Bayesian Multivariate GARCH Models with Dynamic Correlations and Asymmetric Error Distributions

José A. Fioruci[a], Ricardo S. Ehlers[a,*], Marinho G. Andrade Filho[a]

[a] *Universidade de São Paulo, São Carlos, Brazil*

Abstract

The main goal in this paper is to develop and apply stochastic simulation techniques for GARCH models with multivariate skewed distributions using the Bayesian approach. Both parameter estimation and model comparison are not trivial tasks and several aproximate and computationally intensive methods (Markov chain Monte Carlo) will be used to this end. We consider a flexible class of multivariate distributions which can model both skewness and heavy tails. Also, we do not fix tail behaviour when dealing with fat tail distributions but leave it subject to inference.

Keywords: Multivariate GARCH, multivariate skewed distributions, Markov chain Monte Carlo, Metropolis-Hastings.

1 Introduction

There is considerable empirical evidence in the literature in various fields that points to leptokurtic and asymmetric distributions. Probability distributions that are capable of modelling the presence of asymmetry in the distribution of the phenomenon under study have recently been the focus of great interest (see for example Genton 2004 for a review on this topic). In particular, when

*Corresponding author. Email: ehlers@icmc.usp.br

studying financial time series the returns can frequently show some degree of asymmetry and the recent literature provides evidence that modelling this asymmetry in the volatility process many times fails in capturing all the asymmetry in the distribution of these returns (e.g. Gourieroux 1997 and Alexander 2001). Therefore, approaches that also acommodate asymmetry as well as fat tails should be taken into account in the modelling process. This can be accomplished by allowing for some degree of asymmetry in the distribution of the model error. The main contribution of this paper is to use a new class of flexible multivariate skewed distributions in a multivariate GARCH context that will be described in the sequel, and propose Bayesian inference procedures.

There are a number of proposals in the literature to introduce skewness in unimodal symmetric distributions, both uni and multivariate (e.g. Azzalini 1985, Fernandez and Steel 1998, Branco and Dey 2001, Azzalini and Capitanio 2003 and Jones and Faddy 2003). In particular, in the recent financial econometrics literature Cappuccio, Lubian, and Raggi (2004) adopted a Bayesian approach to estimate univariate stochastic volatility models with skew exponential power distributions and Forsberg and Bollerslev (2002) used a normal inverse Gaussian distribution for the error term in univariate GARCH models. Bauwens and Laurent (2005) proposed a new class of skew multivariate Student t distributions while Bauwens, Laurent, and Rombouts (2006) provide a good review of various specifications for multivariate GARCH models. In these last two papers the authors apply (quasi-) maximum likelihood methods for parameter estimation.

In this paper we adopt a fully Bayesian approach to estimate all unknown parameters and compare multivariate GARCH models used for joint modelling financial returns time series allowing for asymmetry in the error terms distribution.

For a multivariate time series $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{kt})'$ the model is given by,

$$\boldsymbol{y}_t = \boldsymbol{H}_t^{1/2} \boldsymbol{\epsilon}_t \tag{1}$$

where $\boldsymbol{H}_t^{1/2}$ is any $k \times k$ positive definite matrix such that the conditional variance of $\boldsymbol{y}_t$ is $\boldsymbol{H}_t$ and which depends on a finite vector of parameters $\boldsymbol{\theta}$. The $k \times 1$ error vectors are assumed independent and identically distributed with $E(\boldsymbol{\epsilon}_t) = 0$ and $Var(\boldsymbol{\epsilon}_t) = \boldsymbol{I}_k$, where $\boldsymbol{I}_k$ is the identity matrix of order k. The literature on multivariate GARCH models has developed significantly over the past few years.

There are different possible specifications for $\boldsymbol{H}_t$ (for example the VEC model of Bollerslev, Engle, and Wooldridge 1988 and the BEKK model of Engle and Kroner 1995). In this paper, we focus on the so called conditional correlation models which allow to specify separately the individual conditional variances and the conditional correlation matrix (or any other measure of dependence between the individual series such as the copula of the conditional joint distribution). Bollerslev (1990) proposed a parsimonious approach in which the conditional covariances are proportional to the product of the corresponding conditional standard deviations. The constant conditional correlation (CCC) model is defined as,

