UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 14-17
Statistics and Econometrics Series 11
June 2014

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

BAYESIAN ESTIMATION OF A DYNAMIC CONDITIONAL CORRELATION MODEL WITH MULTIVARIATE SKEW-SLASH INNOVATIONS

Cristina G. de la Fuente, Pedro Galeano, and Michael P. Wiper*

Abstract

Financial returns often present a complex relation with previous observations, along with a slight skewness and high kurtosis. As a consequence, we must pursue the use of flexible models that are able to seize these special features: a financial process that can expose the intertemporal relation between observations, together with a distribution that can capture asymmetry and heavy tails simultaneously. A multivariate extension of the GARCH such as the Dynamic Conditional Correlation model with Skew-Slash innovations for financial time series in a Bayesian framework is proposed in the present document, and it is illustrated using an MCMC within Gibbs algorithm performed on simulated data, as well as real data drawn from the daily closing prices of the DAX, CAC40, and Nikkei indices.

Keywords: Bayesian inference; Dynamic Conditional Correlation; Financial time series; Infinite mixture; Kurtosis; MCMC; Skew-Slash.

* Department of Statistics, Universidad Carlos III de Madrid, Spain. E-mail addresses: cgfuente@est-econ.uc3m.es (Cristina G. de la Fuente), pedro.galeano@uc3m.es (Pedro Galeano), and mwiper@est-econ.uc3m.es (Michael P. Wiper). **Acknowledgements**: We acknowledge financial support by Spanish Ministry of Economy and Competition Grant ECO2012-38442.

Bayesian Estimation of Dynamic Conditional Correlation Models With Skew-Slash Innovations

Cristina G. de la Fuente Pedro Galeano

Michael P. Wiper

Department of Statistics

Universidad Carlos III de Madrid

Abstract

Financial returns often present a complex relation with previous observations, along with a slight skewness and high kurtosis. As a consequence, we must pursue the use of flexible models that are able to seize these special features: a financial process that can expose the intertemporal relation between observations, together with a distribution that can capture asymmetry and heavy tails simultaneously. A multivariate extension of the GARCH such as the Dynamic Conditional Correlation model with Skew-Slash innovations for financial time series in a Bayesian framework is proposed in the present document, and it is illustrated using an MCMC within Gibbs algorithm performed on simulated data, as well as real data drawn from the daily closing prices of the DAX, CAC40, and Nikkei indices.

Keywords: Bayesian inference; Dynamic Conditional Correlation; Financial time series; Infinite mixture; Kurtosis; MCMC; Skew-Slash.

1 Introduction

Working with financial returns often presents the challenge of having to model data with a complex relation to previous observations. This kind of data usually reflects a heteroskedastic behavior, that could be reasonably dealt with by means of a Generalized Autoregressive Conditional Heteroskedastic (GARCH) process, introduced by Bollerslev (1986).

Financial data sets have been widely studied assuming Gaussianity for their stochastic component; nevertheless, this kind of data usually presents a slight skewness and high kurtosis. As a consequence, the normal distribution is not able to properly perform its modeling task. Instead, we should pursue the use of more flexible models that allow us to capture the characteristics of the studied data more adequately.

A number of alternatives to Gaussian innovations have been proposed. For the univariate case, Bollerslev (1987) proposed to capture the high kurtosis in the innovations with a Student's-t distribution and, with the same objective, Nelson (1991) considered the Generalized Error Distribution (GED), and Bai, Russell, and Tiao (2003) assumed a mixture of two zero mean Gaussian distributions. For the multivariate case, Galeano and Ausín (2010) present a finite mixture of Normal distributions. Extensions have also been developed to capture skewness such as the Skew-Normal, the Skew-t, and the Skew-GED distributions of Fernández and Steel (1998), among others. We can also find in Fioruci, Ehlers, and Andrade (2014) the study of multidimensional financial returns with Skew-t innovations.

Nevertheless, it still has not been possible to show the existence of a single parametric distribution that adequately describes the behavior of financial returns in all situations.

A couple of suitable options capable to capture both the skewness and high kurtosis could be the Skew-t distribution developed by Jones and Faddy (2003), as proposed by Fioruci, Ehlers, and Andrade (2014), or the Skew-Slash distribution by

Wang and Genton (2006), among others. Note that the Skew-Slash distribution has already been applied successfully to describe data with similar characteristics; one example can be found in Lachos, Garibay, Labra, and Aoki (2009 a).

We must keep in mind that the Skew-Slash distribution is an infinite mixture of Skew-Normal distributions and, as a consequence, its probability density function (pdf) presents a complicated form that would make it very difficult to perform Maximum Likelihood, via either constrained optimization or the EM algorithm, appropriately in the financial framework, while Bayesian inference is more powerful and is able to undertake the problem we want to present. Additionally, the Bayesian framework naturally provides the possibility to take into account the intrinsic uncertainty presented by the correlations and volatilities of the assets under study, as well as the one to incorporate expert information, when available. Hence, our proposal for the multivariate case is to perform inference for multidimensional financial returns by means of a Dynamic Conditional Correlation model with Skew-Slash innovations, from a Bayesian approach.

The rest of the paper is structured as follows. In Section 2, the multivariate Skew-Slash distribution is introduced, and we give closed forms for its mean, variance, and kurtosis coefficient. In Section 3, we present the Dynamic Conditional Correlation model with Skew-Slash innovations. The methodology implemented to perform the Bayesian estimation of the proposed model is detailed in Section 4. In section 5, we illustrate the framework presented by estimating the proposed model for simulated data and, afterwards, we carry out the pertinent estimations for real data conformed by the log-returns of the DAX, CAC40, and Nikkei indices. Finally, the paper ends with some conclusions and we expose some extensions that would be desirable in future research in section 6.

2 The multivariate Skew-Slash distribution

This section introduces several distributions that are required for later developments and that help us achieve a better understanding of the Skew-Slash distribution.

2.1 Normal/independent family of distributions

We say that a d-dimensional random vector, $\mathbf{V}$, follows a Normal/independent distribution, as defined by Lange and Sinsheimer (1993), with location parameter $\boldsymbol{\eta}$ and scale matrix $\boldsymbol{\Sigma}$ if its pdf is given by

$$f_{\mathbf{V}}(\mathbf{v}) = \int_{\mathbb{R}^+} \phi_d\left(\mathbf{v}|\boldsymbol{\eta}, u^{-2}\boldsymbol{\Sigma}\right) dH(u); \mathbf{v} \in \mathbb{R}^d,$$

where $\boldsymbol{\eta} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is a symmetric positive definite matrix, and $H(u|\nu)$ is the cumulative distribution function (cdf) of a unidimensional positive random variable U, indexed by the parameter ν, and we denote it as $\mathbf{V} \sim \mathcal{NI}_d(\boldsymbol{\eta}, \boldsymbol{\Sigma}; H)$. Also, $\phi_d(\cdot|\boldsymbol{\eta}, \boldsymbol{\Sigma})$ denotes the d-dimensional multivariate normal density with mean $\boldsymbol{\eta}$ and covariance matrix $\boldsymbol{\Sigma}$.

Its alternative stochastic representation is given by

$$\mathbf{V} \equiv \boldsymbol{\eta} + U^{-1}\mathbf{X},$$

where $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$. Also, U and $\mathbf{X}$ are independent. We assume that $\equiv$ denotes equivalence in distribution.

Let us notice that, if $\mathbf{V} \sim \mathcal{NI}_d(\boldsymbol{\eta}, \boldsymbol{\Sigma}; H)$, then we can find general expressions for its mean, given by

$$E(\mathbf{V}) = \boldsymbol{\eta},$$

and for its variance-covariance matrix, given by

$$V(\mathbf{V}) = E\left(U^{-2}\right)\boldsymbol{\Sigma},$$

if $E(U^{-2}) < \infty$.

The Normal/independent family of distributions consists of symmetric distributions that allow for heavier tails than the Gaussian one. It includes, among others, the Student's-t, the Slash, the Power Exponential, and the Contaminated-Normal distributions, all of which have heavier tails than the Gaussian distribution.

2.2 Skew-Normal distribution

According to Azzalini and Dalla Valle (1996), we say that a d-dimensional random vector, $\mathbf{Z}$, follows a Skew-Normal distribution with location parameter $\boldsymbol{\eta}$, scale matrix $\boldsymbol{\Sigma}$, and skewness parameter $\boldsymbol{\lambda}$, denoted as $\mathbf{Z} \sim \mathcal{SN}_d\left(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}\right)$, if its pdf is given by

$$f_{\mathbf{Z}}\left(\mathbf{z}\right) = 2\phi_d\left(\mathbf{z}|\boldsymbol{\eta}, \boldsymbol{\Sigma}\right)\Phi_1\left(\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1/2}\left(\mathbf{z}-\boldsymbol{\eta}\right)\right); \mathbf{z} \in \mathbb{R}^d, \tag{1}$$

where $\boldsymbol{\eta} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is a symmetric positive definite matrix, and $\boldsymbol{\lambda} \in \mathbb{R}^d$. Also, $\Phi_1\left(\cdot\right)$ denotes the cdf of the univariate standard normal distribution.

Let us notice that, when $\boldsymbol{\lambda} = \mathbf{0}$, the expression in (1) reduces to the normal density. Also, it is useful to take into account the stochastic representation for Z, proposed by Azzalini and Dalla Valle (1996), given by

$$\mathbf{Z} \equiv \boldsymbol{\eta} + \boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{\delta}\left|X_0\right| + \left(\mathbf{I}-\boldsymbol{\delta}\boldsymbol{\delta}'\right)^{1/2}\mathbf{X}_1\right), \tag{2}$$

where $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{\lambda}}$, which implies that $\left\|\boldsymbol{\delta}\right\|_2 < 1$, and $X_0 \sim \mathcal{N}_1\left(0,1\right)$ independent from $\mathbf{X}_1 \sim \mathcal{N}_d\left(\mathbf{0},\mathbf{I}\right)$.

Let us also notice that, if $\mathbf{Z} \sim \mathcal{SN}_d\left(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}\right)$, then its mean is given by

$$E\left(\mathbf{Z}\right) = \boldsymbol{\eta} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta},$$

and its variance-covariance matrix is given by

$$V\left(\mathbf{Z}\right) = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{1/2}.$$

We may say that this is a more flexible distribution than the Gaussian one because it allows for the presence of skewness.

2.3 Skew-Slash distribution

According to the definition given by Wang and Genton (2006), we say that, if $\mathbf{Z} \sim \mathcal{SN}_d(\boldsymbol{\lambda})$ is a standard Skew-Normal random vector, independent from $U \sim \mathcal{B}e(\nu, 1)$, then let us define

$$\mathbf{W} \equiv \boldsymbol{\eta} + U^{-1}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}. \tag{3}$$

In this case, we say that $\mathbf{W}$ follows a d-dimensional Skew-Slash distribution with location parameter $\boldsymbol{\eta}$, scale matrix $\boldsymbol{\Sigma}$, skewness parameter $\boldsymbol{\lambda}$, and kurtosis parameter ν, and it will be denoted as $\mathbf{W} \sim \mathcal{SSL}_d(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$.

The pdf of a Skew-Slash random vector is given by

$$f_{\mathbf{W}}(\mathbf{w}) = \int_0^1 2\nu u^{\nu-1}\phi_d\left(\mathbf{w}|\boldsymbol{\eta}, u^{-2}\boldsymbol{\Sigma}\right)\Phi_1\left(u\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1/2}(\mathbf{w}-\boldsymbol{\eta})\right)du; \mathbf{w} \in \mathbb{R}^d,$$

where $\boldsymbol{\eta} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is a symmetric positive definite matrix, $\boldsymbol{\lambda} \in \mathbb{R}^d$, and $\nu > 0$.

From this expression, we can see that the d-dimensional Skew-Slash distribution is a scale-mixture of a Skew-Normal distribution. This means that we are allowing the possibility of different variances for different members of the population.

