UNIVERSIDAD CARLOS III DE MADRID

Dissertation

Modeling Financial Returns with Skew-Slash Innovations

Author: **Cristina García de la Fuente**

Advisors:

Pedro Galeano San Miguel

Michael P. Wiper

Universidad Carlos III de Madrid

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in the

Graduate program in Mathematical Engineering

Department of Statistics

Madrid, June 2014

“Perhaps it is good to have a beautiful mind, but an even greater gift is to discover a beautiful heart.”

- Beautiful mind

Contents

List of Figures

List of Tables

Acknowledgments

To my advisors, Pedro Galeano and Michael P. Wiper, for all your time, for all your effort, because we have been a team the past few years. Thank you for your encouraging words during rough patches. Thank you for being part of this great achievement; clearly, it would not have been possible without you.

To María Durbán, Elisa Molanes, and Gabriel Núñez, for all your support and advise. To Javi, Goretti, Eli, Diego, Jorge, Peter, Rasi, Dae Jin, Gilberto, Alberto, Joanna, Raqui, Carlo, Nuria, and Sofi for being my friends and companion, and for sharing this journey with me.

To Silvano Espíndola, Marcela González, Guillermo Grabinsky, Rubén Hernández Cid, Manuel Mendoza, and Rosario Sarmiento. Not only did you teach me what you know, but you passed on to me the love for knowledge and teaching. Thank you for your affection and for helping me believe in myself.

To Manu, Arely, Déborah, and Rafa. Thank you for your support, acceptance, and everything we have shared. To Lois, the choirs from Guadalupe, Carmina, Enrique, Corina, Daniel, José, Begoña, Paula, and Clara, because you have become part of my family and especially helped me to get through the past few years.

To Ramón, Beatriz†, León†, J. María†, Soledad†, Lourdes, Andrea, Ricardo, Sofía, Bony, Fernando, Mónica, Renata, Alejandro, Marián, Ainara, Nico, Paulina, Natalia, Pablo, Susana, Gustavo, Gustavo, Susana, Santiago, Diego, Ramón, Mónica, Nicolás, Regina, Paola, Rafa. It has been very tough to be away from you; I love you with all my heart; you are the reason of who I am.

To Aitor$^{\aleph}$. Thank you for your love, for believing in me, for listening, for hoping with me and for me, for always wanting to help, for your companionship, for your support, for helping me move forward, and for wanting to become better together. You are my engine. I love you.

To my parents. Thank you for believing in me, for always being there for me, for every sacrifice you have made for me, as well as the ones we have made together, for always encouraging me and helping me move on. You gave me life, and you keep on giving me everything you have to give. I hope I am worthy of it, and that I can make you proud.

To God. Just thank you. There is nothing I can give that did not come to me from you.

Resumen

Los rendimientos financieros presentan con frecuencia una relación compleja con observaciones previas, así como una ligera asimetría y alta kurtosis. Como consecuencia, debemos utilizar modelos más flexibles que sean capaces de asir estos rasgos especiales: un proceso estocástico que sea capaz de manifestar la relación intertemporal entre las observaciones, así como una distribución que pueda capturar la asimetría y las colas pesadas de manera simultánea.

La distribución Gaussiana ha sido ampliamente utilizada en la literatura para estudiar diferentes tipos de datos; sin embargo, en algunos casos, como el estudio de rendimientos financieros, encontramos la necesidad de incorporar modelos más flexibles, debido a la frecuente presencia de un comportamiento ligeramente asimétrico y colas pesadas. Para enfrentar este problema, encontramos varias propuestas de familias de distribuciones.

Por ejemplo, Azzalini (1985) presentó la distribución Skew-Normal, que es capaz de capturar la falta de simetría subyacente en ciertos tipos de datos. Además, Lange y Sinsheimer (1993), a través de la distribución Slash, nos enseñan una manera de captar la kurtosis, una habilidad de gran importancia en este tipo estudios. Para unir estas dos características, se han presentado otras opciones, como la distribución Skew-Slash, propuesta por Wang y Genton (2006), que nos otorga la posibilidad de percibir tanto la asimetría como la presencia de colas pesadas.

En esta tesis, en primer lugar, se propone utilizar un proceso GARCH (Generalized Autoregressive Conditional Heteroskedastic), propuesto por Bollerslev (1986), con innovaciones Skew-Slash para modelar series temporales de rendimientos de datos financieros en el caso univariante y, en segundo lugar, un modelo de Correlación Condicional Dinámica, propuesto por Tse y Tsui (2002), con innovaciones Skew-Slash en el caso multivariante.

En el caso univariante, derivamos expresiones explícitas de los momentos de órdenes altos para la distribución propuesta, de modo que pueda mostrarse su capacidad para incorporar al mismo tiempo una ligera falta de simetría y alta kurtosis. Además, obtendremos los estimadores máximo verosímiles y propondremos un procedimiento de inferencia Bayesiana para el modelo GARCH con innovaciones Skew-Slash e ilustramos el desempeño de la metodología propuesta utilizando tanto simulaciones como un ejemplo de datos reales a través de los log-rendimientos del índice de Standard & Poor's desde el 3 de enero de 2000 hasta el 28 de diciembre de 2013.

Para el caso multivariante, proponemos una extensión adecuada a través del modelo de Correlacion Condicional Dinámica con innovaciones Skew-Slash multivariantes para trabajar con rendimientos financieros desde el punto de vista Bayesiano, el cual se ilustra ejecutando un algoritmo MCMC (Markov Chain Montecarlo) para datos simulados, así como datos reales tomados de los precios de cierre diarios de los índices de Dow Jones y NASDAQ desde el 2 de enero de 1996 hasta el 29 de diciembre de 2006 en un primer ejemplo, además de los log-rendimientos diarios de los índices DAX, CAC40 y Nikkei desde el 10 de octubre de 1991 hasta el 30 de diciembre de 1997 en un segundo ejemplo.

Abstract

Financial returns often present a complex relation with previous observations, along with a slight skewness and high kurtosis, which can not typically be captured via Gaussian distributions. As a consequence, we need to develop flexible models that are able to capture these features. To respond to this problem, several families of distributions have been proposed.

For example, Azzalini (1985) presented the Skew-Normal distribution, which is able to capture the underlying skewness. Also, Lange and Sinsheimer (1993) show us a way to pick up the kurtosis by means of the Slash distribution, a much needed feature in this type of study. Other distributions such as the Skew-Slash proposed by Wang and Genton (2006) allow us to capture both skewness and heavy tails.

In this thesis, we begin by proposing the use of a Generalized Autoregressive Conditional Heteroskedastic (GARCH) process, introduced by Bollerslev (1986), with Skew-Slash innovations to model univariate financial time series of returns. In this case, we derive formulae for the higher order moments of this distribution, which show that this distribution can incorporate both moderate skewness and high kurtosis. We also obtain the Maximum Likelihood estimations and we propose a Bayesian inference procedure for the GARCH model with Skew-Slash innovations, and illustrate the performance of our proposed methodology using simulations, as well as a real data example using the log-returns of the Standard & Poor's index from January 3^{rd}, 2000 to December 28^{th}, 2013.