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t,$$

where $\boldsymbol{D}_t = \text{diag}(h_{11,t}^{1/2}, \ldots, h_{kk,t}^{1/2})$, $\boldsymbol{R}$ is a symmetric positive definite matrix which elements are the (constant) conditional correlations ρ_{ij}, $i, j = 1, \ldots, k$ (with $\rho_{ij} = 1$, for $i = j$). So, each conditional covariance is given by $h_{ij,t} = \rho_{ij}\sqrt{h_{ii,t}\, h_{jj,t}}$. Besides, each conditional variance in $\boldsymbol{D}_t$ is specified as a univariate GARCH model. Here we specify a GARCH(1,1) model for each conditional variance, i.e.

$$h_{ii,t} = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{ii,t-1}, \ i = 1, \ldots, k, \tag{2}$$

with $\omega_i > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ and $\alpha_i + \beta_i < 1$, $i = 1, \ldots, k$. Note that this model contains $k(k+5)/2$ parameters. It is clear that $\boldsymbol{H}_t$ is positive definite if and only if $h_{ii,t} > 0, i = 1, \ldots, k$ and $\boldsymbol{R}$ is positive definite.

Engle (2002), Christodoulakis and Satchell (2002) and Tse and Tsui (2002) independently proposed generalizations of the CCC model by allowing the conditional correlation matrix to be time dependent which is then called the dynamic conditional correlation (DCC) model. So, in this case $\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t$. Here we adopt the same approach in Engle (2002) by setting the following parsimonious formulation,

$$\boldsymbol{R}_t = \text{diag}(\boldsymbol{Q}_t)^{-1/2}\ \boldsymbol{Q}_t\ \text{diag}(\boldsymbol{Q}_t)^{-1/2},$$

where $\boldsymbol{Q}_t$ are $k \times k$ symmetric positive-definite matrices given by,

$$\boldsymbol{Q}_t = (1 - \alpha - \beta)\boldsymbol{R} + \alpha \boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}' + \beta \boldsymbol{Q}_{t-1}, \tag{3}$$

$\boldsymbol{u}_t = \boldsymbol{D}_t^{-1}\boldsymbol{y}_t$ are the standardized returns, $\boldsymbol{R}$ is the unconditional covariance matrix of $\boldsymbol{u}_t$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. After some algebra it is not difficult to see that the conditional covariances are given by $h_{ij,t} = q_{ij,t}\sqrt{h_{ii,t}h_{jj,t}}/\sqrt{q_{ii,t}q_{jj,t}}$. So, the matrix $\boldsymbol{Q}_t$ is written as a GARCH(1,1)-type equation and then transformed to give the correlation matrix $\boldsymbol{R}_t$.

Estimation

The conditional likelihood function of model (1), for a sample of observations $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ can be written as

$$\begin{aligned} l(\boldsymbol{\theta}) &= \prod_{t=1}^{n} |\boldsymbol{H}_t|^{-1/2}\ p_\epsilon(\boldsymbol{H}_t^{-1/2}\boldsymbol{y}_t) \\ &= \prod_{t=1}^{n} \left[\prod_{i=1}^{k} h_{ii,t}^{-1/2}\right] |\boldsymbol{R}_t|^{-1/2}\ p_\epsilon((\boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t)^{-1/2}\boldsymbol{y}_t) \end{aligned} \tag{4}$$

where $p_{\boldsymbol{\epsilon}}$ is the joint density function for $\boldsymbol{\epsilon}_t$. The set of all model parameters is represented by $\boldsymbol{\theta} = (\omega_1, \alpha_1, \beta_1, \ldots, \omega_k, \alpha_k, \beta_k, \rho_{12}, \ldots, \rho_{k-1,k})$.

The natural first choice for the error distribution in (1) would be a standardized multivariate normal distribution. However, normality is rejected in most applications as the unconditional distribution of most financial asset returns has fatter tails than implied by this model with normal errors. In the univariate case, the excess of (unconditional) kurtosis has been most commonly accommodated with Student t distributed errors (e.g. Baillie and Bollerslev 1989). A natural alternative in the multivariate case is then the multivariate Student t distribution (Fiorentini et al. 2003) which has the extra degrees of freedom parameter ν to be estimated. We assume that $\nu > 2$ so that $\boldsymbol{H}_t$ can always be interpreted as a conditional covariance matrix. The density function for this multivariate t distribution is given by

$$p(\boldsymbol{\epsilon}_t) = \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})[\pi(\nu-2)]^{k/2}} \left[1 + \frac{\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t}{\nu-2}\right]^{-\frac{\nu+k}{2}},$$

where $\Gamma(\cdot)$ is the Gamma function, and the likelihood function is easily obtained by applying (4). We note that this is the standardized version of the multivariate Student t distribution, i.e. $E(\boldsymbol{\epsilon}_t) = 0$ and $Var(\boldsymbol{\epsilon}_t) = \boldsymbol{I}_k$. Both this and the standard multivariate normal distributions are spherically symmetric about the origin, in which case $\boldsymbol{y}_t$ has an elliptically symmetric distribution and $p(\boldsymbol{y}_t) \propto |\boldsymbol{H}_t|^{-1/2} g(\boldsymbol{y}_t'\boldsymbol{H}_t^{-1}\boldsymbol{y}_t)$ for some nonnegative scalar function $g(\cdot)$. Elliptical distributions are able to model heavy tails but fail to capture asymmetric dependence structures. Finally, it is worth mentioning that there are in fact many candidates for the multivariate generalization of the Student t distribution (Kotz and Nadarajah 2004).