Given the definition of the Skew-Slash random vector in (3), and taking into account the alternative stochastic representation for a random vector that follows a Skew-Normal distribution given in (2), we get a more elaborate alternative stochastic representation for the Skew-Slash distribution

$$\mathbf{W} \equiv \boldsymbol{\eta} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}X + U^{-1}\left[\boldsymbol{\Sigma}^{1/2}(\mathbf{I}-\boldsymbol{\delta}\boldsymbol{\delta}')\boldsymbol{\Sigma}^{1/2}\right]^{1/2}\mathbf{X}_1, \tag{4}$$

where

$$\begin{aligned}
\boldsymbol{\delta} &= \frac{1}{\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{\lambda}}}\boldsymbol{\lambda} \\
X &= U^{-1}|X_0|; X_0 \sim \mathcal{N}_1(0,1) \\
\mathbf{X}_1 &\sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}); \mathbf{X}_1 \perp X_0 \\
U &\sim \mathcal{B}e(\nu, 1).
\end{aligned}$$

Let us notice that, with this definition, we have that $\|\boldsymbol{\delta}\|_2 < 1$.

This stochastic representation is very useful for Bayesian inference; it is also useful to acknowledge some of the central moments of the Skew-Slash distribution, and it makes simulation very easy to execute.

In this case, we are able to provide closed expressions for the mean vector, the variance-covariance matrix, and the kurtosis coefficient; nevertheless, the skewness coefficient is intractable for the Skew-Slash distribution.

If the random vector $\mathbf{W} \in \mathbb{R}^d$ follows a Skew-Slash distribution, as defined in (3), according to Wang and Genton (2006), its expectation is given by

$$E\left(\mathbf{W}\right) = \boldsymbol{\eta} + \sqrt{\frac{2}{\pi}}\frac{\nu}{\nu-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}, \text{ for } \nu > 1, \tag{5}$$

and its variance-covariance matrix is given by

$$V\left(\mathbf{W}\right) = \boldsymbol{\Sigma}^{1/2}\left\{\frac{\nu}{\nu-2}\mathbf{I} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}\boldsymbol{\delta}'\right\}\boldsymbol{\Sigma}^{1/2}, \text{ for } \nu > 2. \tag{6}$$

Finally, we can give a closed expression for the kurtosis coefficient using Proposition 7 from Lachos, Labra, and Ghosh (2007), defined as

$$\gamma_2\left(W\right) = E\left[\left\{\left(\mathbf{W} - \boldsymbol{\mu}_W\right)'\boldsymbol{\Sigma}_W^{-1}\left(\mathbf{W} - \boldsymbol{\mu}_W\right)\right\}^2\right],$$

as established by Mardia (1974), that represents the extension of the kurtosis coefficient proposed by Azzalini and Capitanio (1999) for the Skew-Normal distribution. In this case, the kurtosis coefficient is given by

$$\gamma_2\left(W\right) = \frac{\left(\nu-2\right)^2}{\nu\left(\nu-4\right)}a_1 - 4\frac{\left(\nu-2\right)^2}{\nu\left(\nu-3\right)}a_2 + a_3 - d\left(d+2\right), \text{ for } \nu > 4,$$

where $\boldsymbol{\mu}_\mathbf{W} = E\left(\mathbf{W} - \boldsymbol{\eta}\right) = \sqrt{\frac{2}{\pi}}\frac{\nu}{\nu-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$, $\boldsymbol{\Sigma}_\mathbf{W} = V\left(\mathbf{W}\right)$, and

$$\begin{aligned}
a_1 &= d\left(d+2\right) + 2\left(d+2\right)\boldsymbol{\mu}_\mathbf{W}'\boldsymbol{\Sigma}_\mathbf{W}^{-1}\boldsymbol{\mu}_\mathbf{W} + 3\left(\boldsymbol{\mu}_\mathbf{W}'\boldsymbol{\Sigma}_\mathbf{W}^{-1}\boldsymbol{\mu}_\mathbf{W}\right)^2, \\
a_2 &= \left(d + \frac{2\left(\nu-1\right)}{\nu}\right)\boldsymbol{\mu}_\mathbf{W}'\boldsymbol{\Sigma}_\mathbf{W}^{-1}\boldsymbol{\mu}_\mathbf{W} \\
&\quad + \left(1 + \frac{2\left(\nu-1\right)}{\nu} - \frac{\pi}{2}\frac{\left(\nu-1\right)^2}{\nu\left(\nu-2\right)}\right)\left(\boldsymbol{\mu}_\mathbf{W}'\boldsymbol{\Sigma}_\mathbf{W}^{-1}\boldsymbol{\mu}_\mathbf{W}\right)^2, \\
a_3 &= 2\left(d+2\right)\boldsymbol{\mu}_\mathbf{W}'\boldsymbol{\Sigma}_\mathbf{W}^{-1}\boldsymbol{\mu}_\mathbf{W} + 3\left(\boldsymbol{\mu}_\mathbf{W}'\boldsymbol{\Sigma}_\mathbf{W}^{-1}\boldsymbol{\mu}_\mathbf{W}\right)^2.
\end{aligned}$$

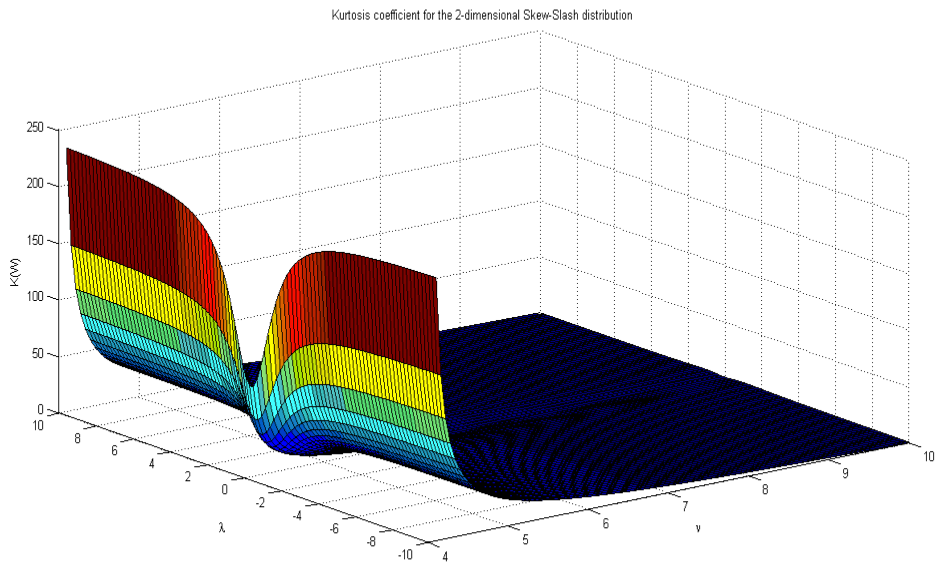


Figure 1: Kurtosis coefficient of the 2-dimensional Skew-Slash distribution as a function of λ and ν

To illustrate the ability of the Skew-Slash distribution to generate high kurtosis, Figure 1 shows some values of the kurtosis coefficient for a 2-dimensional Skew-Slash random vector and, analogously, Figure 2 does the same for a 3-dimensional Skew-Slash random vector. In both cases, we set $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ to satisfy $E(\mathbf{W}) = \mathbf{0}$ and $V(\mathbf{W}) = \mathbf{I}$. In order to be able to plot the surface, we set $\boldsymbol{\lambda}$ to be proportional to a vector of ones. That way, we can plot the size of the elements in $\boldsymbol{\lambda}$ and ν against the corresponding kurtosis coefficient.

It can be seen that the kurtosis gets higher as ν decreases. Besides, the size of the elements in the illustrated $\boldsymbol{\lambda}$ vectors has only a small effect on the kurtosis. We are also able to notice that, as the dimension gets higher, the kurtosis allowed for the same parameter values gets higher as well.

Let us notice that the Skew-Slash distribution, as well as the Skew-t distribution, defined by Azzalini and Capitanio (2003), among others, are particular members of the Skew-Normal/independent family of distributions, defined by Lachos, Labra and Ghosh (2007).

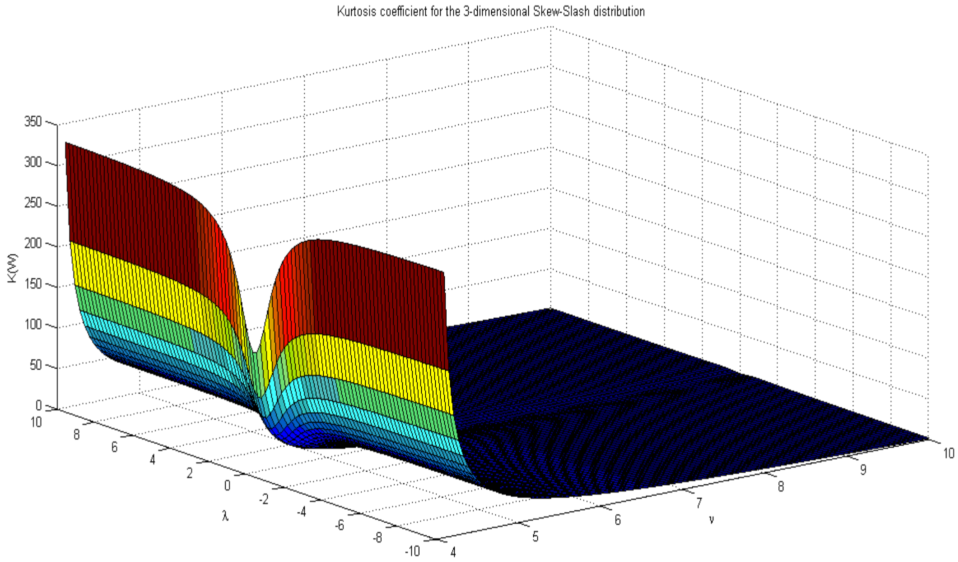


Figure 2: Kurtosis coefficient of the 3-dimensional Skew-Slash distribution as a function of λ and ν

3 Dynamic Conditional Correlation model with Skew-Slash innovations

A popular model used to describe the behavior of financial time series is the Generalized Autoregressive Conditional Heteroskedastic (GARCH) process. The issue with the construction of a conditional heteroskedastic model in the multivariate case is that it must be flexible enough to be able to capture the joint behavior of the assets, but it cannot be too highly parameterized.

To respond to the problem we pose, we decide to consider the Dynamic Conditional Correlation model introduced by Tse and Tsui (2002). In the way we present this model, we find that its structure is flexible enough to perform well with the data sets that we are interested in, while having a reasonable amount of parameters. Also, this model has already been successfully used to work with financial returns by Galeano and Ausín (2010) using a finite mixture of normal distributions for the innovations.

Several models for multiple returns have been proposed. Among others, Bollerslev (1990), Jeantheau (1998), Engle and Sheppard (2001), Engle (2002), Cappiello, Engle, and Sheppard (2006), Billio, Caporin, and Gobbo (2006), and Aielli (2013) propose some of the most popular approaches in the class of Dynamic Conditional Correlation models.

The Skew-Slash distribution has much heavier tails than the Skew-Normal distribution, as Wang and Genton (2006) explain in their paper. Also, we have identified the analytical expressions of the mean vector and variance-covariance matrix for the Skew-Slash distribution, which allows us to establish the restrictions that the Dynamic Conditional Correlation model subjects us to.

Taking into account the mentioned features presented by financial returns, as well as the power offered by the Skew-Slash distribution, we propose a dynamic conditional correlation model with d-dimensional Skew-Slash innovations, defined as

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{H}_t^{1/2}\boldsymbol{\varepsilon}_t, \tag{7}$$

where $\boldsymbol{\mu}$ is the unconditional mean of the process, $\mathbf{H}_t \in \mathbb{R}^{d\times d}$ is the conditional covariance matrix of $\mathbf{y}_t$, given the past observations $\{\mathbf{y}_{t-1}, \ldots, \mathbf{y}_0\}$, and $\boldsymbol{\varepsilon}_t$ is the innovation at time t.