Afterwards, for the multivariate case, we propose to use an extension of the GARCH process, such as the Dynamic Conditional Correlation model, introduced by Tse and Tsui (2002), with multivariate Skew-Slash innovations for financial returns in a Bayesian framework, and it is illustrated using a Markov Chain Montecarlo (MCMC) within Gibbs algorithm performed on simulated data, as well as real data drawn from the daily closing prices of the Dow Jones and NASDAQ indices from January 2^{nd}, 1996 until December 29^{th}, 2006 on a first example, and the daily log-returns of the DAX, CAC40, and Nikkei indices between October 10^{th}, 1991 and December 30^{th}, 1996 in a second example.

Chapter 1

Introduction

Modeling financial data sets can be a complicated task because working with financial returns often presents the challenge of having to model data with a complex relation to previous observations. Besides this, we often find that the residuals, after fitting a suitable model, usually exhibit a slight lack of symmetry as well as a high kurtosis.

Financial returns often reflect a structure that may be reasonably explained with conditional heteroskedastic models, such as the Generalized Autoregressive Conditional Heteroskedastic (GARCH) process, proposed by Bollerslev (1986), the Exponential Generalized Autoregressive Conditional Heteroskedastic (EGARCH) model of Nelson (1991), the Glosten - Jagannathan - Runkle GARCH (GJR-GARCH) model by Glosten, Jagannathan, and Runkle (1993), or the Threshold GARCH (TGARCH) model by Zakoian (1994), in the univariate case.

For the multivariate case, it is important to consider an appropriate generalization, and several proposals can be found in the literature. First, Dynamic Conditional Correlation Models have been proposed by Tse and Tsui (2002)

– that will be used later in this thesis –; Bollerslev (1990); Jeantheau (1998); Engle and Sheppard (2001); Engle (2002); Cappiello, Engle, and Sheppard (2006); Billio, Caporin, and Gobbo (2006), and Aielli (2013).

Other models for multiple returns have been proposed. Among others, Bollerslev, Engle, and Woolridge (1988) propose a VEC model based on conditional covariance matrices; Engle, Granger, and Kraft (1984) present a multivariate ARCH model; Engle, Ng, and Rothschild (1990) present a factor structure for the conditional covariance matrix; Engle and Kroner (1995) present a restricted version of the VEC model with their Baba-Engle-Kraft-Kroner (BEKK) model; Alexander and Chibumba (1997) present the Orthogonal GARCH (O-GARCH) model, later geneneralized by van der Weide (2002) with the Generalized Orthogonal GARCH (GO-GARCH) model; Vrontos, Dellaportas, and Politis (2003) present their Full Factor GARCH model, and Kawakatsu (2006) presents a generalization to the EGARCH model of Nelson (1991) by means of the Matrix Exponential GARCH model. See Tsay (2010) for a detailed review on financial return models. See also Silvennoien and Teräsvirta (2008) for an extended review on multivariate GARCH models.

In most of the previous models, it is assumed that the standardized residuals of the models follow a Gaussian distribution. This is a very appealing tool in statistical and probabilistic modeling; nevertheless, this type of model is inevitably binded to symmetry and does not allow for heavy tails, among other properties. Because of this, in the past few decades, the necessity of more flexible distributions has become patent.

If we focus particularly on the skewness and kurtosis, that are precisely the main characteristics that are very often found in financial time series of returns, we can find some alternative families of distributions proposed in the literature. For instance, Azzalini (1985) developed the Skew-Normal distribution, later generalized to the multivariate case by Azzalini and Dalla Valle (1996) and by Arellano-Valle, Bolfarine, and Lachos (2005), that allows the presence of skewness. On the other hand, one way to capture thick tails is presented by Lange and Sinsheimer (1993) with the family of Normal/independent distributions. Each one of these distributions present certain advantages; however, they lack the possibility of capturing both skewness and kurtosis at the same time, and we should notice that these characteristics can certainly present concurrently in some data sets, just like it happens in the case of interest displayed in the present thesis.

To solve this problem, we can find several proposals, a number of which can be included as particular cases of the multivariate family of distributions presented by Lachos, Labra, and Ghosh (2007): the family of Skew-Normal/independent distributions. The elements in this family combine the properties of the Skew-Normal distribution with the idea presented by Lange and Sinsheimer (1993) with the Normal/independent family of distributions. As a result, they obtain a new family of distributions that are more flexible than the Gaussian one and allow us to capture skewness and heavy tails simultaneously.

As it could be expected, the case where the stochastic component of a conditional heteroskedastic process is assumed to be Gaussian has been widely explored. However, it does not seem to be able to capture the essence of the returns in an adequate way because, as we have expressed before, the residuals after fitting conditional heteroskedastic models usually exhibit moderate 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 in a more appropriate manner.

In the financial literature, 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.

Some 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. Let us mention that these two distributions can be understood as particular cases of the Skew-Normal/independent family.

An interesting feature of the Skew-Slash distribution is that it is easy to simulate observations of this distribution by means of its alternative stochastic representation because we would only need to generate Gaussian random variables and a Beta random variable, and there are algorithms widely available to undertake this task. This is important to perform Bayesian inference as it is done in chapters 3 and 4. Also, we may use the analytical expressions for the mean and variance to ensure a structure that meets the restrictions required by this kind of process of null mean and unit variance. As well as the mentioned perks, we find that the Skew-Slash distribution consists on an infinite scale-mixture of Skew-Normal distributions, which allows the assumption that the variance is not fixed for all the members of the population, and we could not possibly believe that all of the factors that affect a financial asset have the same variance, or that they always have the same strength to affect the financial returns. Additionally, the multivariate Skew-Slash distribution presents with the advantage of being closed under linear transformations. This means that not only are linear transformations of this variables also distributed according to a Skew-Slash, but also the marginal distributions are ensured to follow this same probability model, as shown in Proposition 3 of Wang and Genton (2006).

Given these findings, we propose the Skew-Slash distribution, presented by Wang and Genton (2006), as an alternative model for the innovations in a GARCH model or a Dynamic Conditional Correlation model, depending on the framework we are working in (univariate or multivariate). Additionally, let us notice that the Skew-Slash distribution has also been applied successfully to describe data with similar characteristics in other settings; see e.g. Lachos, Garibay, Labra, and Aoki (2009).

For the univariate case, we consider the fact that inference for financial return models can be carried out using either the maximum likelihood or the Bayesian approach, and we shall consider both methods. Firstly, maximum likelihood estimation is carried out using a direct constrained optimization algorithm. Second, we introduce an approach to Bayesian inference for the Skew-Slash distribution which is based on that of Lachos, Dey, and Cancho (2009) with some modifications that enable us to account for the moment restrictions inherent to the GARCH framework.