2 Skewed Distributions

It is often the case that financial time series can be skewed and many empirical studies found evidence that modelling asymmetry in the volatility process often fails to capture skewness in the returns distribution. Therefore, estimation methods that accommodate skewness should be taken into account too. This can be dealt with by allowing a certain amount of skewness in the error distribution.

There are a number of proposals in the literature to introduce skewness in unimodal symmetric distributions (e.g. Azzalini 1985, Fernandez and Steel 1998, Branco and Dey 2001, Azzalini and Capitanio 2003 and Jones and Faddy 2003). Basically, skewing a symmetric distribution it is possible to preserve some of its properties and different skewing mechanisms will preserve a different set of properties (see Genton 2004 for a review). In the financial econometrics literature Cappuccio, Lubian, and Raggi (2004) use the Bayesian approach to estimate univariate stochastic volatility models with skew generalized error distributions. Bauwens and Laurent (2005) proposed a new class of multivariate skew Student t distribution and Bauwens, Laurent, and Rombouts (2006) review many existing multivariate GARCH especifications. Both these papers are based on (quasi-) maximum likelihood methods for estimation purposes.

In this paper we adopt a Bayesian approach to estimate DCC models with skewed and heavy tailed distributions for the errors. We begin by describing a general tool for the univariate case which has the advantages of simplicity and clear interpretation of all parameters. Fernandez and Steel (1998) presented a general method for transforming any continuous unimodal and symmetric distribution into a skewed one by changing the scale at each side of the mode. They proposed the following class of skewed distributions indexed by a shape parameter $\gamma > 0$, which describes the degree of asymmetry,

$$s(x|\gamma) = \frac{2}{\gamma + 1/\gamma}\left\{ f\left(\frac{x}{\gamma}\right) I_{[0,\infty)}(x) + f(x\gamma) I_{(-\infty,0)}(x) \right\}, \qquad (5)$$

where $f(\cdot)$ is a univariate density symmetric around zero and $I_C(\cdot)$ is an indicator function on C. Note that $\gamma = 1$ yields the symmetric distribution as $s(x|\gamma = 1) = f(x)$, and values of $\gamma > 1$ (< 1) indicate right (left) skewness. Also, the mode of this density remains at zero irrespective of the particular value of γ. Mean and variance of $s(x|\gamma)$ depend on γ and are given by

$$\mu_\gamma = m_1(\gamma - 1/\gamma) \quad \text{and} \quad \sigma_\gamma^2 = (m_2 - m_1^2)(\gamma^2 + 1/\gamma^2) + 2m_1^2 - m_2$$

where

$$m_r = 2\int_0^\infty x^r f(x)dx,$$

is the r-th absolute moment of $f(x)$ on the positive real line. In this paper, $f(x)$ is a standardized distribution in which case $m_2 = 1$ so that mean and variance of the non-standardized skew density are given by $\mu_\gamma = m_1(\gamma - 1/\gamma)$ and $\sigma_\gamma^2 = (1 - m_1^2)(\gamma^2 + 1/\gamma^2) + 2m_1^2 - 1$. In particular, for the Student t distribution we obtain that

$$m_1 = \frac{2\sqrt{\nu-2}}{\nu-1}\,\frac{\Gamma(1/2)\,\Gamma(\nu/2)}{\Gamma((\nu+1)/2)}.$$

So, after some algebraic manipulations, the mean and the variance of the non-standardized skew t distribution are given by

$$\mu_\gamma = \frac{\Gamma((\nu-1)/2)\,\sqrt{\nu-2}\,(\gamma-1/\gamma)}{\sqrt{\pi}\,\Gamma(\nu/2)} \quad \text{and} \quad \sigma_\gamma^2 = (\gamma^2+1/\gamma^2) - \mu_\gamma^2 - 1. \quad (6)$$

Now, defining $y = (x - \mu_\gamma)/\sigma_\gamma$ then

$$p(y) = \frac{2\sigma_\gamma}{\gamma+1/\gamma}\left\{ f\left(\frac{\sigma_\gamma y + \mu_\gamma}{\gamma}\right) I\left(y \geq \frac{-\mu_\gamma}{\sigma_\gamma}\right) + f(\gamma(\sigma_\gamma y + \mu_\gamma)) I\left(y < \frac{-\mu_\gamma}{\sigma_\gamma}\right)\right\}$$

is its standardized version.