We specify

$$\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t, \tag{8}$$

where

$$\mathbf{D}_t = diag\left(\left\{h_{it}^{1/2}\right\}_{i=1}^{d}\right) \in \mathbb{R}^{d\times d} \tag{9}$$

is a diagonal matrix that contains the d conditional standard deviations, denoted as $h_{it}^{1/2}$ for $i = 1, \ldots, d$, and $\mathbf{R}_t \in \mathbb{R}^{d\times d}$ is the matrix of conditional correlations. Let us notice that $\mathbf{H}_t$ is a symmetric positive definite matrix if and only if $\mathbf{D}_t$ has a positive diagonal and $\mathbf{R}_t$ is symmetric positive definite itself. We also define

$$h_{it} = \omega_i + \alpha_i\left(y_{t-1,i} - \mu_i\right)^2 + \beta_i h_{t-1,i}, \tag{10}$$

$$\mathbf{R}_t = \left(1 - \theta_1 - \theta_2\right)\mathbf{R} + \theta_1\mathbf{R}_{t-1} + \theta_2\mathbf{I}, \tag{11}$$

where $\mathbf{R}$ is a symmetric positive definite correlation matrix with unit diagonal elements, and off-diagonal elements denoted by ρ_{ij}. Further, we assume that $\mathbf{y}_0, \mathbf{h}_0 \in \mathbb{R}^d$ are known constants, which is not a restrictive assumption from a practical point of view, because financial data sets usually present elevated sample sizes. We also take $\alpha_i, \beta_i > 0$ and $\alpha_i + \beta_i < 1$, for all $i \in \{1, \ldots, d\}$ to ensure positivity of $\mathbf{h}_t$ and covariance stationarity; besides, $\theta_1, \theta_2 > 0$ and $\theta_1 + \theta_2 < 1$. Let us remark that, under this structure, $\mathbf{R}_t$ is a symmetric positive definite matrix and $\mathbf{D}_t$ has positive diagonal elements; hence, $\mathbf{H}_t$ is indeed a symmetric positive definite matrix.

Finally, we will say that the innovations are independent identically distributed random vectors, and they follow a d-dimensional Skew-Slash distribution such that

$$\boldsymbol{\varepsilon}_t \sim \mathcal{SSL}_d\left(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu\right),$$

with $E\left(\boldsymbol{\varepsilon}_t\right) = \mathbf{0}$ and $V\left(\boldsymbol{\varepsilon}_t\right) = \mathbf{I}$ for all $t \in \{1, \ldots, T\}$, where T denotes the size of our time series.

4 Bayesian Inference

First of all, we must acknowledge the restrictions intrinsic to this model. We have already stated that the innovations will be modeled through a Skew-Slash distribution, and we have established that $E\left(\boldsymbol{\varepsilon}_t\right) = \mathbf{0}$ and $V\left(\boldsymbol{\varepsilon}_t\right) = \mathbf{I}$. Incorporating this restriction to (5) and (6), we have that $\boldsymbol{\eta} = -\sqrt{\frac{2}{\pi}\frac{\nu}{\nu-1}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1} = \frac{\nu}{\nu-2}\mathbf{I} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}\boldsymbol{\delta}'$. Therefore,

$$\boldsymbol{\eta} = -\sqrt{\frac{2}{\pi}\frac{\nu}{\nu-1}}\left\{\frac{\nu}{\nu-2}\mathbf{I} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}\boldsymbol{\delta}'\right\}^{-1/2}\boldsymbol{\delta} \tag{12}$$

and

$$\boldsymbol{\Sigma} = \left\{\frac{\nu}{\nu-2}\mathbf{I} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}\boldsymbol{\delta}'\right\}^{-1}. \tag{13}$$

This means that $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ will depend of $\boldsymbol{\delta}$ and ν.

Considering that

$$\boldsymbol{\Sigma}^{-1} = \frac{\nu}{\nu-2}\mathbf{I} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}\boldsymbol{\delta}',$$

let us notice that we can easily find a closed form for its inverse, given by

$$\boldsymbol{\Sigma} = \frac{\nu - 2}{\nu}\mathbf{I} - \frac{2\left(\nu - 2\right)^2}{\pi\left(\nu - 1\right)^2 - 2\nu\left(\nu - 2\right)\boldsymbol{\delta}'\boldsymbol{\delta}}\boldsymbol{\delta}\boldsymbol{\delta}'.$$

Let us denote $\mathbf{M}_D = \mathbf{I} - \boldsymbol{\delta}\boldsymbol{\delta}'$, with inverse $\mathbf{M}_D^{-1} = \mathbf{I} + \frac{1}{1-\boldsymbol{\delta}'\boldsymbol{\delta}}\boldsymbol{\delta}\boldsymbol{\delta}'$, expressions that will allow for a more compact notation in the expressions to come.

Introducing the parameter restrictions found in (12) and (13) into the alternative stochastic representation for the Skew-Slash distribution presented in (4), we can express the innovations alternatively as

$$\boldsymbol{\varepsilon}_t \equiv -\sqrt{\frac{2}{\pi}\frac{\nu}{\nu - 1}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}X_t + U_t^{-1}\left[\boldsymbol{\Sigma}^{1/2}\left(\mathbf{I} - \boldsymbol{\delta}\boldsymbol{\delta}'\right)\boldsymbol{\Sigma}^{1/2}\right]^{1/2}\mathbf{X}_{1t},$$

where

$$\begin{aligned}
&\boldsymbol{\delta} = \frac{1}{\sqrt{1 + \boldsymbol{\lambda}'\boldsymbol{\lambda}}}\boldsymbol{\lambda},\\
&X_t = U_t^{-1}\left|X_{0t}\right|; X_{0t} \sim \mathcal{N}_1\left(0, 1\right),\\
&\mathbf{X}_{1t} \sim \mathcal{N}_d\left(\mathbf{0}, \mathbf{I}\right); \mathbf{X}_{1t} \perp X_{0t},\\
&U_t \sim \mathcal{B}e\left(\nu, 1\right),\\
&\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\delta}, \nu).
\end{aligned}$$

It is useful to take into account the fact that, because $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$, we can define $\mathbf{H}_t^{1/2}$, such that $\mathbf{H}_t = \mathbf{H}_t^{1/2}\mathbf{H}_t^{1/2}$, as $\mathbf{H}_t^{1/2} = \mathbf{D}_t^{1/2}\mathbf{R}_t^{1/2}\mathbf{D}_t^{1/2}$, where $\mathbf{D}_t^{1/2} = diag\left(\left\{h_{it}^{1/4}\right\}_{i=1}^d\right)$ and $\mathbf{R} = \mathbf{R}^{1/2}\mathbf{R}^{1/2}$. Let us also define a new matrix $\tilde{\mathbf{D}}_t = \mathbf{D}_t^{-1/2} = diag\left(\left\{h_{it}^{-1/4}\right\}_{i=1}^d\right)$. Now, we can define $\mathbf{H}_t^{-1/2} = \tilde{\mathbf{D}}_t\mathbf{R}_t^{-1/2}\tilde{\mathbf{D}}_t$. Finally, from now on we consider $a_{xt} = x_t - \sqrt{\frac{2}{\pi}\frac{\nu}{\nu-1}}$, also for compacting purposes.

Let us assume that we have observed a series of returns $\{\mathbf{y}_t\}_{t=1}^T$. Then, under the proposed model, the likelihood function is given by

$$\begin{aligned}
&f\left(\{\mathbf{y}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T | \boldsymbol{\delta}, \nu, \boldsymbol{\mu}, \{\omega_i, \alpha_i, \beta_i\}_{i=1}^d, \theta_1, \theta_2, \mathbf{R}\right)\\
&\propto \nu^T\left\{\frac{\frac{\nu}{\nu-2} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}'\boldsymbol{\delta}}{1-\boldsymbol{\delta}'\boldsymbol{\delta}}\right\}^{T/2}\left[\prod_{t=1}^T\frac{u_t^{\nu+d}}{\sqrt{\det\left(\mathbf{R}_t\right)}}\prod_{i=1}^d\frac{1}{\sqrt{h_{it}}}\right]\\
&e^{-\frac{1}{2}\sum_{t=1}^T u_t^2\left[x_t^2 + \left(\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{H}_t^{1/2}a_{xt}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\right)'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\left(\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{H}_t^{1/2}a_{xt}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\right)\right]},
\end{aligned}$$

for $\{x_t\}_{t=1}^T > 0$, $0 < \{u_t\}_{t=1}^T < 1$, $\|\boldsymbol{\delta}\|_2 < 1$, $\nu > 2$, $\{\omega_i\}_{i=1}^d > 0$, $\alpha_i + \beta_i < 1$ with $\alpha_i, \beta_i > 0$ for all $i \in \{1, \ldots, d\}$, $\theta_1 + \theta_2 < 1$ for $\theta_1, \theta_2 > 0$, and $\mathbf{R}$ symmetric positive definite

To perform the Bayesian estimation of the model, we decided to sample the joint posterior parameter distribution through the individual sampling of the posterior distributions (exposed further in the present document) by designing a Metropolis-Hastings within Gibbs algorithm.

In order to move forward towards the Bayesian estimation, we need to establish the prior distributions for all of the model parameters. For simplicity, we consider prior distributions that are non-informative and have simple known forms that can be easily understood intuitively, besides being able to incorporate the insight of an expert, when available; also, we assume independence between $\boldsymbol{\lambda}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, $\omega_1, \ldots, \omega_d$, $\{\alpha_i, \beta_i\}$ for all $i \in \{1, \ldots, d\}$, $\{\theta_1, \theta_2\}$, and $\mathbf{R}$. We define $\mathcal{H}$ as the collection of all the hyperparameters in the model. The *a priori* distributions are defined as

$$f(\boldsymbol{\delta}|\mathcal{H}) \propto \mathbf{I}(\|\boldsymbol{\delta}\| < 1),$$

$$f(\nu|\mathcal{H}) \propto (\nu-2)^{a_\nu - 1} \exp\left\{-\frac{\nu}{b_\nu}\right\} \mathbf{I}(\nu > 2);\ a_\nu \gg b_\nu,$$

$$\boldsymbol{\mu}|\mathcal{H} \sim \mathcal{N}_d(\mathbf{m}_\mu, \mathbf{S}_\mu),$$

$$\omega_i|\mathcal{H} \sim \mathcal{IG}(a_\omega, b_\omega), \text{ for } i \in \{1, \ldots, d\},$$

$$f(\alpha_i, \beta_i|\mathcal{H}) = \frac{\Gamma(c_{\alpha\beta})}{\Gamma(c_{\alpha\beta}p_\alpha)\Gamma(c_{\alpha\beta}p_\beta)\Gamma(c_{\alpha\beta}(1-p_\alpha-p_\beta))} \alpha_i^{c_{\alpha\beta}p_\alpha - 1} \beta_i^{c_{\alpha\beta}p_\beta - 1} (1-\alpha_i-\beta_i)^{c_{\alpha\beta}(1-p_\alpha-p_\beta)-1},$$

for $i \in \{1, \ldots, d\}$,

$$f(\theta_1, \theta_2|\mathcal{H}) = \frac{\Gamma(c_\theta)}{\Gamma(c_\theta p_1)\Gamma(c_\theta p_2)\Gamma(c_\theta(1-p_1-p_2))} \theta_1^{c_\theta p_1 - 1} \theta_2^{c_\theta p_2 - 1} (1-\theta_1-\theta_2)^{c_\theta(1-p_1-p_2)-1},$$

$$\mathbf{R}|\mathcal{H} \sim \mathcal{U}\left(\text{s.p.d. matrices with unit diagonal} \in \mathbb{R}^{d\times d}\right).$$

Note that the prior distributions for all pairs (α_i, β_i), $i = 1, \ldots, d$, and (θ_1, θ_2) are derived assuming a Dirichlet-like prior distribution. Also, $\mathcal{IG}(\cdot, \cdot)$ denotes the inverse Gamma distribution.