For the multivariate case, 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 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 thesis is structured as follows. In chapter 2, we introduce the univariate and multivariate versions of the Skew-Slash distribution along with some of its properties and moments, as a particular member of the Skew-Normal/independent family of distributions, and we also show the way in which this family is constructed, along with some of its properties as a whole. In chapter 3, the GARCH process with univariate Skew-Slash innovations is detailed together with the methodologies implemented to perform inference from the Maximum Likelihood and Bayesian points of view; to illustrate the performance of the present models, we present the fitting of the parameters in simulation studies, and we also estimate a model to analyze the Standard & Poor's index between January 3rd, 2000 and December 28th, 2013. In chapter 4, we outline the Dynamic Conditional Correlation model with multivariate Skew-Slash innovations and we present the implementation of a Bayesian methodology to perform the inference; to illustrate the performance of the model we propose, as well as its respective methodology, we fit the parameters of a 2-dimensional and a 3-dimensional simulated data sets, besides showing the estimation results of the analysis of two real data sets: firstly, we study the Dow Jones and NASDAQ returns between January 2nd, 1996 and December 29th, 2006; afterwards, we work with the log-returns of the DAX, CAC40, and Nikkei indices from October 10th, 1991 until December 30th, 1997. Finally, we present our conclusions and some future lines of research in chapter 5.

Chapter 2

Some flexible distributions

In the present chapter, we present a review of a number of distributions that are generalizations or extensions of the Gaussian distribution with the purpose of allowing for more flexibility in their structure. In particular, we focus on the possibility of capturing the skewness and kurtosis underlying in certain types of data sets.

To do this, we will take up a journey that begins in the normal or Gaussian distribution and leads to a better understanding of the Skew-Slash distribution, proposed by Wang and Genton (2006), seen as a particular member of the Skew-Normal/independent family of distributions, defined by Lachos, Labra, and Ghosh (2007).

In the pursuit of a better comprehension of the Skew-Slash distribution, we will be interested in its structure, certain moments and an alternative stochastic representation that allows us to simulate observations from it in an easy way, as well as playing an important role in the design of a Bayesian methodology for the inference in models that might incorporate the use of this distribution.

2.1 Univariate distributions

We have mentioned before that the uses of the Gaussian distribution have been widely explored in different frameworks. In fact, it is a fundamental tool in probability and statistics and, even though it is not a very flexible distribution because of its symmetry and the fact that it does not allow for very heavy tails, among other properties, it is still a fundamental distribution in the construction of other distributions that allow for more flexibility. In the present section, we will review the Gaussian distribution as well as other more flexible distributions based on it, on our way of defining the Skew-Slash distribution, that is the center of the present thesis.

2.1.1 Normal distribution

We say that a random variable, X, follows a normal or Gaussian distribution with location parameter η and dispersion parameter σ, denoted by $X \sim \mathcal{N}_1(\eta, \sigma)$, if its probability density function is given by

$$\phi_1(x|\eta,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\eta}{\sigma}\right)^2\right\}; x \in \mathbb{R},$$

where $\eta \in \mathbb{R}$ and $\sigma \geq 0$.

Let us notice that, if $X \sim \mathcal{N}_1(\eta, \sigma)$, then its mean is given by

$$E(X) = \eta,$$

its variance is given by

$$V(X) = \sigma^2,$$

its skewness coefficient is given by

$$S(X) = 0,$$

and kurtosis coefficient given by

$$K(X) = 3.$$

As we can see, the Gaussian distribution is symmetric around its mean, and has a fixed kurtosis, regardless of the value of its parameters.

It is well known that it is not possible to find a closed form for the cumulative distribution function of the normal distribution, but, if X is a standard normal random variable, i.e., $X \sim \mathcal{N}_1(0, 1)$, then we will denote its cumulative distribution function as $\Phi_1(x)$. In this case, we can also compact the notation for the standard normal density and just write $\phi_1(x)$.

2.1.2 Student's-t distribution

Another recurred distribution that allows for heavier tails than the Gaussian one is the Student's-t distribution.

We say that a random variable, V, follows a Student's-t distribution with location parameter η, scale parameter σ, and kurtosis parameter ν (also known as the degrees of freedom), denoted by $V \sim \mathcal{T}_1(\eta, \sigma, \nu)$, if its probability density function is given by

$$t_1(v|\eta, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma}\left[1 + \nu^{-1}\left(\frac{v-\eta}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}} ; w \in \mathbb{R},$$

where $\eta \in \mathbb{R}$, $\sigma \geq 0$, and $\nu \in \mathbb{R}^+$, and we will denote its cumulative distribution function as $T_1(v|\eta, \sigma, \nu)$.

The Student's-t distribution admits an alternative stochastic representation, given by

$$V \equiv \eta + \sqrt{\nu}U^{-1/2}X,$$

where $\equiv$ signifies equivalence in distribution, and $U \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{1}{2}\right)$, independent from X, a normal random variable with zero mean and standard deviation σ.

Analogously, we can write

$$V \equiv \eta + U^{-1/2} X,$$

where $U \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ independent from $X \sim \mathcal{N}_1(0, \sigma)$.

Let us notice that, if $V \sim \mathcal{T}_1(\eta, \sigma, \nu)$, then its mean is given by

$$E(V) = \eta,$$

its variance is given by

$$V(V) = \frac{\nu}{\nu - 2}\sigma^2 \text{ for } \nu > 2,$$

its skewness coefficient is given by

$$S(V) = 0,$$

and its kurtosis coefficient is given by

$$K(V) = 3\frac{(\nu - 1)^2}{(\nu - 2)(\nu - 4)} \text{ for } \nu > 4.$$

As we can see, the Student's-t distribution is symmetric, but, unlike the Gaussian distribution, it is able to present different levels of kurtosis, depending on its parameters. Furthermore, let us notice that the Student's-t distribution is a particular member of the Normal/independent family of distributions, that will be detailed in section 2.1.4.

2.1.3 Slash distribution

According to the definition given by Lange and Sinsheimer (1993), we say that a random variable, V, follows a Slash distribution with location parameter η, scale parameter σ, and kurtosis parameter ν, denoted as $V \sim \mathcal{SL}_1(\eta, \sigma, \nu)$, if its probability density function is given by

$$f_V(v) = \nu \int_0^1 u^{\nu-1} \phi_1\left(v|\eta, u^{-1}\sigma\right) du; v \in \mathbb{R},$$

where $\eta \in \mathbb{R}$, $\sigma \geq 0$, and $\nu > 0$.

This distribution admits an alternative stochastic representation given by

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

with $U \sim \mathcal{B}e(\nu, 1)$ independent from $X \sim \mathcal{N}_1(0, \sigma)$.