We note that this approach entirely separates the effects of the skewness and tail parameters thus making prior independence between the two a plausible assumption, and hence facilitates the choice of their prior distributions. Another important feature is that it retains the mode at zero. The parameter γ controls the allocation of mass to each side of the mode as it is easy to verify that $P(X_i \geq 0) = \gamma^2/(1+\gamma^2)$, which does not depend on $f(\cdot)$.

2.1 Multivariate Skew Densities

Bauwens and Laurent (2005) proposed to generalize the method described above to the multivariate case, i.e. to construct a multivariate skew distribution from a symmetric one. The idea is to use the same method of changing the scale on each side of the mode for each coordinate of the multivariate density. They show that this new class of multivariate skewed densities can be written as,

$$s(\boldsymbol{x}|\boldsymbol{\gamma}) = 2^k \left(\prod_{i=1}^k \frac{\gamma_i}{1+\gamma_i^2}\right) f(\boldsymbol{x}^*) \quad (7)$$

where $f(\cdot)$ is a symmetric multivariate density, $\boldsymbol{x}^* = (x_1^*, \ldots, x_k^*)$, $x_i^* = x_i/\gamma_i$ if $x_i \geq 0$ and $x_i^* = x_i\gamma_i$ if $x_i < 0$, $i = 1, \ldots, k$. It is clear that when $k = 1$ we recover the univariate skew density (5). Also, for each margin the interpretation of γ_i remains the same as in Fernandez and Steel (1998), i.e. $\gamma_i^2 = Pr(X_i \geq 0)/Pr(X_i < 0)$ and right (left) marginal skewness corresponds to $\gamma_i > 1$ ($\gamma_i < 1$). Note that the mode of (7) is still at the origin. Extending the univariate case, the existence of the moments of (7) depends only on the existence of marginal moments $E(X_i^r)$ and not on the particular values of $\gamma_1, \ldots, \gamma_k$.

In particular, for the multivariate t distribution with ν degrees of freedom, and given $\gamma_1, \ldots, \gamma_k$ we can compute the mean μ_{γ_i} and variance $\sigma^2_{\gamma_i}$ for each margin using (6). In this case, a vector with elements $(x_i^* - \mu_{\gamma_i})/\sigma_{\gamma_i}$ is its standardized version and density (7) can be written as,

$$s(\boldsymbol{x}|\boldsymbol{\gamma}) = 2^k \left(\prod_{i=1}^{k} \frac{\gamma_i \sigma_{\gamma_i}}{1+\gamma_i^2} \right) \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})[\pi(\nu-2)]^{k/2}} \left[1 + \frac{\boldsymbol{x}^{*\prime}\boldsymbol{x}^*}{\nu-2} \right]^{-\frac{\nu+k}{2}},$$

where $x_i^* = (x_i\sigma_{\gamma_i} + \mu_{\gamma_i})/\gamma_i$ if $x_i \geq -\mu_{\gamma_i}/\sigma_{\gamma_i}$ and $x_i^* = (x_i\sigma_{\gamma_i} + \mu_{\gamma_i})\gamma_i$ if $x_i < -\mu_{\gamma_i}/\sigma_{\gamma_i}$. This is then the standardized multivariate skew Student t density, which allows for a tail behaviour (common to all margins) heavier than would be implied by a multivariate skew normal distribution. Also, the standardized symmetric multivariate Student t density would result if $\gamma_i = 1$, $i = 1, \ldots, k$. Finally, $\nu \to \infty$ would be equivalent to assume that $f(\boldsymbol{x}^*)$ in (7) is a standard multivariate normal density, in which case we would obtain a standardized multivariate skew normal density.

