student's $t(5)$, and GHST$(0, \Sigma_t, -0.03, \nu)$. The DCC models use the correct specification of the distribution. For example, the DCC model for the Student's $t(5)$ DGP is estimated using a Student's t distribution (with estimated degrees of freedom) for the error terms. The boldface numbers show the models with the smallest MAE for a given DGP. Monte Carlo standard errors of the MAEs are provided in parentheses.

	normal		$t(5)$		GHST	
	DCC	model (6)–(9)	DCC	model (6)–(9)	DCC	model (6)–(9)
Constant	**0.0037**	**0.0037**	0.0057	**0.0050**	**0.0053**	**0.0053**
	(0.0028)	(0.0028)	(0.0043)	(0.0043)	(0.0038)	(0.0038)
Sine	0.1348	**0.1331**	0.1387	**0.1292**	0.1506	**0.1334**
	(0.0121)	(0.0121)	(0.0112)	(0.0104)	(0.0181)	(0.0119)
Fast Sine	**0.2254**	0.2261	0.2553	**0.2187**	0.2543	**0.2210**
	(0.0051)	(0.0046)	(0.0019)	(0.0056)	(0.0081)	(0.0054)
Step	0.0658	**0.0648**	0.0684	**0.0673**	0.0941	**0.0700**
	(0.0098)	(0.0099)	(0.0098)	(0.0095)	(0.0237)	(0.0097)
Ramp	0.1593	**0.1568**	**0.1587**	0.1645	0.1684	**0.1638**
	(0.0123)	(0.0125)	(0.0105)	(0.0095)	(0.0161)	(0.0094)

Table 3: Data descriptive statistics.

	Coca-Cola	IBM	Merck	JP Morgan
Mean $\times 10^4$	6.3327	5.2667	5.6934	7.5582
Median	0.0000	0.0000	0.0000	0.0000
Standard Deviation	0.0156	0.0189	0.0182	0.0262
Sample Variance	0.0002	0.0004	0.0003	0.0007
Skewness	0.2289	0.2942	-0.1193	0.7417
Skewness test Prob.	0.0000	0.0000	0.0000	0.0000
Kurtosis	5.0476	6.8934	5.9988	11.3329
Kurtosis test Prob.	0.0000	0.0000	0.0000	0.0000
Range	0.2436	0.2871	0.2777	0.4582
Minimum	-0.1048	-0.1554	-0.1474	-0.2073
Maximum	0.1388	0.1316	0.1303	0.2510
Omnibus K^2 test Prob.	0.0000	0.0000	0.0000	0.0000

The descriptive statistics for the data sets used in the empirical application section 6. The data are collected in CRSP between January 1989 and December 2009. All observations are daily returns, namely $y_t = \log p_t - \log p_{t-1}$. This data set contains four stocks of major companies in four different industries, which are all composite of the DJ30 index. The Omnibus K^2 statistic test refers to the test for non-normality due to either skewness or kurtosis, and the Skewness/Kurtosis tests are constructed as in D'Agostino et al. (1990).

Table 4: Estimation Results DCC(1,1) and DGH(1,1).