Let us notice that, if $V \sim \mathcal{SL}_1(\eta, \sigma, \nu)$, then its mean is given by

$$E(V) = \eta,$$

its variance is given by

$$V(V) = \sigma^2 \frac{\nu}{\nu - 2} \text{ for } \nu > 2,$$

its skewness coefficient is given by

$$S(V) = 0,$$

and its kurtosis coefficient is given by

$$K(V) = 3 \frac{(\nu - 2)^2}{\nu(\nu - 4)} \text{ for } \nu > 4.$$

As we can see, the Slash distribution is always symmetrical, and allows for different kurtosis coefficients, depending on the value of ν. In fact, it has a similar structure than the Student's-t distribution because, just like it, the Slash distribution is also a particular member of the Normal/independent family of distributions, detailed below.

2.1.4 Normal/independent family of distributions

We say that a random variable, V, follows a Normal/independent distribution, as defined by Lange and Sinsheimer (1993), with location parameter η and scale parameter σ if its probability density function is given by

$$f_V(v) = \int_{\mathbb{R}} \phi_1\left(v|\eta, u^{-1}\sigma\right) dH(u); v \in \mathbb{R},$$

where $\eta \in \mathbb{R}$, $\sigma \geq 0$, and $H(u|\nu)$ is the cumulative distribution function of a unidimensional positive random variable U, indexed by the parameter ν, and we denote it as $V \sim \mathcal{NI}_1(\eta, \sigma; H)$.

This distribution also offers the possibility of an alternative stochastic representation, given by

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

where $X \sim \mathcal{N}_1(0, \sigma)$. Also, U and X are independent.

Let us notice that, if $V \sim \mathcal{NI}_1(\eta, \sigma; H)$, then we can find some general expressions for several central moments. Its mean is given by

$$E(V) = \eta,$$

its variance is given by

$$V(V) = \sigma^2 E\left(U^{-2}\right)$$

if $E(U^{-2}) < \infty$, its skewness coefficient is given by

$$S(V) = 0,$$

and its kurtosis coefficient is given by

$$K(V) = 3\frac{E(U^{-4})}{E^2(U^{-2})}$$

if $E\left(U^{-4}\right)<\infty$.

As we can see, this family of distributions always presents a symmetric behavior, but allows for different structures of the tails, and the kurtosis coefficient is entirely defined by the behavior of the so-called independent variable, U.

We must acknowledge that the Normal/independent family of distributions includes the Student's-t and the Slash, as we mentioned before, besides the Power Exponential, and the Contaminated Normal distributions, among others, all of which have heavier tails than the Gaussian distribution.

2.1.5 Skew-Normal distribution

According to Azzalini and Dalla Valle (1996),we say that a random variable, Z, follows a Skew-Normal distribution with location parameter η, scale parameter σ, and skewness parameter λ, denoted as $Z \sim \mathcal{SN}_1\left(\eta,\sigma,\lambda\right)$, if its probability density function is given by

$$f_Z\left(z\right)=2\phi_1\left(z|\eta,\sigma\right)\Phi_1\left(\lambda\frac{z-\eta}{\sigma}\right);z\in\mathbb{R},$$

where $\eta\in\mathbb{R}$, $\sigma\geq 0$, and $\lambda\in\mathbb{R}$. See also O'Hagan and Leonard (1976) and Azzalini (1985).

Let us notice that, in particular, positive values of λ imply positive skewness, while negative values of λ imply negative skewness, and the distribution reduces to the normal when $\lambda=0$.

One useful property of the Skew-Normal distribution is that it admits an alternative stochastic representation, as shown in Henze (1986), given by

$$Z\equiv\eta+\sigma\left(\delta\left|X_0\right|+\sqrt{1-\delta^2}X_1\right), \tag{2.1}$$

where $\delta = \lambda/\sqrt{1+\lambda^2}$, $|\delta| < 1$, and X_0 and X_1 are independent identically distributed standard normal random variables.

Let us notice that, if $Z \sim \mathcal{SN}_1(\eta, \sigma, \lambda)$, then its mean is given by

$$E(Z) = \eta + \sqrt{\frac{2}{\pi}}\sigma\delta,$$

its variance is given by

$$V(Z) = \sigma^2\left(1 - \frac{2}{\pi}\delta^2\right),$$

its skewness coefficient is given by

$$S(Z) = \sqrt{\pi}\frac{4-\pi}{\left(\pi - 2\delta^2\right)^{3/2}}\delta^3,$$

and its kurtosis coefficient is given by

$$K(Z) = \frac{3\pi^2 - 12\pi\delta^2 + (8\pi - 12)\,\delta^2}{\left(\pi - 2\delta^2\right)^2}.$$

As we can see, this is a more flexible distribution, that allows for both diverse skewness and kurtosis coefficients, but we must consider that δ alone, or equivalently λ, determines both coefficients entirely.

2.1.6 Skew-t distribution

We say that a random variable, W, follows a Skew-t distribution, as defined by Azzalini and Capitanio (2003), with location parameter η, scale parameter σ, location parameter λ, and kurtosis parameter ν, and denote it as $W \sim \mathcal{ST}_1(\eta, \sigma, \lambda, \nu)$, if its probability density function is of the form

$$f_W(w) = 2t_1(w|\eta, \sigma, \nu)\, T_1\left(\left.\frac{\sqrt{\nu+1}\lambda(w-\eta)}{\sigma\sqrt{\sigma^2(w-\eta)^2 + \nu}}\right| 0, 1, \nu+1\right); w \in \mathbb{R},$$

where $\eta \in \mathbb{R}$, $\sigma \geq 0$, $\lambda \in \mathbb{R}$, and $\nu \in \mathbb{R}^+$.

The Skew-t distribution also admits an alternative stochastic representation, given by

$$W \equiv \eta + U^{-1}Z,$$

where $Z \sim \mathcal{SN}_1(0, \sigma, \lambda)$ independent from U with probability density function given by

$$f_U(u) = \frac{2\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} u^{\nu-1} \exp\left\{-\frac{\nu}{2}u^2\right\} \mathbf{I}(u > 0).$$

In other words, $U^2 \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$.