Under this framework, exact inference is impossible, but we can derive explicit expressions for all of the posterior distributions. Let us remember that $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\delta}, \nu)$.

$$\begin{aligned} f\ & \left(\boldsymbol{\delta}|\nu, \{\boldsymbol{\varepsilon}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T, \mathcal{H}\right) \\ & \propto \mathbf{I}\left(\|\boldsymbol{\delta}\|_2^2 < 1\right) \left[\frac{\frac{\nu}{\nu-2} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2 \boldsymbol{\delta}'\boldsymbol{\delta}}{1-\boldsymbol{\delta}'\boldsymbol{\delta}}\right]^{T/2} \\ & \exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2 \left\{\boldsymbol{\varepsilon}_t'\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\mathbf{\Sigma}^{-1/2}\boldsymbol{\varepsilon}_t - 2a_{xt}\boldsymbol{\varepsilon}_t'\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta} + a_{xt}^2\boldsymbol{\delta}'\mathbf{M}_D^{-1}\boldsymbol{\delta}\right\}\right\} \end{aligned} \tag{14}$$

$$\begin{aligned} f\ & \left(\nu|\boldsymbol{\delta}, \{\boldsymbol{\varepsilon}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T, \mathcal{H}\right) \\ \propto\ & (\nu-2)^{a_\nu - 1}\exp\left\{-\frac{\nu}{b_\nu}\right\}\mathbf{I}(\nu > 2)\,\nu^T \left[\frac{\nu}{\nu-2} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}'\boldsymbol{\delta}\right]^{T/2}\left[\prod_{t=1}^T u_t^\nu\right] \\ & \exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2 \left\{\begin{array}{l} \boldsymbol{\varepsilon}_t'\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\mathbf{\Sigma}^{-1/2}\boldsymbol{\varepsilon}_t - 2a_{xt}\boldsymbol{\varepsilon}_t'\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta} \\ +2\left(\frac{\nu}{\nu-1}\right)\left[\frac{1}{\pi}\left(\frac{\nu}{\nu-1}\right) - \sqrt{\frac{2}{\pi}}x_t\right]\boldsymbol{\delta}'\mathbf{M}_D^{-1}\boldsymbol{\delta} \end{array}\right\}\right\} \end{aligned} \tag{15}$$

$$\begin{aligned} & f\left(\boldsymbol{\mu}|\boldsymbol{\delta}, \nu, \{\omega_i, \alpha_i, \beta_i\}_{i=1}^d, \theta_1, \theta_2, \mathbf{R}, \{\mathbf{y}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T, \mathcal{H}\right) \\ & \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \mathbf{m}_\mu)'\mathbf{S}_\mu(\boldsymbol{\mu} - \mathbf{m}_\mu)\right\}\left[\prod_{t=1}^T\prod_{i=1}^d \frac{1}{\sqrt{h_{it}}}\right] \\ & \exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2 \left[\begin{array}{l} \mathbf{y}_t'\mathbf{H}_t^{-1/2}\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\mathbf{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\mathbf{y}_t \\ +\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\mathbf{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu} \\ -2\mathbf{y}_t'\mathbf{H}_t^{-1/2}\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\mathbf{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu} \\ -2a_{xt}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta} \\ +2a_{xt}\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\mathbf{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta} \end{array}\right]\right\} \end{aligned} \tag{16}$$

$$
\begin{aligned}
&f\left(\{\omega_i\}_{i=1}^d|\boldsymbol{\delta},\nu,\boldsymbol{\mu},\{\alpha_i,\beta_i\}_{i=1}^d,\theta_1,\theta_2,\mathbf{R},\{\mathbf{Y}_t\}_{t=1}^T,\{X_t\}_{t=1}^T,\{U_t\}_{t=1}^T,\mathcal{H}\right)\\
&\propto\left\{\prod_{i=1}^d\omega_i^{1-a_\omega}\right\}\exp\left\{-\sum_{i=1}^d\frac{b_\omega}{\omega_i}\right\}\mathbf{I}\left(\{\omega_i\}_{i=1}^d>0\right)\left[\prod_{t=1}^T\prod_{i=1}^d\frac{1}{\sqrt{h_{it}}}\right]\\
&\quad\exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2\left[\begin{array}{l}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\mathbf{y}_t\\+\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\-2\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\-2a_{xt}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\\+2a_{xt}\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\end{array}\right]\right\}
\end{aligned}\tag{17}
$$

$$
\begin{aligned}
&f\left(\alpha_{i^*},\beta_{i^*}|\boldsymbol{\delta},\nu,\boldsymbol{\mu},\{\omega_i\}_{i=1}^d,\{\alpha_i,\beta_i\}_{\substack{i=1\\i\neq i^*}}^d,\theta_1,\theta_2,\mathbf{R},\{\mathbf{Y}_t,X_t,U_t\}_{t=1}^T,\mathcal{H}\right)\\
&\propto\alpha_{i^*}^{c_{\alpha\beta}p_\alpha-1}\beta_{i^*}^{c_{\alpha\beta}p_\beta-1}\left(1-\alpha_{i^*}-\beta_{i^*}\right)^{c_{\alpha\beta}\left(1-p_\alpha-p_\beta\right)-1}\\
&\quad\mathbf{I}\left(\alpha_{i^*}+\beta_{i^*}<1;\alpha_{i^*},\beta_{i^*}>0\right)\left[\prod_{t=1}^T\frac{1}{\sqrt{h_{i^*t}}}\right]\\
&\quad\exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2\left[\begin{array}{l}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\mathbf{y}_t\\+\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\-2\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\-2a_{xt}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\\+2a_{xt}\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\end{array}\right]\right\}
\end{aligned}\tag{18}
$$

$$
\begin{aligned}
&f\left(\theta_1,\theta_2|\boldsymbol{\delta},\nu,\boldsymbol{\mu},\{\omega_i,\alpha_i,\beta_i\}_{i=1}^d,\mathbf{R},\{\mathbf{Y}_t\}_{t=1}^T,\{X_t\}_{t=1}^T,\{U_t\}_{t=1}^T,\mathcal{H}\right)\\
&\propto\theta_1^{c_\theta p_1-1}\theta_2^{c_\theta p_2-1}\left(1-\theta_1-\theta_2\right)^{c_\theta(1-p_1-p_2)-1}\\
&\quad\mathbf{I}\left(\theta_1+\theta_2<1;\theta_1,\theta_2>0\right)\left[\prod_{t=1}^T\frac{1}{\sqrt{\det\left(\mathbf{R}_t\right)}}\right]\\
&\quad\exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2\left[\begin{array}{l}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\mathbf{y}_t\\+\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\-2\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\-2a_{xt}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\\+2a_{xt}\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\end{array}\right]\right\}
\end{aligned}\tag{19}
$$

$$
\begin{aligned}
&f\left(\mathbf{R}|\boldsymbol{\delta},\nu,\boldsymbol{\mu},\{\omega_i,\alpha_i,\beta_i\}_{i=1}^{d},\theta_1,\theta_2,\{\mathbf{Y}_t\}_{t=1}^{T},\{X_t\}_{t=1}^{T},\{U_t\}_{t=1}^{T},\mathcal{H}\right)\\
&\propto \mathbf{I}\left(\mathbf{R}\ s.p.d.\right)\left[\prod_{t=1}^{T}\frac{1}{\sqrt{\det\left(\mathbf{R}_t\right)}}\right]\\
&\quad\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}u_t^2\left[\begin{array}{l}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\mathbf{y}_t\\+\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\-2\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\-2a_{xt}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\\+2a_{xt}\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\end{array}\right]\right\}
\end{aligned}
\tag{20}
$$

$$
X_t|\boldsymbol{\delta},\nu,\{\boldsymbol{\varepsilon}_t\}_{t=1}^{T},\{X_1,\ldots,X_{t-1},X_{t+1},\ldots,X_T\},\{U_t\}_{t=1}^{T},\mathcal{H}\sim\mathcal{N}_1\left(\mu_x,\sigma_x^2\right)\mathbf{I}\left(\mathbb{R}^+\right),
\tag{21}
$$

where

$$
\mu_x=\boldsymbol{\varepsilon}_t'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}+\sqrt{\frac{2}{\pi}}\frac{\nu}{\nu-1}\boldsymbol{\delta}'\boldsymbol{\delta}
$$

and

$$
\sigma_x^2=u_t^{-2}\left(1-\boldsymbol{\delta}'\boldsymbol{\delta}\right).
$$

By writing $\mathbf{I}\left(\mathbb{R}^+\right)$, we denote that the marginal posterior distribution for x_t is a normal distribution truncated to $x_t\in(0,\infty)$.

$$
U_t^2|\boldsymbol{\delta},\nu,\{\boldsymbol{\varepsilon}_t\}_{t=1}^{T},\{X_t\}_{t=1}^{T},\{U_1,\ldots,U_{t-1},U_{t+1},\ldots,U_T\},\mathcal{H}\sim\mathcal{G}\left(a_u,b_u\right)\mathbf{I}\left(0,1\right),
\tag{22}
$$

where

$$
a_u=\frac{\nu+d+1}{2}
$$

and

$$
b_u=\frac{1}{2}\left\{x_t^2+\left[\boldsymbol{\varepsilon}_t-a_{xt}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\right]'\boldsymbol{\Sigma}^{-1/2}\mathbf{M}_D^{-1}\boldsymbol{\Sigma}^{-1/2}\left[\boldsymbol{\varepsilon}_t-a_{xt}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\right]\right\}.
$$

By writing $\mathbf{I}\left(0,1\right)$, we denote that the marginal posterior distribution for u_t is a Gamma distribution truncated to $u_t\in(0,1)$.

Let us notice that almost none of the distributions exposed above have known forms. For this reason, we designed an MCMC within Gibbs algorithm, as we mentioned before. Now we exhibit the pseudocode to explain the algorithm.

1. Set $m = 0$ and initial values

   $\boldsymbol{\lambda}^{(0)}$, $\nu^{(0)}$, $\boldsymbol{\mu}^{(0)}$, $\boldsymbol{\omega}^{(0)}$, $\left\{\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}\right\}_{i=1}^{d}$, $\left(\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}\right)$, $\mathbf{R}^{(0)}$.

2. Simulate $u_t^{(m)} \sim \mathcal{B}e\left(\nu^{(m)}, 1\right)$ and $x_0^{(m)} \sim \mathcal{N}_1(0,1)$, and compute $x_t^{(m)} = \left|x_0^{(m)}\right| / u_t^{(m)}$ for all $t \in \{1, \ldots, T\}$.

3. Compute $\boldsymbol{\delta}^{(m)} = \frac{1}{\sqrt{1+\boldsymbol{\lambda}^{(m)\prime}\boldsymbol{\lambda}^{(m)}}}\boldsymbol{\lambda}^{(m)}$.

4. Compute $\left\{\mathbf{h}_t^{(m)}\right\}_{t=1}^{T}$ as in (10), $\left\{\mathbf{D}_t^{(m)}\right\}_{t=1}^{T}$ as in (9), $\left\{\mathbf{R}_t^{(m)}\right\}_{t=1}^{T}$ as in (11), and $\left\{\mathbf{H}_t^{(m)}\right\}_{t=1}^{T}$ as in (8).

5. Compute $\boldsymbol{\eta}^{(m)}$ and $\boldsymbol{\Sigma}^{(m)}$ to satisfy the null mean and unit variance restrictions as specified in (12) and (13), respectively.

6. Compute $\boldsymbol{\varepsilon}_t^{(m)} = \mathbf{H}_t^{(m)-1/2}\left(\mathbf{y}_t - \boldsymbol{\mu}^{(m)}\right)$ for all $t \in \{1, \ldots, T\}$.

7. Obtain a sample $\boldsymbol{\delta}^{(m+1)}$ from the distribution in (14). Compute $\boldsymbol{\lambda}^{(m+1)} = \frac{1}{\sqrt{1-\boldsymbol{\delta}^{(m+1)\prime}\boldsymbol{\delta}^{(m+1)}}}\boldsymbol{\delta}^{(m+1)}$.

8. Obtain a sample $\nu^{(m+1)}$ from the distribution in (15).

9. Obtain a sample $\boldsymbol{\mu}^{(m+1)}$ from the distribution in (16). Execute step 4 alone.

10. Obtain a sample $\boldsymbol{\omega}^{(m+1)}$ from the distribution in (17). Execute step 4 alone.

11. Obtain samples $\alpha_{i^*}^{(m+1)}, \beta_{i^*}^{(m+1)}$ from the distribution in (18) for all $i^* \in \{1, \ldots, d\}$. Execute step 4 alone.

12. Obtain a sample $\theta_1^{(m+1)}, \theta_2^{(m+1)}$ from the distribution in (19). Execute step 4 alone.

13. Obtain a sample $\mathbf{R}^{(m+1)}$ from the distribution in (20). Execute step 4 alone.

14. Obtain a sample $x_t^{(m+1)}$ from the distribution in (21).

15. Obtain a sample $u_t^{(m+1)}$ from the distribution in (22).

16. Set m = m + 1 and repeat steps 4 thru 15 until $m = M'$ for a large M'.

For the parameters of the model, we apply Metropolis Hastings steps because it is not possible to sample directly from their posterior distributions, but the steps that correspond to the latent variables can be performed using a Gibbs sampler.