As another heavy tailed multivariate distribution we consider the multivariate generalized error distribution (GED) also known as multivariate exponential power distribution. The probability density function of the standardized univariate GED with tail parameter $\delta > 0$ is given by,

$$p(x|\delta) = \left[\frac{\Gamma(3/\delta)}{\Gamma(1/\delta)}\right]^{1/2} \frac{1}{2\,\Gamma((\delta+1)/\delta)} \exp\left\{ - \left[\frac{\Gamma(3/\delta)}{\Gamma(1/\delta)} x^2\right]^{\delta/2} \right\} \qquad (8)$$

with kurtosis given by $\Gamma(1/\delta)\Gamma(5/\delta)/\Gamma(3/\delta)^2 - 3$ so that values $\delta < 2$ lead to leptokurtic distributions while the tails are thinner than the stardard normal if $\delta > 2$. The standard normal is obtained for $\delta = 2$. Extensions to the multivariate case were proposed in Gómez, Gómez-Villegas, and Marín (1998) but in this case the marginal distributions as well as the absolute moments are difficult to obtain. For this reason, we work with the joint distribution of k independent random variables so that the marginal densities are given by (8) with a common tail parameter δ. The joint density is then given by,

$$p(\boldsymbol{x}|\delta) = \left[\frac{\Gamma(3/\delta)}{\Gamma(1/\delta)}\right]^{k/2} \frac{1}{[2\ \Gamma((\delta+1)/\delta)]^k} \exp\left\{ - \left[\frac{\Gamma(3/\delta)}{\Gamma(1/\delta)}\right]^{\delta/2} \sum_{i=1}^{k} |x_i|^{\delta} \right\}. \quad (9)$$

Since density (8) is standardized it follows that $E(\boldsymbol{X}) = \mathbf{0}$ and $Var(\boldsymbol{X}) = \boldsymbol{I}_k$ and we can use the method proposed in Bauwens and Laurent (2005) to introduce asymmetry in the multivariate distribution. We refer to this multivariate distribution as $GED(\mathbf{0}, \boldsymbol{I}_k, \delta)$. It is not difficult to obtain the first absolute moment as $m_1 = \Gamma(2/\delta)/\sqrt{\Gamma(1/\delta)\Gamma(3/\delta)}$ so that the marginal means μ_{γ_i} and variances $\sigma^2_{\gamma_i}$ can be computed accordingly. Finally, the density of this standardized skewed version of the GED, denoted $SSGED(\mathbf{0}, \boldsymbol{I}_k, \boldsymbol{\gamma}, \delta)$, is given by

$$s(\boldsymbol{x}|\boldsymbol{\gamma}) = 2^k \left(\prod_{i=1}^{k} \frac{\gamma_i \sigma_{\gamma_i}}{1+\gamma_i^2} \right) \left[\frac{\Gamma(3/\delta)}{\Gamma(1/\delta)}\right]^{k/2} \frac{\exp\left\{ - \left[\frac{\Gamma(3/\delta)}{\Gamma(1/\delta)}\right]^{\delta/2} \sum_{i=1}^{k} |x_i^*|^{\delta} \right\}}{(2/\delta)^k\ [\Gamma(1/\delta)]^k},$$

where $x_i^* = (x_i\sigma_{\gamma_i} + \mu_{\gamma_i})/\gamma_i$ if $x_i \geq -\mu_{\gamma_i}/\sigma_{\gamma_i}$ and $x_i^* = (x_i\sigma_{\gamma_i} + \mu_{\gamma_i})\gamma_i$ if $x_i < -\mu_{\gamma_i}/\sigma_{\gamma_i}$.

3 Prior Distributions

Following the Bayesian paradigm, we need to complete the model specification by specifying the prior distributions of all parameters of interest. These are assumed to be a priori independent and normally distributed truncated to the intervals that define each one. By Bayes Theorem, the joint posterior density is proportional to the product of the likelihood function (4) by the joint prior density.

For the GARCH(1,1) coefficients in (2), we adopted the same prior distributions proposed in Ardia (2006) which are given by $\omega_i \sim N(\mu_{\omega_i}, \sigma^2_{\omega_i})I_{(\omega_i>0)}$, $\alpha_i \sim N(\mu_{\alpha_i}, \sigma^2_{\alpha_i})I_{(0<\alpha_i<1)}$ and $\beta_i \sim N(\mu_{\beta_i}, \sigma^2_{\beta_i})I_{(0<\beta_i<1)}$, $i = 1, \ldots, k$. A prior distribution for the tail parameter is assigned as $\nu \sim N(\mu_\nu, \sigma^2_\nu)I_{(\nu>2)}$ or $\delta \sim N(\mu_\delta, \sigma^2_\delta)I_{(\delta>0)}$ when using the multivariate Student t or GED respectively. Finally, a similar approach is adopted for the parameters α and β in equation