Let us notice that, if $W \sim \mathcal{ST}_1(\eta, \sigma, \lambda, \nu)$, then its mean is given by

$$E(W) = \eta + \sqrt{\nu\pi}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\sigma\delta,$$

for $\nu > 1$, its variance is given by

$$V(W) = \sigma^2\left\{\frac{\nu}{\nu-2} - \frac{\nu}{\pi}\left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^2 \delta^2\right\},$$

for $\nu > 2$, its skewness coefficient is given by

$$S(W) = \sqrt{\frac{\nu}{\pi}}\delta\frac{3\Gamma^2\left(\frac{\nu}{2}\right)a_{11} + \delta^2\frac{\nu}{2}a_{21}}{\left[\frac{\nu}{\nu-2}\Gamma^2\left(\frac{\nu}{2}\right) - \frac{\nu}{\pi}\delta^2\Gamma^2\left(\frac{\nu-1}{2}\right)\right]^{3/2}}, \tag{2.2}$$

for $\nu > 3$, where

$$a_{11} = \frac{\nu}{2}\Gamma\left(\frac{\nu-3}{2}\right) - \frac{\nu}{\nu-2}\Gamma\left(\frac{\nu-1}{2}\right)$$

and

$$a_{21} = \frac{4}{\pi}\Gamma^3\left(\frac{\nu-1}{2}\right) - \Gamma^2\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu-3}{2}\right)$$

and its kurtosis coefficient is given by

$$K(W) = \frac{3\frac{\nu^2}{(\nu-2)(\nu-4)}\Gamma^4\left(\frac{\nu}{2}\right) + 6\frac{\nu}{\pi}\delta^2 a_{12} + \frac{\nu^2}{\pi}\delta^4 a_{22}}{\nu^2\left[\frac{1}{(\nu-2)^2}\Gamma^4\left(\frac{\nu}{2}\right) - \frac{2}{\pi(\nu-2)}\delta^2\Gamma^2\left(\frac{\nu}{2}\right)\Gamma^2\left(\frac{\nu-1}{2}\right) + \frac{1}{\pi^2}\delta^4\Gamma^4\left(\frac{\nu-1}{2}\right)\right]}, \tag{2.3}$$

for $\nu > 4$, where

$$a_{12} = \Gamma^2\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu-1}{2}\right)\left[\frac{\nu}{\nu-2}\Gamma\left(\frac{\nu-1}{2}\right) - \nu\Gamma\left(\frac{\nu-3}{2}\right)\right],$$

$$a_{22} = \Gamma\left(\frac{\nu-1}{2}\right)\left[2\Gamma^2\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu-3}{2}\right) + \frac{3}{\pi}\Gamma^3\left(\frac{\nu-1}{2}\right)\right].$$

As we can see, the Skew-t distribution allows for a much more flexible structure than the distributions that we have reviewed up until this moment because, depending on the values of λ and ν, we can have different values of the skewness and kurtosis coefficients, just like the Skew-Slash distribution, that is detailed below.

Furthermore, let us notice that both the Skew-t distribution and the Skew-Slash distribution are both members of the Skew-Normal/independent distribution, that will be explained in section 2.1.8.

To illustrate the ability of generating asymmetric densities, as well as distributions with heavy tails, in Figure 2.1 we show the value of the skewness coefficient of the Skew-t distribution for different values of $\lambda \in [-10, 10]$ and $\nu \in [3.1, 10]$, in the same way that Figure 2.2 illustrates the kurtosis coefficient for values of $\lambda \in [-10, 10]$ and $\nu \in [4.1, 10]$.

In Figure 2.1, we can see that, except for values of ν lower than 4, the skewness coefficient is negligible, and it does not change importantly when the parameters stay away from the asymptotic behavior. On the other hand, in Figure 2.2 we find that λ does not affect highly the values of the kurtosis coefficient.

2.1.7 Skew-Slash distribution

According to Wang and Genton (2006) we say that, if Z is a Skew-Normal random variable and U is an independent, beta distributed random variable,

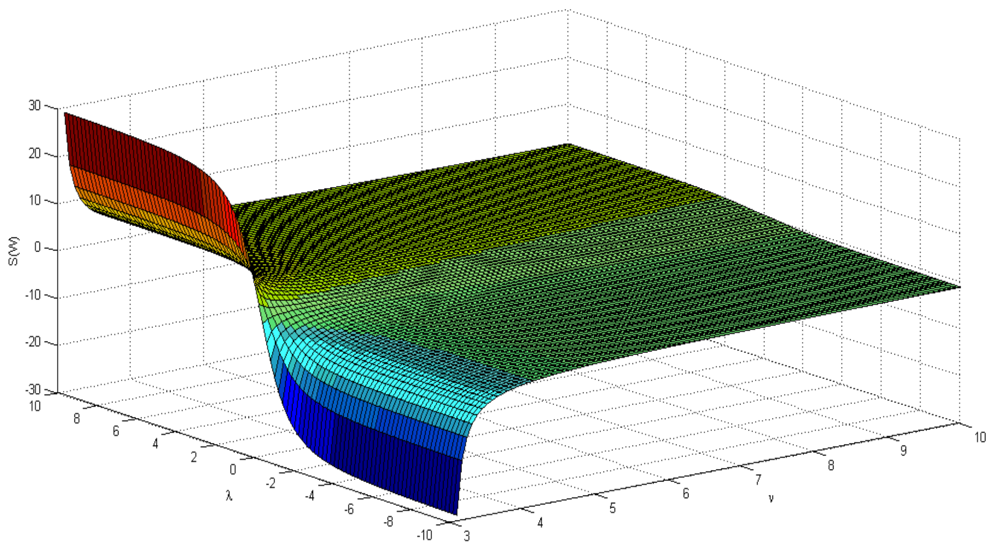


Figure 2.1: Skewness coefficient of the Skew-t distribution as a function of λ and ν

$U \sim \mathcal{B}e\,(\nu, 1)$, where $\nu > 0$, then we will say that $W \equiv U^{-1}Z$ follows a standard Skew-Slash distribution, denoted as $W \sim \mathcal{SSL}_1\,(\lambda, \nu)$[1], with $\lambda \in \mathbb{R}$.

We can also extend this notion to a four parameter distribution. We say that a random variable, W, follows a Skew-Slash distribution with location parameter η, scale parameter σ, skewness parameter λ, and kurtosis parameter ν if

$$W \equiv \eta + \sigma U^{-1} Z,$$

where $Z \sim \mathcal{SN}_1\,(0, 1, \lambda)$, independent from U. In this case, we shall write $W \sim \mathcal{SSL}_1\,(\eta, \sigma, \lambda, \nu)$.

The probability density function of a Skew-Slash random variable is given

[1]In fact, Wang and Genton (2006) define the Skew-Slash variable using $U^{-1/2}$, but this definition is equivalent.

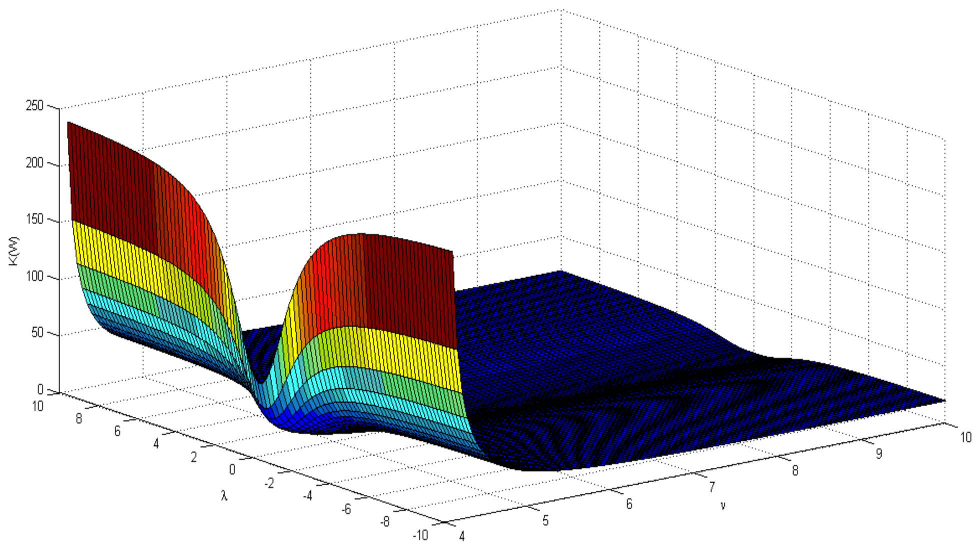


Figure 2.2: Kurtosis coefficient of the Skew-t distribution as a function of λ and ν