We simulate $\boldsymbol{\delta}$ through a transformation of a normal random vector to ensure that the restriction is satisfied. ν, $\boldsymbol{\omega}$, (α_i, β_i) for $i \in \{1, \dots, d\}$, $\boldsymbol{\theta}$ are sampled by means of logarithmic transformations based on their current values, also ensuring that their restrictions are met. The candidate for $\boldsymbol{\mu}$ is proposed using a normal distribution centered on the current value. Finally, the potential new values for $\mathbf{R}$ are proposed by standardizing a covariance matrix that takes into account the current value for this correlation matrix. In the case of x_t and u_t, for $t \in \{1, \dots, T\}$, the posterior distributions have known forms, so sampling is straightforward.

5 Examples

To illustrate our proposal, we present two approaches. First, we estimate the parameters of simulated data in order to evaluate our procedure. Afterwards, we model a real data set that contains information about the joint behavior of the DAX, CAC40, and Nikkei indices. In both cases, we specify the Bayesian prior hyperparameters as $a_\nu = 100$, $b_\nu = 0.01$, $\mathbf{m}_\mu = \mathbf{0}$, $\mathbf{S}_\mu = 0.07\mathbf{I}$, $a_\omega = 0.001$, $b_\omega = 0.001$, $c_{\alpha\beta} = 10$, $p_\alpha = 0.1$, $p_\beta = 0.85$, $c_\theta = 10$, $p_1 = 0.9$, $p_2 = 0.05$. The MCMC within Gibbs sampler that we designed is run for 5000 iterations to burn in and 10000 iterations in equilibrium for the simulation case, and 20000 iterations to burn in and 25000 in equilibrium for the real data.

5.1 Simulated data

With the purpose of testing our work, we started by modeling simulated data sets.

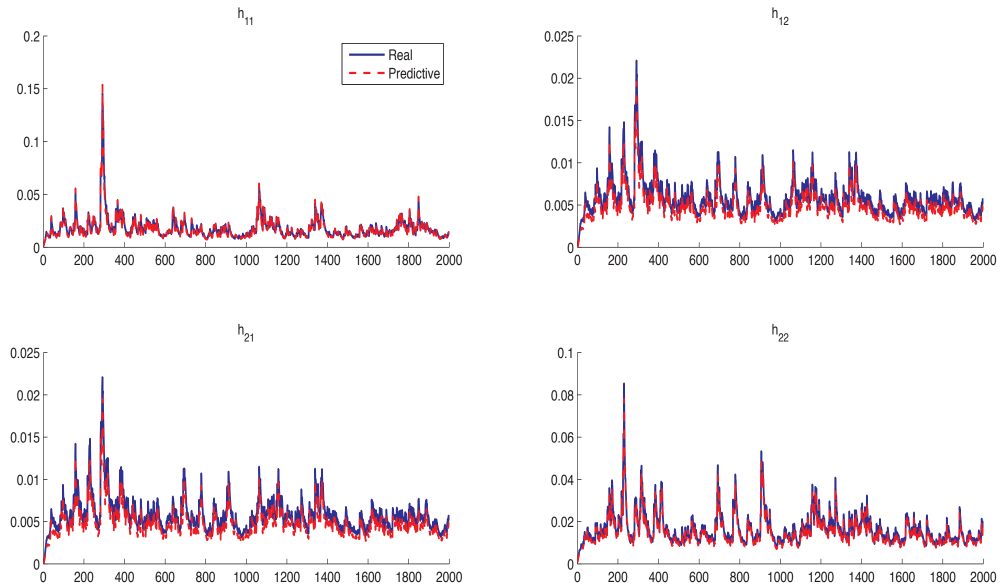


Figure 3: True (solid line) and fitted (dotted line) volatilities for the 2-dimensional simulated data.

In every case, we obtained fitted volatilities as the mean value of the fitted volatilities for the 10000 parameter values of the MCMC sampler.

First, we generated a 2-dimensional time series of returns from a multivariate 2-dimensional Skew-Slash DCC model with $T = 2000$ observations, Skew-Slash parameters $\boldsymbol{\eta} = (0.0769, 0.0769)'$, $\boldsymbol{\Sigma}$ with diagonal elements $(0.6036, 0.6036)$ and off-diagonal element 0.0036, $\boldsymbol{\lambda} = (-0.1, -0.1)'$, $\nu = 5$, and GARCH parameters $\boldsymbol{\mu} = (0, 0)'$, $\boldsymbol{\omega} = (0.001, 0.001)'$, $(\alpha_1, \beta_1)' = (\alpha_2, \beta_2)' = (0.1, 0.85)'$, $\boldsymbol{\theta} = (0.9, 0.05)'$, and $\mathbf{R}$ with unit diagonal and off-diagonal element $\rho_{12} = 0.7$. Notice that $\boldsymbol{\Sigma}$ and $\mathbf{R}$ are symmetric matrices; also, keep in mind that $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ are set to verify the zero mean and unit variance of the innovations.

Figure 3 shows the real volatilities (blue solid line) and the Bayesian posterior mean volatility estimates (red dashed line), and we can see that they are almost indistinguishable. Figure 4 illustrates the comparison between the theoretical marginal densities of the innovations (solid blue line), and the mean predictive marginal densities (dashed red line), both compared to the marginal histograms of the real inno-

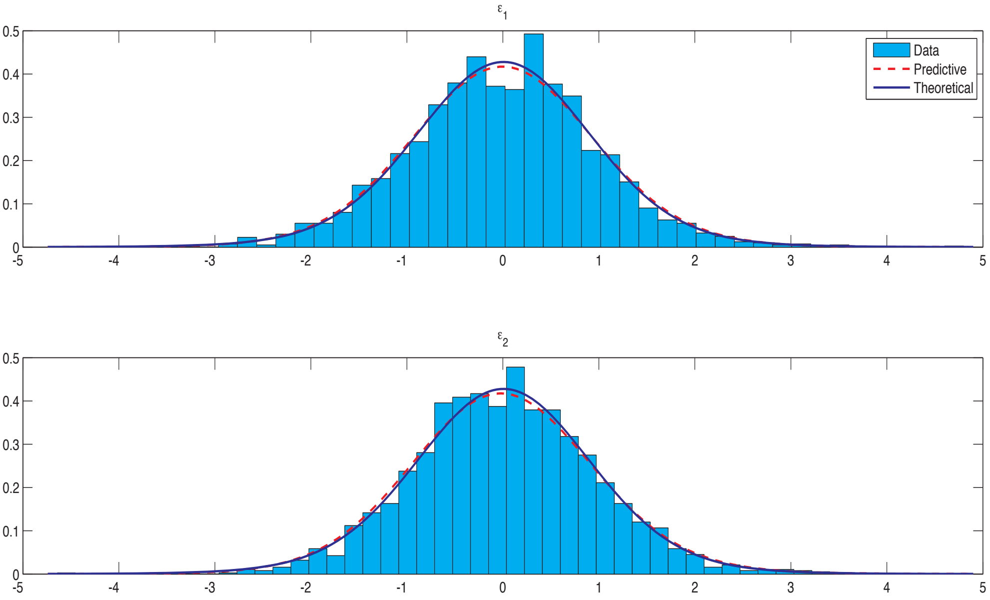
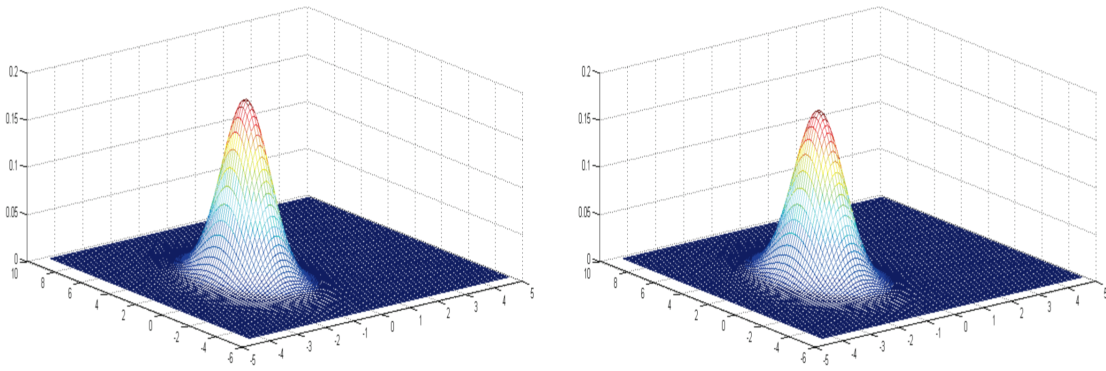
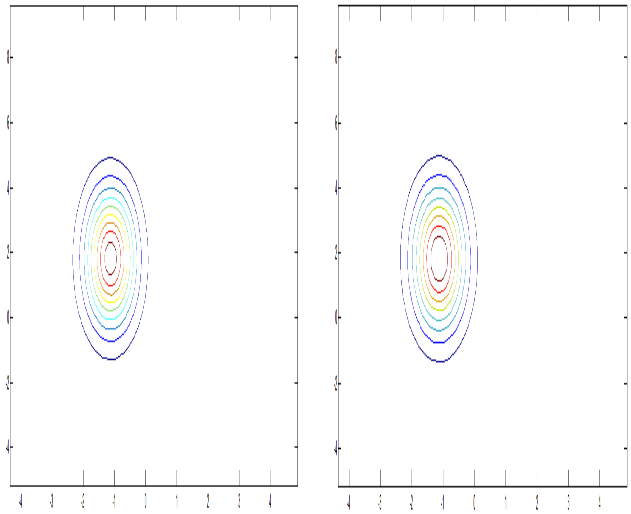


Figure 4: True (solid line) and fitted (dotted line) predictive innovation marginal densities compared to their histograms for the 2-dimensional data set.



(a) (b)

Figure 5: True (a) and fitted (b) predictive innovation joint densities for the 2-dimensional simulated data.



(a) (b)

Figure 6: Contour plots of the true (a) and fitted (b) predictive innovation densities for the 2-dimensional simulated data.

vations presented by the 2-dimensional simulated data. Here, we find that not only are both marginal densities very similar, but the estimation results are apparently very good; actually, it is hard to see which function seems closer to the observations. Finally, Figure 5 shows the joint theoretical density of the innovations in 5(a) next to their joint predictive density in 5(b), while Figure 6 shows the contour plot of the joint theoretical density of the innovations in 6(a) together with their contour plot of the joint predictive density in 6(b). From both sets of figures, we can see that, even though we are not able to superimpose them, their similarity is clear.

Second, we generate a 3-dimensional time series of returns from a multivariate Skew-Slash DCC model with $T = 2000$ observations, Skew-Slash parameters $\boldsymbol{\eta} = (0.0768, 0.0768, 0.0768)'$, $\boldsymbol{\Sigma}$ with diagonal elements all equal to 0.6035 and off-diagonal elements all with value 0.0035, $\boldsymbol{\lambda} = (-0.1, -0.1, -0.1)'$, $\nu = 5$, and GARCH parameters $\boldsymbol{\mu} = (0, 0, 0)'$, $\boldsymbol{\omega} = (0.001, 0.001, 0.001)'$, $(\alpha_1, \beta_1)' = (\alpha_2, \beta_2)' = (\alpha_3, \beta_3)' = (0.1, 0.85)'$, $\boldsymbol{\theta} = (0.9, 0.05)'$, and $\mathbf{R}$ with unit diagonal and off-diagonal elements $\rho_{12} = 0.5$, $\rho_{13} = 0.7$, and $\rho_{23} = 0.3$. Notice that $\boldsymbol{\Sigma}$ and $\mathbf{R}$ are symmetric matrices,

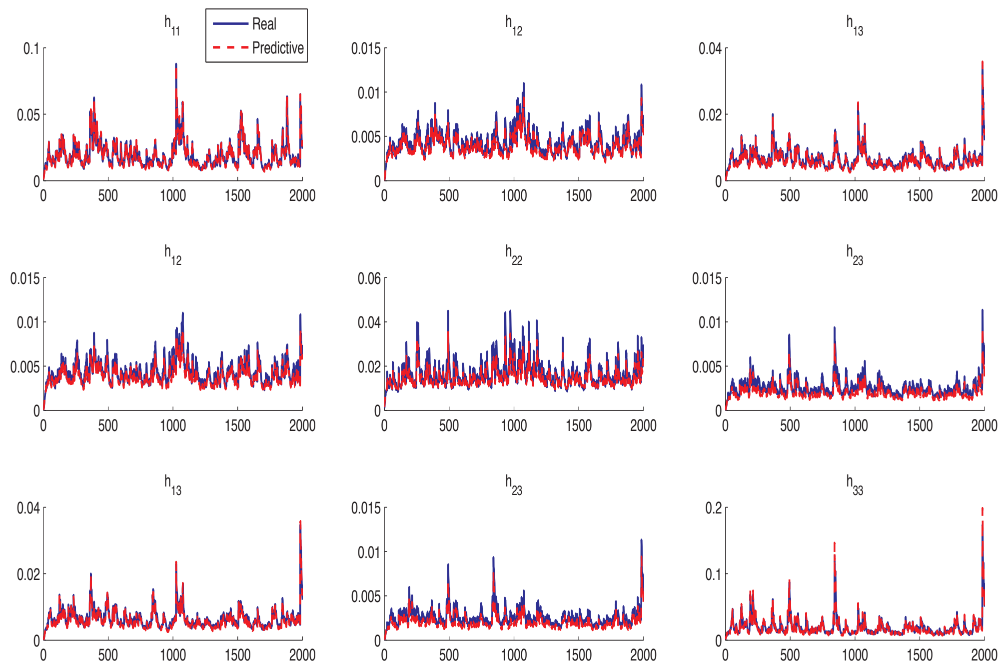


Figure 7: True (solid line) and fitted (dotted line) volatilities for the 3-dimensional data set.

also keep in mind that $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ are set to verify the zero mean and unit variance of the innovations.