(3), i.e. $\alpha \sim N(\mu_\alpha, \sigma^2_\alpha) I_{(0<\alpha<1)}$ and $\beta \sim N(\mu_\beta, \sigma^2_\beta) I_{(0<\beta<1)}$. For these prior distributions, the values of the hyperparameters were kept fixed. It is worth mentioning that using a truncated normal distribution as prior facilitates the insertion of information in certain regions of the parameter space since the hyperparameters no longer represent the mean and variance but still control the region of higher probability mass. In the application in Section 4 we specified the hyperparameters as, $\mu_{\omega_i} = \mu_{\alpha_i} = \mu_{\beta_i} = \mu_\nu = \mu_\delta = \mu_\alpha = \mu_\beta = 0$ and $\sigma^2_{\omega_i} = \sigma^2_{\alpha_i} = \sigma^2_{\beta_i} = \sigma^2_\nu = \sigma^2_\delta = \sigma^2_\alpha = \sigma^2_\beta = 100$.

As for the skewness parameters we find it reasonable to choose a prior that is centered around the symmetric version of the skewed distribution and gives approximately equal weights to left and right skewness. Extending the univariate approach in Fernandez and Steel (1998), we shall assume that the components of the random vector $\boldsymbol{\gamma}$ are independent and use a Gamma(a, b) prior on each γ_i^2. If we choose the hyperparameters a and b such that $E(\gamma_i) = 1$ then it is not difficult to see that this implies choosing $b = [\Gamma(a+1/2)/\Gamma(a)]^2$ and a can be elicited by controlling the prior variance and prior mass of γ_i on the interval (0,1). Fernandez and Steel (1998) found that $a = 1/2$, which leads to $Var(\gamma_i) \approx 0.57$ and $P(0 < \gamma_i < 1) \approx 0.58$, is a reasonable choice. This is our default choice here too. We note that this particular choice is equivalent to setting $\gamma_i \sim N(0, 0.64^{-1})$ truncated to $\gamma_i > 0$ (i.e. a half-normal prior for γ_i).

Denoting the set of all unknown parameters by $\boldsymbol{\theta}$ the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is analytically intractable. Therefore, we adopt MCMC sampling strategies for obtaining samples from the joint posterior distributions. Since the full conditional posterior distributions are not of known form, Metropolis steps provide the easiest black-box sampling strategy to yield the required realizations of $\pi(\boldsymbol{\theta}|\boldsymbol{y})$.

4 Application

As an empirical example of our Bayesian procedure we consider the daily indices of stock markets in Frankfurt (DAX), Paris (CAC40) and Tokyo (Nikkei), from October 10, 1991 until December 30, 1997 (a total of 1627 days). These stock market data were analysed before in Franses and van Dijk (2000) and are freely available from the website http://www.robjhyndman.com/TSDL/data/FVD1.dat. If I_{it} is the value of the ith index at time t then we model the multivariate series of returns

$y_{it} = 100\log(\frac{I_{it}}{I_{i,t-1}})$.

For the MCMC updates we adopt a Metropolis-Hastings algorithm where all the parameters are updated as a block. In particular, we used a random walk Metropolis algorithm where at each iteration we generate new values from a multivariate normal distribution centered around the current value with a variance-covariance proposal matrix which was calculated from a pilot tunning procedure. This pilot tunning was carried out by running one-dimensional random walk Metropolis updates with univariate normal candidate distributions whose variances were calibrated to obtain good acceptance rates. We then ran a total of 200,000 iterations discarding the first 30,000 realizations as burn-in and thinning to every 5th iteration. Posterior results are then based on 34,000 realizations of the Markov chain with the prior distributions as explained in Section 3.

The simulated Markov chains were checked for convergence and good mixing by visual inspection of the marginal traces, density estimates, autocorrelations and formal tests. These convergence diagnostics did not indicate lack of convergence and the acceptance rates lied within the interval 0.20-0.50.

Figure 1 shows the estimated posterior densities of the skewness parameters γ_i for the multivariate skew normal, t and GED distributions. They clearly indicate skewness in the marginal distributions for the DAX and CAC40 indices while symmetry is more likely for the NIKKEI index.