by

$$f_W(w) = \int_0^1 2\nu u^{\nu-1}\phi_1\left(w|\eta, u^{-2}\sigma^2\right)\Phi_1\left(\frac{\lambda u\,(w-\eta)}{\sigma^2}\right)du; w \in \mathbb{R},$$

where $\eta \in \mathbb{R}$, $\sigma \geq 0$, $\lambda \in \mathbb{R}$, and $\nu \in \mathbb{R}^+$.

Consequently, we can see that the Skew-Slash distribution is a scale-mixture of a variable with a Skew-Normal distribution.

A more elaborate version of an alternative stochastic representation for the Skew-Slash distribution is given by

$$W \equiv \eta + \sigma U^{-1}\left(\delta\,|X_0| + \sqrt{1-\delta^2}X_1\right), \tag{2.4}$$

where $\delta = \lambda/\sqrt{1+\lambda^2}$, which implies that $-1 < \delta < 1$. As we have mentioned before, this representation will be very useful in the context of Bayesian inference.

Moments of the univariate Skew-Slash distribution

Although expressions for the mean and variance of the Skew-Slash distribution were given in Wang and Genton (2006), they do not provide explicit expressions for the higher order moments. In order to show that the Skew-Slash distribution is able to capture both moderate skewness and high kurtosis, we shall derive the higher order moments below. To do this, the results given in the following lemma are useful.

Lemma 1 *For $j \in \mathbb{N}$,*

1. *If $U \sim \mathcal{B}e\left(\nu, 1\right)$, then*

$$E\left(U^{-j}\right) = \frac{\nu}{\nu - j}, \qquad \text{for } \nu > j.$$

2. *If $Z \sim \mathcal{SN}_1\left(\lambda\right)$, then*

$$E\left(Z^j\right) = 2^{(j-2)/2}\ \pi^{-1}\left(\frac{1}{1+\lambda^2}\right)^{j/2}\sum_{k=0}^{j} a_{kj}\lambda^k,$$

where

$$a_{kj} = \binom{j}{k}\left(1 + (-1)^{j-k}\right)\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{j-k+1}{2}\right),$$

for $k = 0, 1, \ldots, j$. In particular, $a_{kj} = 0$ if $j - k$ is an odd number.

Proof

1. For $j < \nu$, $E\left(U^{-j}\right) = B\left(\nu, 1\right)/B\left(\nu - j, 1\right)$, where $B(\cdot, \cdot)$ is the beta function and the result follows.

2. See e.g. Henze (1986).

■

These results can be used to derive the expectation and central moments of the Skew-Slash distribution, as summarized in the following proposition. The proof for the mean is available in Wang and Genton (2006) and the proof of the formula for the central moments is given in the Appendix A.

Proposition 2 *If* $W \sim \mathcal{SSL}_1(\eta, \sigma, \lambda, \nu)$, *then*

1. *The expectation of* W *is given by*

$$E(W) = \eta + \sigma \left(\frac{2}{\pi}\right)^{1/2} \delta \, \frac{\nu}{\nu - 1}, \tag{2.5}$$

for $\nu > 1$.

2. *For* $k \in \mathbb{N}$, *the central moments of the Skew-Slash distribution are given by*

$$m_j[W] = E\Big\{[W - E(W)]^j\Big\} = 2^{(j-2)/2}\sigma^j \left(\frac{1}{1+\lambda^2}\right)^{j/2} \sum_{l=0}^{j} c_{lj}\lambda^l, \tag{2.6}$$

for $\nu > j$, *where*

$$c_{lj} = \sum_{m=0}^{l} b_{m,j-l+m,j},$$

and

$$b_{m,j-l+m,j} = (-1)^{l-m}\left[1 + (-1)^{j-l}\right]\pi^{-\frac{l-m+2}{2}} \frac{\nu}{\nu - (k-l+m)} \left(\frac{\nu}{\nu-1}\right)^{l-m} \binom{j}{l}\binom{l}{m}\Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{j-l+1}{2}\right)$$

for fixed j, *fixed* $l \in \{0, 1, \ldots, j\}$, *and* $m \in \{0, 1, \ldots, l\}$, *respectively. In particular, note that* $c_{lj} = 0$ *if* $j - l$ *is an odd number.*

Given the general expression of the central moments of the Skew-Slash distribution given in (2.6), it is straightforward to obtain the variance and the skewness and kurtosis coefficients of W, as summarized in the following Corollary. Note that the formula for the variance was also derived in Wang and Genton (2006).

Corollary 3

1. *The variance of the Skew-Slash distribution is given by*

$$V(W) = m_2[W] = \sigma^2 \frac{c_{02} + c_{22}\lambda^2}{1+\lambda^2},$$

where

$$c_{02} = \frac{\nu}{\nu-2}, \tag{2.7}$$

$$c_{22} = \frac{\nu}{\nu-2} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2. \tag{2.8}$$

Thus,

$$V(W) = \sigma^2 \left(\frac{\nu}{\nu-2} - \delta^2 \left(\frac{\nu}{\nu-1}\right)^2 \frac{2}{\pi}\right) \quad \text{for } \nu > 2. \tag{2.9}$$

2. *The skewness coefficient of the Skew-Slash distribution is given by*

$$S(W) = \frac{m_3[W]}{m_2[W]^{3/2}} = 2^{1/2} \frac{c_{13}\lambda + c_{33}\lambda^3}{(c_{02} + c_{22}\lambda^2)^{3/2}}, \tag{2.10}$$

for $\nu > 3$, where c_{02} and c_{22} are given in (2.8) and

$$c_{13} = \frac{3}{\pi^{1/2}}\left(\frac{\nu}{\nu-3} - \frac{\nu}{\nu-2}\frac{\nu}{\nu-1}\right),$$

$$c_{33} = \frac{4}{\pi^{3/2}}\left(\frac{\nu}{\nu-1}\right)^3 - \frac{3}{\pi^{1/2}}\frac{\nu}{\nu-2}\frac{\nu}{\nu-1} + \frac{2}{\pi}\frac{\nu}{\nu-3}.$$