Figure 7 shows the real volatilities (blue solid line) and the Bayesian posterior mean volatility estimates (red dashed line), and we can see that they are almost indistinguishable. Figure 8 illustrates the comparison between the theoretical marginal densities of the innovations (solid blue line), and the mean predictive marginal densities (dashed red line), both compared to the marginal histograms of the real innovations presented by the 3-dimensional simulated data. Here, we also find that both densities are almost indistinguishable, and the estimation results appear to be very good as well. Finally, Figure 9 shows the 2-variable marginal theoretical densities of the innovations in 9(a) for $(\varepsilon_1, \varepsilon_2)$, 9(c) for $(\varepsilon_1, \varepsilon_3)$, and 9(e) for $(\varepsilon_2, \varepsilon_3)$ next to their 2-variable marginal predictive densities in 9(b) for $(\varepsilon_1, \varepsilon_2)$, 9(d) for $(\varepsilon_1, \varepsilon_3)$, and 9(f) for $(\varepsilon_2, \varepsilon_3)$.

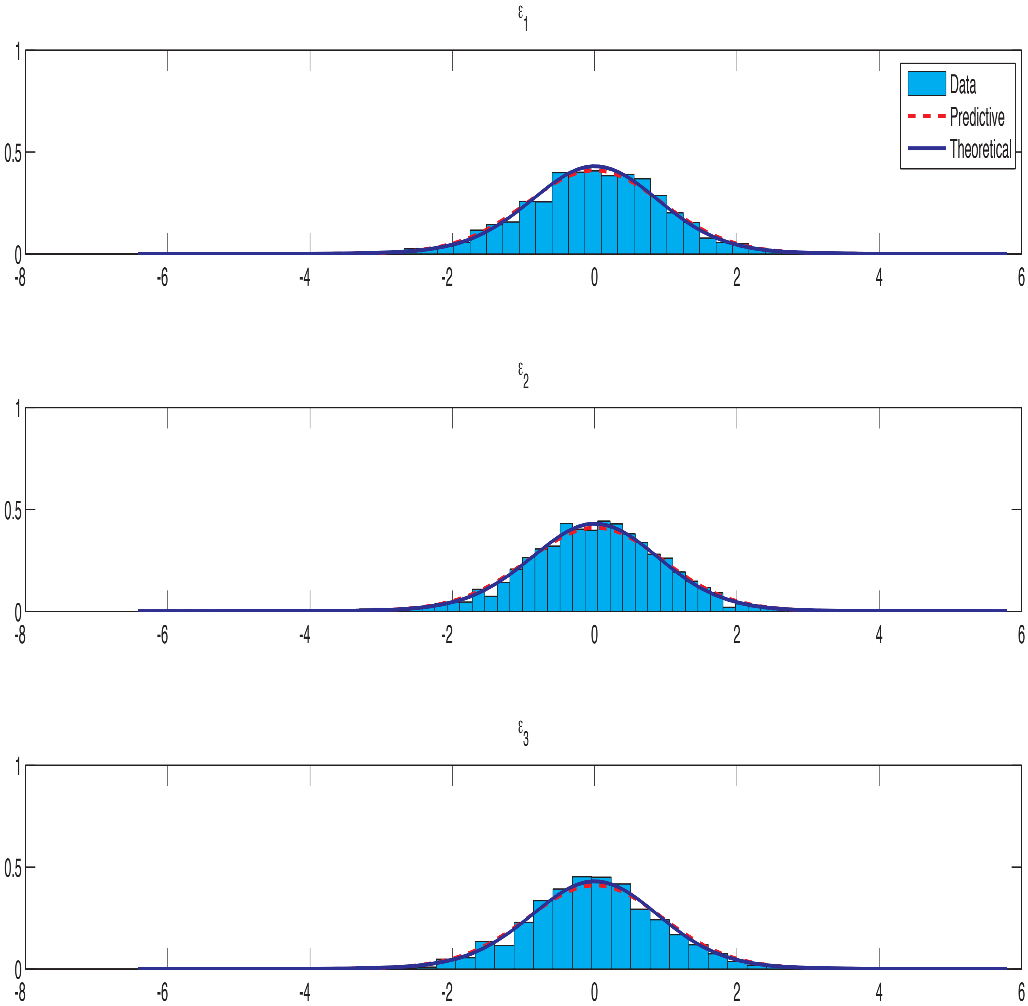
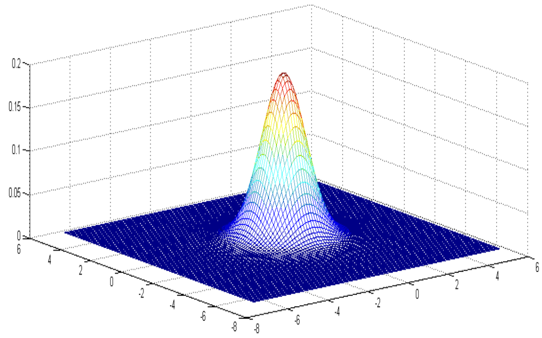
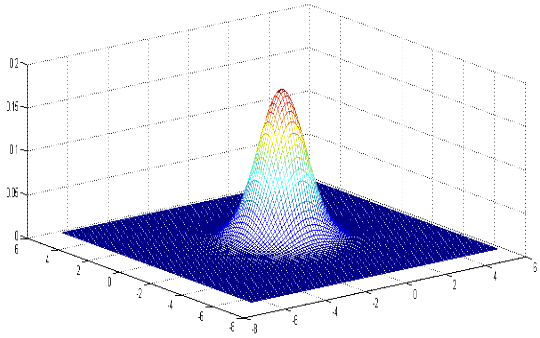


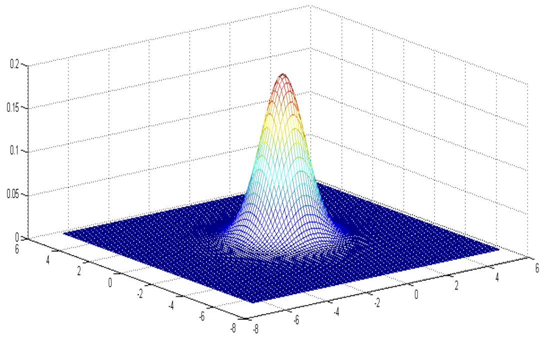
Figure 8: True (solid line) and fitted (dotted line) predictive innovation marginal densities compared to their histograms for the 3-dimensional data set.



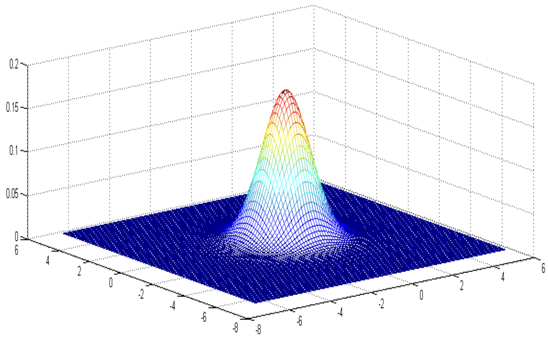
(a)



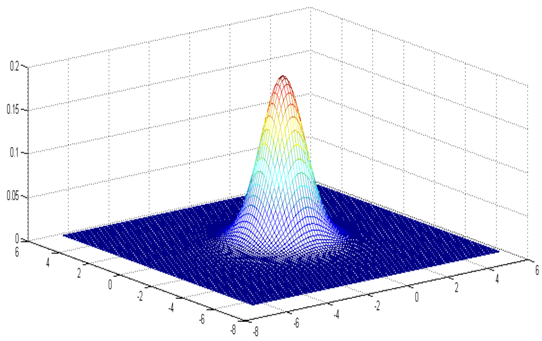
(b)



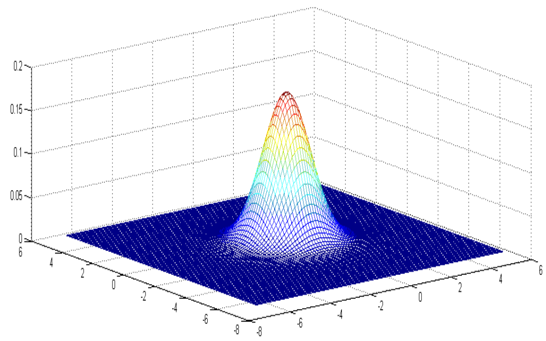
(c)



(d)



(e)



(f)

Figure 9: True (left) and fitted (right) predictive innovation 2-variable marginal densities for the 3-dimensional simulated data.

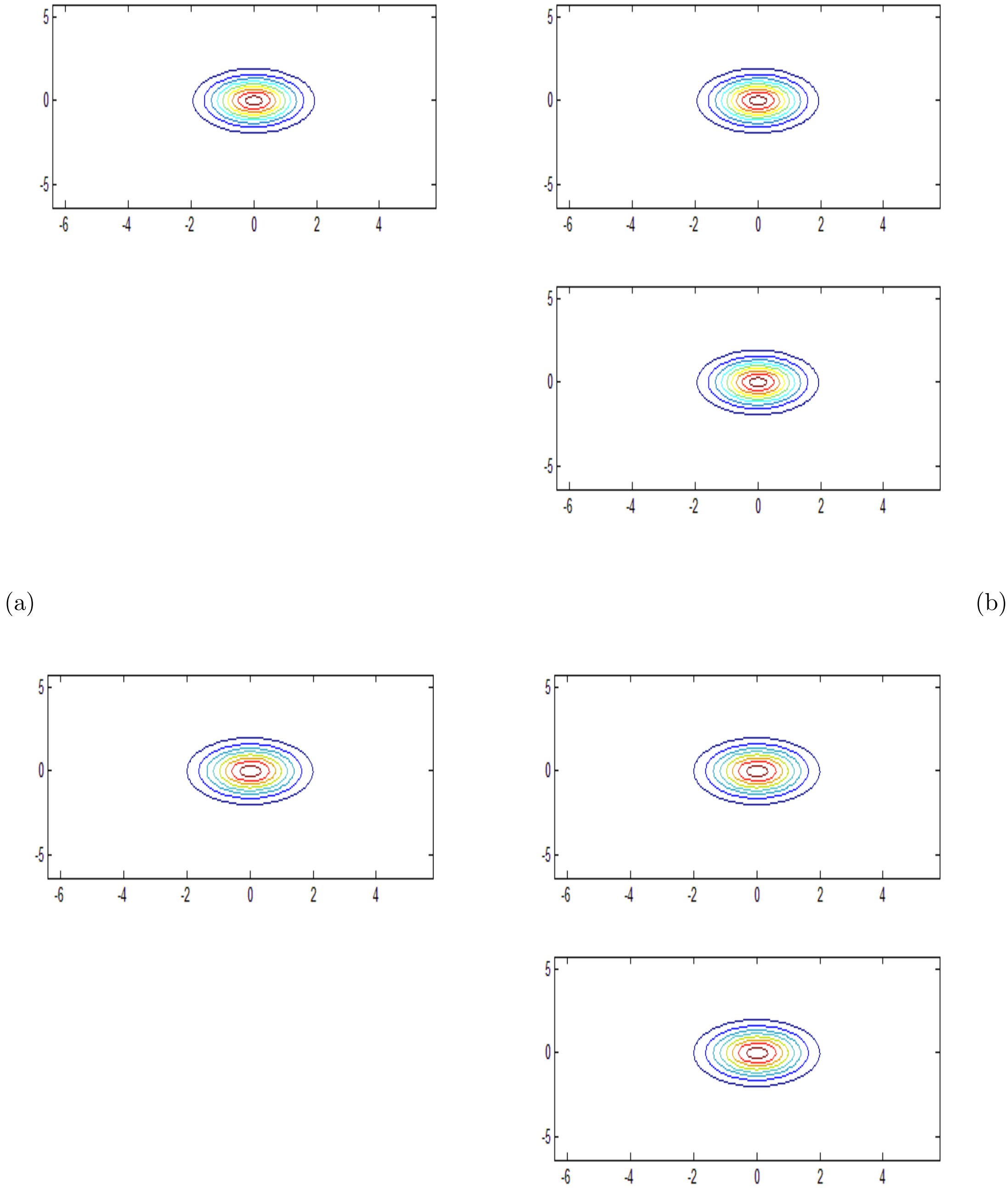


Figure 10: Contour plots of the true (a) and fitted (b) predictive innovation 2-variable marginal densities for the 3-dimensional simulated data.