	DCC-G		DCC-GHST		DGH-GHST-G		DGH-GHST	
Parms	Value	S.E.	Value	S.E.	Value	S.E.	Value	S.E.
A_{d1}	0.0372 *	0.0041	0.0287 *	0.0035	0.0224 *	0.0025	0.0322 *	0.0040
A_{d2}	0.0355 *	0.0047	0.0262 *	0.0031	0.0174 *	0.0017	0.0339 *	0.0038
A_{d3}	0.0376 *	0.0072	0.0297 *	0.0052	0.0090 *	0.0013	0.0384 *	0.0052
A_{d4}	0.0575 *	0.0050	0.0528 *	0.0055	0.0368 *	0.0035	0.0520 *	0.0046
A_{ϕ}	0.0104 *	0.0010	0.0097 *	0.0011	0.0071 *	0.0008	0.0098 *	0.0010
B_{d1}	0.9594 *	0.0044	0.9691 *	0.0038	0.9938 *	0.0011	0.9957 *	0.0014
B_{d2}	0.9589 *	0.0053	0.9686 *	0.0036	0.9947 *	0.0008	0.9944 *	0.0016
B_{d3}	0.9128 *	0.0179	0.9563 *	0.0082	0.9878 *	0.0029	0.9887 *	0.0034
B_{d4}	0.9387 *	0.0052	0.9438 *	0.0057	0.9928 *	0.0011	0.9936 *	0.0015
B_{ϕ}	0.9857 *	0.0015	0.9854 *	0.0019	0.9959 *	0.0008	0.9964 *	0.0009
c_{d1}	2.4433 *	0.4496	2.5601 *	0.6043	0.8451 *	0.0756	1.4419 *	0.0877
c_{d2}	3.0548 *	0.4008	2.0451 *	0.2783	0.8934 *	0.0791	1.5347 *	0.0780
c_{d3}	3.2550 *	0.1125	2.6745 *	0.1995	1.4845 *	0.0253	1.6826 *	0.0490
c_{d4}	5.1516 *	1.1859	6.5864 *	1.7241	1.0527 *	0.0917	2.0121 *	0.0976
$c_{\phi 12}$	0.2739 *	0.0432	0.3198 *	0.0391	0.7748 *	0.0378	0.7792 *	0.0503
$c_{\phi 13}$	0.3589 *	0.0401	0.4317 *	0.0356	0.7520 *	0.0369	0.7570 *	0.0486
$c_{\phi 14}$	0.3021 *	0.0418	0.3108 *	0.0369	0.7744 *	0.0354	0.7771 *	0.0491
$c_{\phi 23}$	0.1600 *	0.0439	0.3055 *	0.0404	0.7974 *	0.0402	0.8015 *	0.0525
$c_{\phi 24}$	0.3024 *	0.0414	0.3842 *	0.0364	0.7759 *	0.0370	0.7811 *	0.0502
$c_{\phi 34}$	0.2183 *	0.0431	0.3152 *	0.0374	0.8007 *	0.0387	0.8048 *	0.0525
ν	-	-	6.4343 *	0.2351	5.9633 *	0.2324	6.3205 *	0.2449
γ_1	-	-	0.0377	0.0233	0.0395	0.0220	0.0729 *	0.0236
γ_2	-	-	0.0212	0.0228	0.0233	0.0215	0.0631 *	0.0234
γ_3	-	-	-0.0316	0.0233	-0.0209	0.0220	-0.0190	0.0232
γ_4	-	-	0.0463 *	0.0229	0.0474 *	0.0220	0.0707 *	0.0234
Log-Likelihood	-39990.98		-38787.07		-38877.96		-38683.78	

* Significant at 5% level.

The estimation results for the parameters in the DGH(1,1) and DCC(1,1) model, using four stock return data between January 1989 and December 2009. The factor model is subject to the identifying restrictions in Section 2. The DCC model is defined the same as in Engle (2002). We estimate the DCC model under the Gaussian and GH distributions. The DGH model is estimated under GH distributions, but the DGH steps are chosen to be GH and Gaussian. The standard errors and the log-likelihood are reported.

Notation: c_{di}s with $i = 1, 2, 3, 4.$ are the unconditional mean of the volatilities for stock returns. $c_{\phi ij}$ is the unconditional mean of the hypersphere coordinates coefficient ϕij in the correlation matrix, as in Equation (12). γ is the skewness parameter vector in the GH-skewed$t$ density. μ is the mean vector in the Gaussian density and in the GHST distribution. ν is the degree of freedom. A_{di} and $A_{\phi ij}$ report the diagonal element of A and B_{di} and $B_{\phi ij}$ are the diagonal elements of B. A and B refer to the coefficient matrices in the DGH equation (6). $(i = 1, 2, 3)$