3. The kurtosis coefficient of the Skew-Slash distribution is given by

$$K(W) = \frac{m_4[W]}{m_2[W]^2} = 2\frac{c_{04} + c_{24}\lambda^2 + c_{44}\lambda^4}{(c_{02} + c_{22}\lambda^2)^2}, \tag{2.11}$$

for $\nu > 4$*, where* c_{02} *and* c_{22} *are given in (2.8) and*

$$c_{04} = \frac{3}{2}\frac{\nu}{\nu-4},$$

$$c_{24} = \frac{6}{\pi}\frac{\nu}{\nu-2}\left(\frac{\nu}{\nu-1}\right)^2 - \frac{12}{\pi}\frac{\nu}{\nu-3}\frac{\nu}{\nu-1} + 3\frac{\nu}{\nu-4},$$

$$c_{44} = -\frac{6}{\pi^2}\left(\frac{\nu}{\nu-1}\right)^4 + \frac{12}{\pi}\frac{\nu}{\nu-2}\left(\frac{\nu}{\nu-1}\right)^2 - \frac{8}{\pi}\frac{\nu}{\nu-3}\frac{\nu}{\nu-1} + \frac{3}{2}\frac{\nu}{\nu-4},$$

respectively.

A direct consequence of these results is that the Skew-Slash distribution is able to generate moderate skewness and high kurtosis. To illustrate this, Figures 2.3 and 2.4 show some values of the skewness and the kurtosis coefficients for values of λ and ν in a grid of points uniformly distributed in the set $[-10, 10] \times [4.1, 10]$ and where η and σ are chosen to have zero mean and variance equal to 1.

Let us notice that the skewness and kurtosis coefficients do not depend on η or σ, but only on λ and ν. Also, observe that, as with the Skew-Normal distribution, the skewness coefficient is 0 for $\lambda = 0$, positive for positive values of λ, and negative for negative values of λ. Note also that ν has a small effect on the skewness. Also, the kurtosis coefficient gets larger as ν decreases. Besides, λ has only a small effect on the kurtosis. Also, it can be observed that, except for small values of ν between 4 and 5, the skewness changes very little with ν and, equally, the dependence of the kurtosis on λ is relatively low.

Additionally, Figure 2.5 shows the values of the skewness and kurtosis obtained for the points in the grid. The image shows that, even for moderate

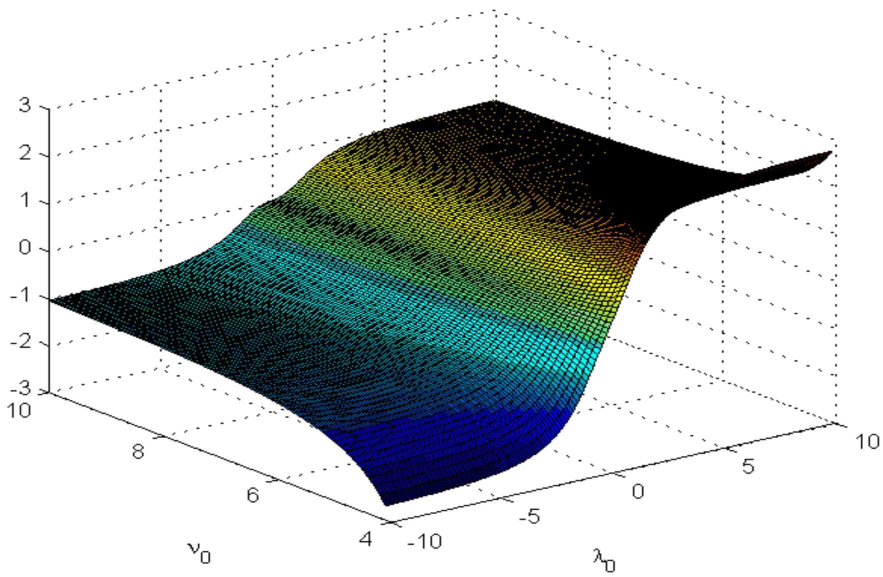


Figure 2.3: Skewness coefficient of the Skew-Slash distribution as a function of λ and ν

skewness, the kurtosis can be very high. Finally, notice that the Skew-Slash distribution is not able to generate very high skewness. However, this is not a drawback for the analysis of financial returns, in which the skewness is typically slight. In fact, this is precisely the kind of characteristic that we are looking for, as for financial returns the skewness is typically slight.

Let us notice that, as we mentioned before, just like the Skew-t distribution, the Skew-Slash distribution is a particular member of the Skew-Normal/independent family of distributions, that is detailed in section 2.1.8.

These two distributions clearly share a common structure, but they also have differences. As we can see from Figures 2.1 and 2.3, the structure of the skewness coefficient in both distributions is quite distinct, specially taking into account the fact that the Skew-t distribution remains almost constant

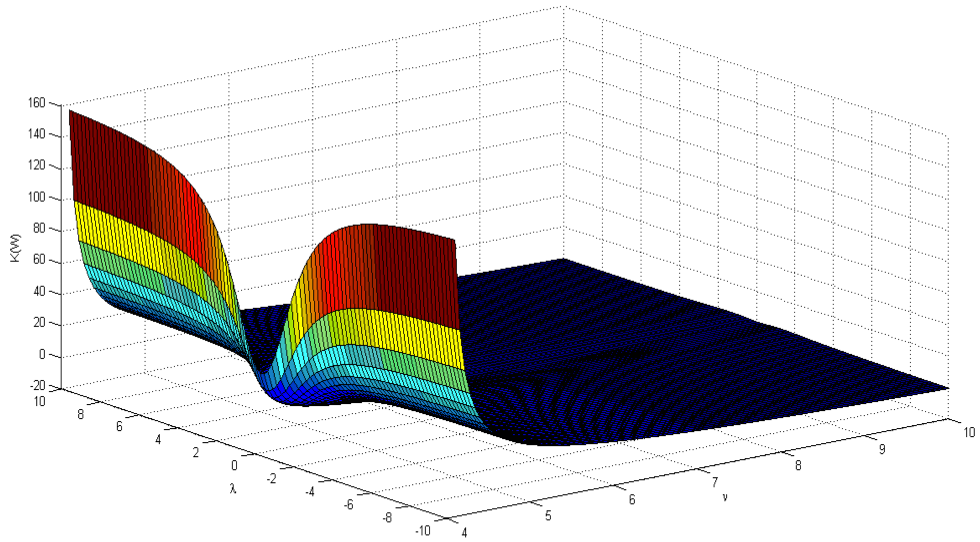


Figure 2.4: Kurtosis coefficient of the Skew-Slash distribution as a function of λ and ν