Figure 10 shows the contour plots of the 2-variable marginal theoretical densities of the innovations in 10(a) together with their contour plots of the 2-variable marginal predictive densities in 10(b). From both sets of figures, we can see that, even though we are not able to superimpose them, their similarity is clear.

5.2 Real data example

To illustrate the usefulness of the approach proposed in the present document, in this section, we analyze the daily log-returns of the DAX, CAC40, and Nikkei stock market indices. Besides this, we obtained fitted volatilities as the mean value of the fitted volatilities of the 25000 parameter values of the MCMC sampler.

The original information consists on the daily closing prices of the stock market indices in Frankfurt (DAX), Paris (CAC40), and Tokyo (Nikkei) from October 10$^{\text{th}}$, 1991 until December 30$^{\text{th}}$, 1997, which leads to 1624 observations[1]. To obtain the daily log-returns, we make a transformation such that the daily log-return at time t, for a certain stock market index, is given by 100 times the logarithm of the increase rate of the closing price of the day t, respect to the closing price of the day $t-1$. This way, we lose the first observation and end up with a total of $T = 1623$ log-returns.

Figure 11 shows the plot of the time series generated by the log-returns of our three stock market indices, and we will associate the DAX index to the first innovation component, ε_1, the CAC40 index to the second innovation component, ε_2, and the Nikkei index to the third (and last) innovation component, ε_3. We can see that the data is clearly heteroskedastic, and it also presents a perturbation in the end that calls for a model that can capture heavy tails in the innovations, as we propose with our 3-dimensional Skew-Slash DCC model. Also, we can already see that the data is mostly symmetric in all three cases.

In fact, the univariate sample means are $(0.0614, 0.0295, 0.0601)'$, the standard deviations of the separate series are $(0.9988, 1.0893, 1.6101)'$, the individual skewness

[1] The data is freely available in http://robjhyndman.com/tsdldata/data/FVD1.dat

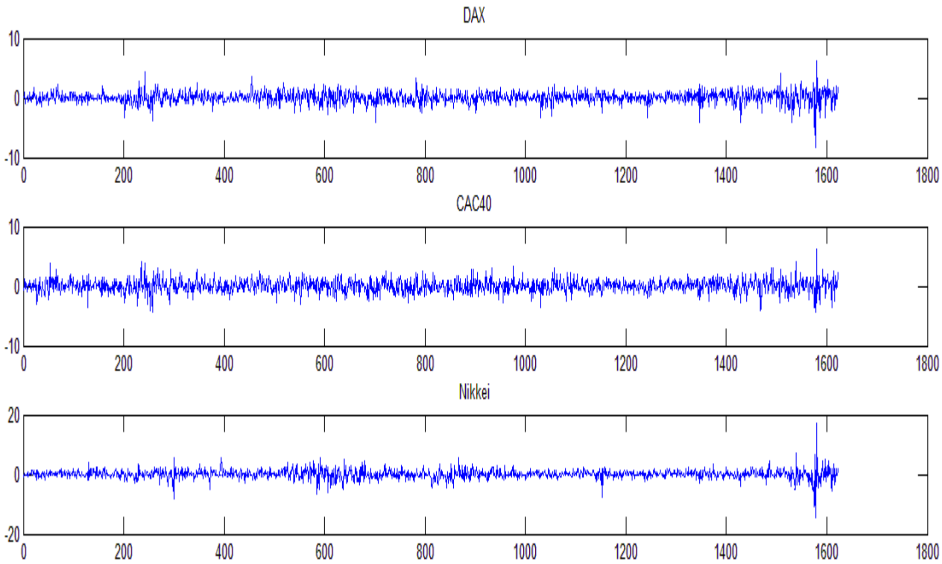


Figure 11: DAX, CAC40, and Nikkei log-returns between October 10$^{\text{th}}$, 1991 and December 30$^{\text{th}}$, 1997.

coefficients are $(-0.5799, -0.0434, -0.2078)'$, and the univariate kurtosis coefficients of the three considered log-return series are given by $(8.5006, 4.4835, 18.5369)'$.

Clearly, the normal distribution could not possibly be able to reflect the behavior exhibited by this data set. This leads us to believe that it makes sense to use the proposed Dynamic Conditional Correlation model with Skew-Slash innovations for this data because all of the indices considered have a slight skewness and a high kurtosis, especially the Nikkei stock market index.

Figure 12 shows the fitted volatilities, while Figure 13 shows the estimated innovation marginal densities. On the other hand, in Figure 14, we can see the 2-variable marginal predictive densities of the innovations, and Figure 15 shows their corresponding contour plots.

Now, we proceed to compare the performance of our model against some of the alternative proposals of Fioruci, Ehlers, and Andrade (2014). They perform their main estimation using another member of the Skew-Normal/independent family: the Skew-T (and they compare it to the performance of other reference distributions).

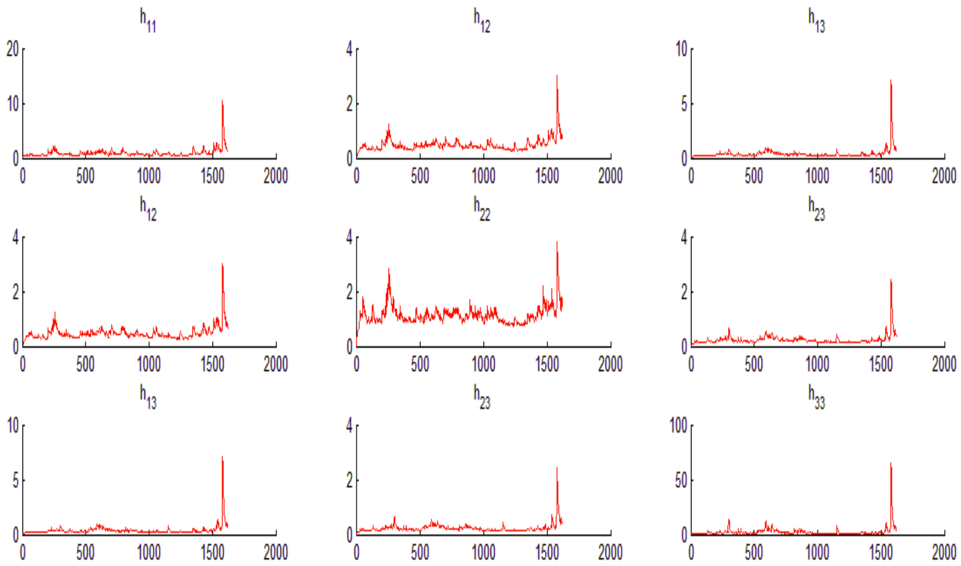


Figure 12: Volatility estimates for the DAX, CAC40, and Nikkei log-returns.

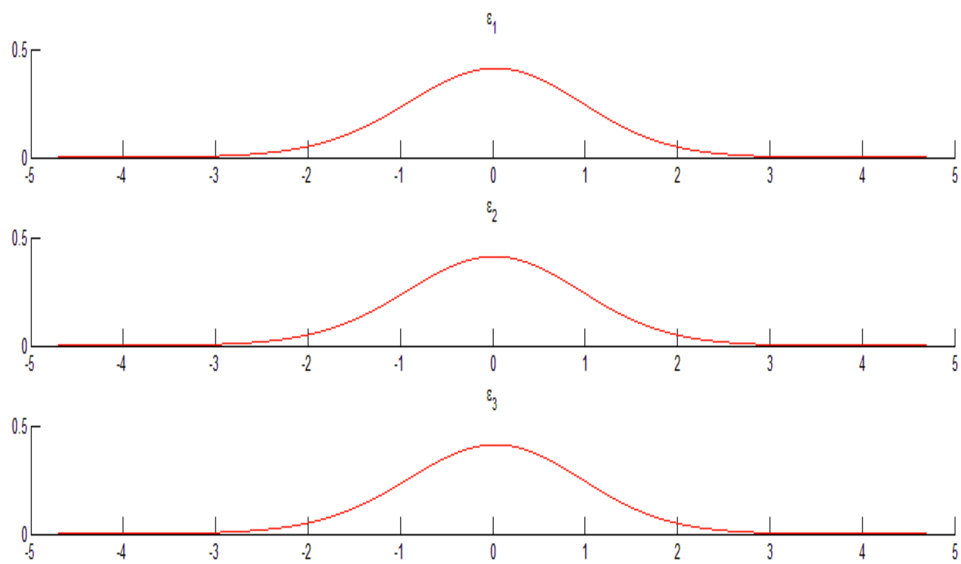
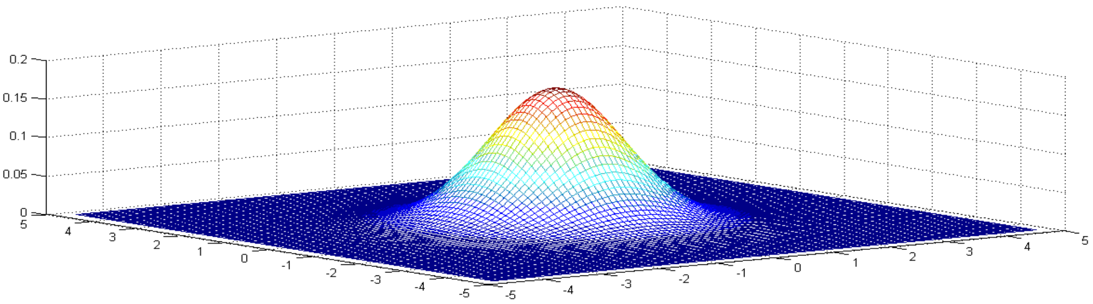


Figure 13: Estimated innovation marginal densities for the DAX, CAC40, and Nikkei log-returns.

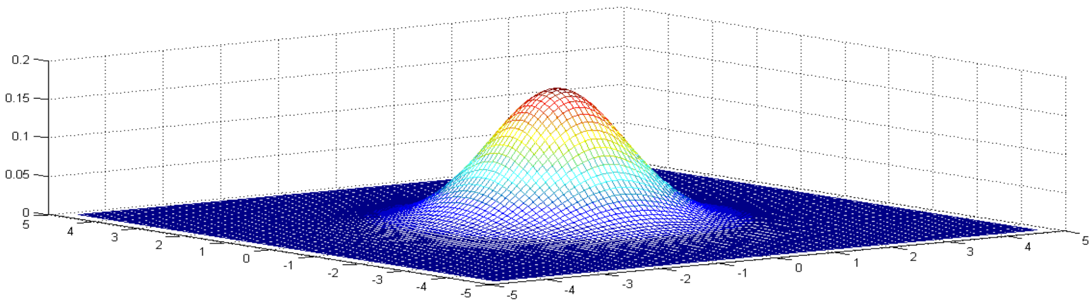
(a)

DAX-CAC40



(b)

DAX-Nikkei



(c)

CAC40-Nikkei

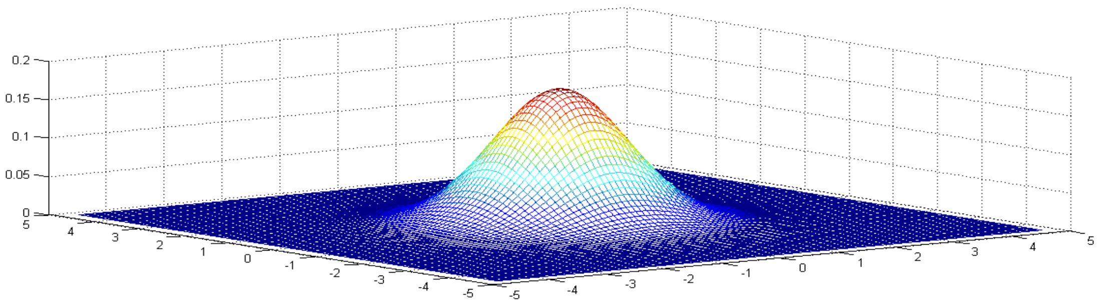


Figure 14: Estimated innovation 2-variable marginal densities for the DAX, CAC40, and Nikkei log-returns.