The models for the error terms, including the symmetric version of the distribution, were then compared according to the deviance information criterion (DIC). This is given by $DIC = 2E[D(\boldsymbol{\theta}_M)] - D(E[\boldsymbol{\theta}_M])$ where $\boldsymbol{\theta}_M$ is the set of parameters in model M and $D(\cdot)$ is the deviance function defined as minus twice the log-likelihood function. So, given a sample from the posterior distribution of $\boldsymbol{\theta}_M$, it is straightforward to approximate the DIC. Table 1 shows the DIC values computed for the six multivariate GARCH models compared, assigning the error term with a multivariate normal, t, GED and their skewed counterparts as described in Section 2. As expected, we obtained lower values of DIC for the distributions with tails heavier than the normal and in particular the Skew t multivariate model is selected as the best one. We note however some similarity with the DIC for the model with symmetric multivariate t errors. Since the DIC is subject to Monte Carlo sampling error (it is a function of stochastically simulated quantities), this might cast some doubt whether the inclusion of skewness is substantially improving model fit. One way around this problem, is to use DIC weights obtained by subtracting from each DIC the value associated with the "best" model (DIC_B) and then setting

$$w_M \propto \exp(-(DIC_M - DIC_B)/2).$$

These weights are then normalized to sum to 1 over the models under consideration. This approach was first suggested in Burnham and Anderson (1998) Section 4.2 for the Akaike's Information Criterion (AIC) where the differences are interpreted as the strength of evidence. The approach can be extended to be used with the DIC which, as pointed out in Spiegelhalter, Best, Carlin, and van der Linde (2002), can be viewed as a Bayesian analogue of AIC. The normalized weights are shown in the second column of Table 1 and model comparison is much easier in this transformed scale. We can see that the Skew t multivariate model is actually very far from all other models which receive very small weights. We then proceed showing the results for this model only.

Table 1: DIC values for the six multivariate GARCH models compared.

Model	DIC	weights
Normal	13957.53	0.0000
Student	13810.36	0.0287
GED	13828.97	0.0000
Skew Normal	13947.62	0.0000
Skew t	13803.32	0.9712
skew GED	13823.48	0.0000

A summary of the MCMC simulations for the model with multivariate Skew t errors is shown in Table 2. Posterior means, medians and 95% credible intervals of the skewness parameters indicate high asymmetry for the DAX returns while for the CAC40 returns there is a slight skewness to the right. The NIKKEI returns do not show asymmetry. Also, the estimate of the tail parameter ν (last row in Table 2) indicates that a fat tail distribution is appropriate for the error term. Finally, it is of practical interest to check whether the correlations change over time thus justifying the additional complexity introduced by the DCC model. The estimates in the bottom of Table

2 indicate that a constant conditional correlation (CCC) model hypothesis ($\alpha = \beta = 0$) can be rejected on the basis of the marginal posterior distributions of α and β, although the posterior distribution of $\alpha + \beta$ does not indicate a too strong persistence in equation (3).

Table 2: Summary of the MCMC simulations for the model with multivariate Skew t errors.

		Mean	SD	2.5%	50%	97.5%
DAX	γ_1	0.8989	0.0292	0.8425	0.8985	0.9571
	w_1	0.0276	0.0093	0.0129	0.0265	0.0488
	a_1	0.0675	0.0138	0.0439	0.0662	0.0978
	b_1	0.9071	0.0198	0.8634	0.9088	0.9405
	$a_1 + b_1$	0.9746	0.0108	0.9505	0.9757	0.9922
CAC40	γ_2	1.0450	0.0336	0.9803	1.0443	1.1119
	w_2	0.0442	0.0166	0.0225	0.0409	0.0848
	a_2	0.0392	0.0096	0.0230	0.0383	0.0606
	b_2	0.9270	0.0187	0.8827	0.9299	0.9539
	$a_2 + b_2$	0.9662	0.0140	0.9324	0.9687	0.9850
NIKKEI	γ_3	1.0079	0.0308	0.9486	1.0072	1.0691
	w_3	0.0378	0.0127	0.0173	0.0364	0.0667
	a_3	0.0853	0.0138	0.0614	0.0842	0.1154
	b_3	0.8997	0.0162	0.8639	0.9009	0.9275
	$a_3 + b_3$	0.9849	0.0087	0.9661	0.9856	0.9998
	α	0.0433	0.0148	0.0163	0.0426	0.0741
	β	0.6495	0.1426	0.2938	0.6747	0.8548
	$\alpha + \beta$	0.6928	0.1385	0.3365	0.7199	0.8834
	ν	8.1533	0.8477	6.6878	8.0850	9.9988

In order to get some assessment on the robustness of the results in Table 2 the hyperparameter σ_ν^2 for the prior on the degrees of freedom was changed to 50 and 10. The results obtained were similar to those reported above.

5 Concluding Remarks

In this paper we propose to extend skewed distributions which have appeared recently in the univariate GARCH models literature to the multivariate case and describe how to carry out Bayesian inference. They form a flexible family of distributions that can accommodate asymmetric data with heavy tails, which are frequently observed in financial time series.