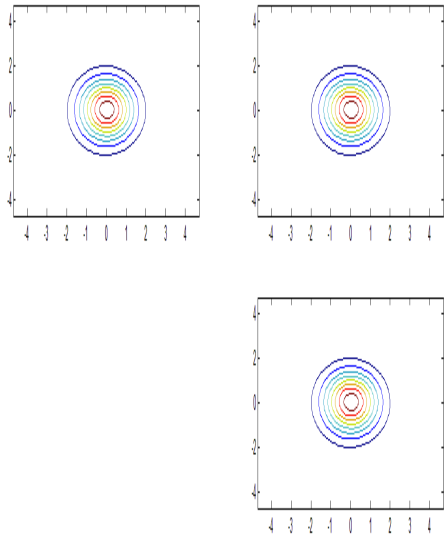


Figure 15: Contour plots of the estimated innovation 2-variable marginal densities for the innovations of the DAX, CAC40, and Nikkei log-returns.

A difference between our approaches is that they establish a null drift parameter, $\boldsymbol{\mu}$, and do not estimate the correlation matrix, while we allow both sets of parameters to define themselves. To compare both models, we decided to set $\boldsymbol{\mu} = \mathbf{0}$ as well, and set the value of $\mathbf{R}$ as the correlation matrix of the data, and leave them fixed during the estimation in order to be able to recreate an analogous scenario.

After we obtained our own estimations, we used the Deviance Information Criterion (DIC), defined by Spiegelhalter, Best, Carlin, and van der Linde (2002), a measurement already provided in their paper, for comparison.

In Table 1 we can find the Deviance Information Criterion (DIC) values for the Dynamic Conditional Correlation model with different innovation distributions: Gaussian, Skew-Normal, Generalized Error Distribution (GED), Skew-GED, Student's-t, Skew-t, and Skew-Slash. In this table, we are able to show the model comparison performed by Fioruci, Ehlers, and Andrade (2014), in which they choose the Skew-t distribution to model the innovations, and also incorporate the comparison with our model.

Table 1: DIC for the Dynamic Conditional Correlation model with several innovation distributions for the DAX, CAC40, and Nikkei data.

	DIC
Normal	13957.53
Skew-Normal	13947.62
GED	13828.97
Skew-GED	13823.48
Student's-t	13810.36
Skew-t	**13803.32**
Skew-Slash	**13801.71**

As they explain, for the DAX, CAC40, and Nikkei data, the distributions with heavier tails exhibit a better behavior than the normal distribution. In fact, the best performance they select is the one exhibited by the mentioned Skew-t distribution, obtaining a $DIC_{SkT} = 13803.32$. We obtained a smaller, but similar value $DIC_{SSL} = 13801.7153$. Therefore, the better performance is attained by the Dynamic Conditional Correlation model with Skew-Slash innovations, although the difference with the Skew-t distribution is small.

6 Conclusions and Extensions

In the present document, we have studied the Skew-Slash distribution, a distribution able to capture the slight skewness and high kurtosis that are often observed in the residuals of financial time series of returns after fitting a conditional heteroskedastic model.

We proposed to model the structure of multidimensional time series of financial returns by means of the Dynamic Conditional Correlation model with Skew-Slash innovations and, while we explained our proposal, we constructed a methodology for the model fitting from a Bayesian point of view.

To illustrate the abilities of our proposed model and methodology, we applied our ideas to simulated series of returns and, later on, we estimated a real data set that has already been studied by Fioruci, Ehlers, and Andrade (2014), using a Dynamic Conditional Correlation model with Skew-t innovations.

We compared our models using the DIC, that manifested a better performance of our model capturing the essence of the data set under study, but not with a very big difference, which lead us to believe that, for this kind of data sets, different members of the Skew-Normal/independent family of distributions might have a similar performance fitting their features, but maybe each data set has its own better partner, to call it some way.

There are a number of subjects that are related to the work presented in this document, and that we find interesting, but have not been able to explore yet.

First of all, we believe that it would be very interesting to open up to the possibility of a Dynamic Conditional Correlation model with Skew-Slash innovations, based in the structure of a model of the form GARCH$(\{p_i, q_i\}_{i=1}^{d})$.

Second, we think that one of the reasons that give relevance to finding a way to model financial data sets that allows us to capture the essence of the behavior of the data in the best way possible is the necessity of the investors for a good assessment of the risk they would be incurring in if they decided to include certain assets in their portfolio. One possible way of answering this inquietude could be to implement our model to the characterization of certain risk measures, such as the Value at Risk or the Conditional Value at Risk.

Finally, it would be interesting to address the possibility of working with a distribution that can not only capture the lack of symmetry or the presence of heavy tails, but allows for more flexibility in the structure of the kurtosis parameter; for instance, we could try to study the possibility of a kurtosis parameter with more than one component.

References

- G. P. Aielli. Dynamic Conditional Correlation: On Properties and Estimation. *Journal of Business & Economic Statistics*, 31(3):282–299, 2013.

- C. O. Alexander and A. M. Chibumba. Multivariate Orthogonal Factor GARCH. *University of Sussex Discussion Papers in Mathematics*, 1997.

- R. B. Arellano-Valle, H. Bolfarine, and V. H. Lachos. Skew-Normal Linear Mixed Models. *Journal of Data Science*, 3:415–438, 2005.

- M. C. Ausín and P. Galeano. Bayesian Estimation of the Gaussian Mixture GARCH Model. *Computational Statistics & Data Analysis*, 51:2636–2652, 2007.

- M. C. Ausín, P. Galeano, and P. Ghosh. A Semiparametric Bayesian Approach to GARCH-type Models with Application to Value at Risk Estimation. *European Journal of Operational Research*, 232:350–358, 2014.

- A. Azzalini. A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, 12(2):171–178, 1985.

- A. Azzalini and A. Capitanio. Statistical Applications of the Multivariate Skew-Normal Distribution. *Journal of the Royal Statistical Society. Series B*, 61(3):579–602, 1999.

- A. Azzalini and A. Capitanio. Distributions Generated by Perturbations of Symmetry with Emphasis on a Multivariate Skew-t Distribution. *Journal of the Royal Statistical Society. Series B*, 65:367–389, 2003.

- A. Azzalini and A. Dalla Valle. The Multivariate Skew-Normal distribution. *Biometrika*, 83(4):715–726, 1996.

- X. Bai, J. R. Russell, and G. C. Tiao. Kurtosis of GARCH and Stochastic Volatility Models with Non-Normal Innovations. *Journal of Econometrics*, 114(2):349–360, 2003.

- M. Billio, M. Caporin, and M. Gobbo. Flexible Dynamic Conditional Correlation Multivariate GARCH Models for Asset Allocation. *Applied Financial Economics Letters*, 2:123–130, 2006.

- T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

- T. Bollerslev. A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics*, 69(3):542–547, 1987.

- T. Bollerslev. Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *The Review of Economics and Statistics*, 72(3):498–505, 1990.

- T. Bollerslev, R. F. Engle, and J. M. Woolridge. A Capital Asset Pricing Model With Time Varying Covariances. *Journal of Political Economy*, 96:116–131, 1988.

- L. Cappiello, R. F. Engle, and K. Sheppard. Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics*, 4:537–572, 2006.

- R. F. Engle. Dynamic Conditional Correlation. A Simple Class of Multivariate GARCH Models. *Journal of Business & Economic Statistics*, 20:339–350, 2002.

- R. F. Engle, C. W. J. Granger, and D. Kraft. Combining Competing Forecasts of Inflation Using a Bivariate ARCH Model. *Journal of Economic Dynamics and Control*, 8:151–165, 1984.

- R. F. Engle and F. K. Kroner. Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, 11:122–150, 1995.

- R. F. Engle, V. K. Ng, and M. Rothschild. Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills. *Journal of Econometrics*, 45:213–238, 1990.

- R. F. Engle and K. Sheppard. Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. *NBER Working Papers, 8554, MA: Nation Bureau of Economic Research*, 2001.

- C. Fernández and M. Steel. On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93:359–371, 1998.

- J. A. Fioruci, R. S. Ehlers, and M. G. Andrade. Bayesian Multivariate GARCH Models with Dynamic Correlations and Asymmetric Error Distributions. *Journal of Applied Statistics*, 41(2):320–331, 2014.

- C. G. de la Fuente, P. Galeano, and M. P. Wiper. Modeling Financial Time Series With the Skew Slash Distribution. *UC3M Working papers. Statistics and Econometrics*, 2012.

- P. Galeano and M. C. Ausín. The Gaussian Mixture Dynamic Conditional Correlation Model: Parameter Estimation, Value at Risk Calculation, and Portfolio Selection. *Journal of Business & Economic Statistics*, 28(4):559–571, 2010.

- M. G. Genton, L. He, and X. Liu. Moments of Skew-Normal Random Vectors and Their Quadratic Forms. *Statistics & Probability Letters*, 51:319–325, 2000.

- T. Jeantheau. Strong Consistency of Estimators for Multivariate ARCH Models. *Econometric Theory*, 14:70–86, 1998.

- M. C. Jones and M. J. Faddy. A Skew Extension of the t-distribution, with Applications. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(1):159–174, 2003.

- H. Kawakatsu. Matrix Exponential GARCH. *Journal of Econometrics*, 134:95–128, 2006.

- V. H. Lachos, D. K. Dey, and V. G. Cancho. Robust Linear Mixed Models with Skew-Normal Independent Distributions from a Bayesian Perspective. *Journal of Statistical Planning and Inference*, 139:4098–4110, 2009.

- V. H. Lachos, V. Garibay, F. V. Labra, and R. Aoki. A Robust Multivariate Measurement Error Model with Skew-Normal/Independent Distributions and Bayesian MCMC Implementation. *Statistical Methodology*, 6:527–541, 2009 a.

- V. H. Lachos, V. Garibay, F. V. Labra, and R. Aoki. The Multivariate Skew-Normal Distribution. *Biometrika*, 83(4):715–726, 2009 b.

- V. H. Lachos, F. V. Labra, and P. Ghosh. Multivariate Skew-Normal/Independent Distributions: Properties and Inference. *Scandinavian Journal of Statistics.*, 2007.

- K. Lange and J. S. Sinsheimer. Normal/independent Distributions and Their Applications in Robust Regression. *Journal of Computational and Graphical Statistics*, 2(2):175–198, 1993.

- K. V. Mardia. *Statistical Distributions in Scientific Work*, volume 3. Patil, G. P., Kotz, S., and Ord, J. K. (D, Reidel Publishing Company), Dordrecht, 1974.

- D. B. Nelson. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2):347–370, 1991.

- A. Silvennoinen and T. Teräsvirta. Multivariate GARCH Models. SSE. Technical report, EFI Working Paper Series in Economics and Finance, 2008.

- D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society. Series B*, 64:583–639, 2002.

- Y. K. Tse and K. C. Tsui. A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations. *Journal of Business & Economic Statistics*, 20(3):351–362, 2002.
- R. van der Weide. GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model. *Journal of Applied Econometrics*, 17:549–564, 2002.
- I. D. Vrontos, P. Dellaportas, and D. N. Politis. A Full-Factor Multivariate GARCH Model. *Econometrics Journal*, 6:312–334, 2003.
- J. Wang and M. G. Genton. The Multivariate Skew-Slash Distribution. *Journal of Statistical Planning and Inference*, 136:209–220, 2006.