Our preference for the skewing mechanism described in this paper over others that appeared in the recent literature is mainly due to its simplicity and generality. In particular, moments calculation is straightforward if the moments of the underlying univariate symmetric distributions are available and it does not require calculation of cumulative distribution functions, which yields faster computations.

This is an ongoing work and we are still investigating the use of skewed distributions with fat tails other than the multivariate Student t and GED. For example, the multivariate slash distribution also has a tail parameter and the multivariate normal as limiting case. Another interesting extension which is under development is the possibility of estimating a different tail parameter for each dimension and check for the influence of potential outliers in the model choice. Model comparison in terms of approximations for the Bayes Factors and/or information criteria based on MCMC output are also under investigation.

Acknowledgements

The work of the first author was funded by CNPq - Brazil. The work of the second author was supported by FAPESP - Brazil, under grant number 2011/22317-0.

References

Alexander, C. (2001). *A Practitioners Guide to Financial Data Analysis*. J. Wiley & Sons.

Ardia, D. (2006). Bayesian estimation of the GARCH(1,1) model with normal innovations. *Student* *5*(3–4), 283–298. URL http://ssrn.com/abstract=1543409.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*, 171–178.

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbations of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society B 65*, 367–389.

Baillie, R. and T. Bollerslev (1989). The message in daily exchange rates: A conditional-variance tale. *Journal of Business and Economic Statistics 7*(3), 297–305.

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*, 346–354.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics 21*(1), 79–109.

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

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy 96*, 116–131.

Branco, M. and D. Dey (2001). A class of multivariate skew-elliptical distribution. *Journal of Multivariate Analysis 79*, 79–113.

Burnham, K. P. and D. R. Anderson (1998).

Cappuccio, N., D. Lubian, and R. Raggi (2004). MCMC Bayesian estimation of a skew-GED stochastic volatility model. *Studies in Nonlinear Dynamics & Econometrics 8*(2).

Christodoulakis, G. A. and S. E. Satchell (2002). Correlated ARCH: Modelling the time-varying correlation between financial asset returns. *European Journal of Operations Research 139*, 351–370.

Engle, R. F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics 20*, 339–350.

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

Fernandez, C. and M. Steel (1998). On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association 93*, 359–371.

Fiorentini, G., E. Sentana, and G. Calzolari (2003). Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with student t innovations. *Journal of Business and Economic Statistics 21*, 532–546.

Forsberg, L. and T. Bollerslev (2002). Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the Euro): the GARCH-NIG model. *Journal of Applied Econometrics 17*(5), 535–548.

Franses, P. H. and D. van Dijk (2000). *Non-linear time series models in empirical finance*. Cambridge University Press: Cambridge.

Genton, M. G. (Ed.) (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. CRC, Chapman & Hall, London.

Gómez, E., M. A. Gómez-Villegas, and J. M. Marín (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics - Theory and Methods 27*(3), 589–600.

Gourieroux, C. (1997). *ARCH Models and Financial Applications*. Springer Series in Statistics.

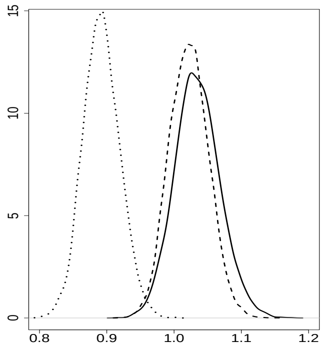
Jones, M. C. and M. J. Faddy (2003). A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society B 65*, 159–174.

Kotz, S. and S. Nadarajah (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press, Cambridge.

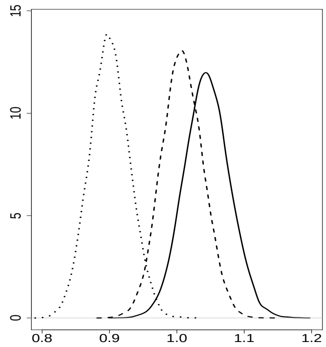
Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B 64*, 1–34.

Tse, Y. K. and A. K. C. Tsui (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics 20*, 351–362.

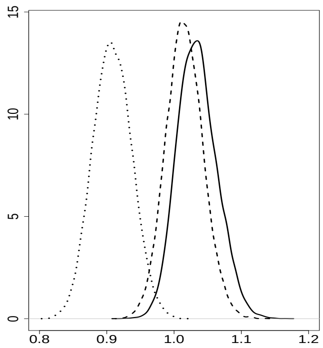
Figure 1: Posterior densities of skewness parameters for the DAX (dotted), CAC40 (full) and NIKKEI (dashed) indices using the skew multivariate normal, t and GED distributions.



(a) Skew Normal



(b) Skew t



(c) Skew GED