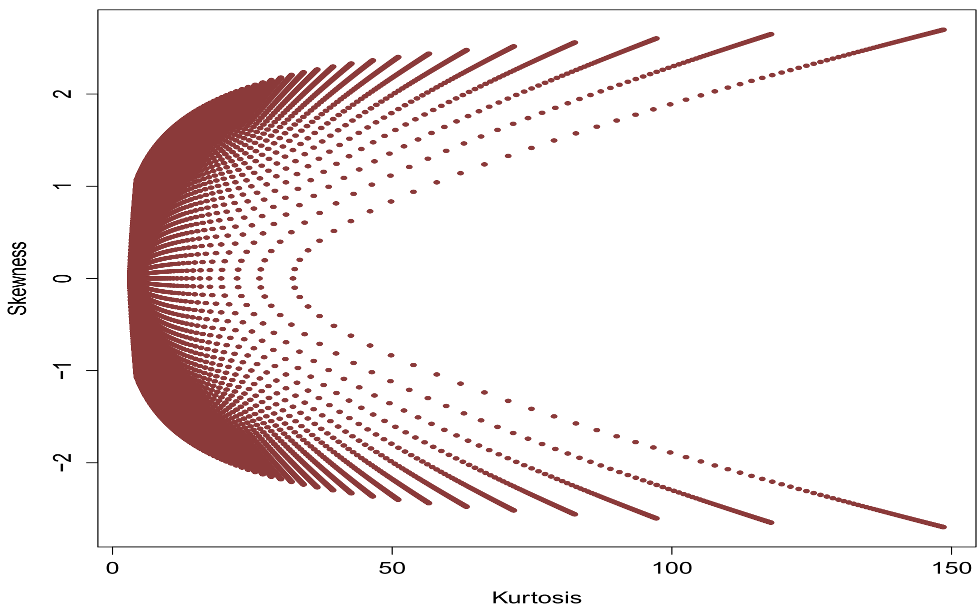


Figure 2.5: Values of the skewness and kurtosis for the Skew-Slash distribution

and very close to 0 when the parameter values are away from the asymptotic behaviors. On the other hand, Figures 2.2 and 2.4 show us that the structure of the kurtosis coefficient is quite similar in both cases, with the difference that the Skew-t distribution appears to allow for heavier tails without the need of getting too close to the asymptote. Furthermore, we can see that the structure of the higher order moments shown in (2.2) and (2.10) for the skewness coefficients, as well as in (2.3) and (2.11) for the kurtosis coefficients, is a bit more simple for the Skew-Slash distribution.

2.1.8 Skew-Normal/independent family of distributions

A generalization of the sort of distributions that we have studied so far, that allows for the presence of skewness and kurtosis, can be found in Lachos, Labra, and Ghosh (2007), and we now take into account a particular case, where our random variables are unidimensional. They propose a new family of distributions: the Skew-Normal/independent.

Considering the definition given by Lachos, Labra, and Ghosh (2007), and setting the univariate case as a particular one from their multivariate proposal, we say that a random variable, W, follows a Skew-Normal/independent distribution with location parameter η, scale parameter σ, and skewness parameter λ if its probability density function is of the form

$$f_W(w) = 2\int_{\mathbb{R}^+} \phi_1\left(w|\eta, u^{-1}\sigma\right) \Phi_1\left(u\lambda\frac{w-\eta}{\sigma}\right) dH(u); w \in \mathbb{R};$$

Its alternative stochastic representation is given by

$$W \equiv \eta + U^{-1}Z,$$

where $Z \sim \mathcal{SN}_1(0, \sigma, \lambda)$, and U and Z are independent[2].

An even more detailed stochastic representation that incorporates (2.1) leads to the following expression.

$$W \equiv \eta + \sigma\delta X + U^{-1}\sigma\sqrt{1-\delta^2}X_1,$$

where $X = U^{-1}|X_0|$, and X_0 and X_1 are independent identically distributed standard normal random variables, and $\delta = \lambda/\sqrt{1+\lambda^2}$; defined this way, $\delta \in (-1, 1)$.

Proposition 7 in the paper by Lachos, Labra, and Ghosh (2007) gives explicit expressions for the mean and variance of a Skew-Normal/independent random variable by explaining us that, if $W \sim \mathcal{SNI}_1(\eta, \sigma, \lambda; H)$, then

- If $E(U^{-1}) < \infty$, then

$$E(W) = \eta + \sqrt{\frac{2}{\pi}}E\left(U^{-1}\right)\sigma\delta;$$

- If $E(U^{-2}) < \infty$, then

$$V(W) = E\left(U^{-2}\right)\sigma^2 - \frac{2}{\pi}E^2\left(U^{-1}\right)\sigma^2\delta^2.$$

Besides this, if $W \sim \mathcal{SNI}_1(\eta, \sigma, \lambda; H)$, then its skewness coefficient is given by

$$S(W) = \left(\frac{2}{\pi}\right)^{1/2} \frac{3\delta\left[E(U^{-3}) - E(U^{-1})E(U^{-2})\right] + \delta^3\left[\frac{4}{\pi}E^3(U^{-1}) - E(U^{-3})\right]}{\left[E(U^{-2}) - \frac{2}{\pi}\delta^2E^2(U^{-1})\right]^{3/2}}$$

[2]Lachos, Labra, and Ghosh (2007) define the stochastic representation of the Skew-Normal/independent random variables or random vectors using $U^{-1/2}$ instead of U^{-1}, but this is an equivalent definition, and it is more convenient for the work we will develop with the Skew-Slash distribution.

if $E\left(U^{-3}\right)<\infty$, and its kurtosis coefficient is given by

$$K(W)=\frac{3E\left(U^{-4}\right)+\frac{12}{\pi}\delta^2 E\left(U^{-1}\right)b_1+\frac{4}{\pi}\delta^4 E\left(U^{-1}\right)b_2}{E^2\left(U^{-2}\right)-\frac{4}{\pi}\delta^2 E^2\left(U^{-1}\right)E\left(U^{-2}\right)+\frac{4}{\pi^2}\delta^4 E^4\left(U^{-1}\right)}$$

if $E\left(U^{-4}\right)<\infty$, where

$$b_1=E\left(U^{-1}\right)E\left(U^{-2}\right)-2E\left(U^{-3}\right),$$
$$\text{and}\quad b_2=2E\left(U^{-3}\right)+\frac{3}{\pi}E^3\left(U^{-1}\right).$$

2.2 Multivariate distributions

Before introducing the multivariate Skew-Slash distribution, we will first review the multivariate normal and the multivariate Student's-t distributions, and we will afterwards examine several flexible distributions, according to the concept presented by Lachos, Labra, and Ghosh (2007), for completeness.

2.2.1 Normal distribution

We say that a random vector, $\mathbf{X}$, follows a d-dimensional normal or Gaussian distribution with location parameter $\boldsymbol{\eta}$ and scale matrix $\boldsymbol{\Sigma}$, denoted by $\mathbf{X}\sim\mathcal{N}_d\left(\boldsymbol{\eta},\boldsymbol{\Sigma}\right)$, if its probability density function is given by

$$\phi_d\left(\mathbf{x}|\boldsymbol{\eta},\boldsymbol{\Sigma}\right)=\frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\eta}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\eta}\right)\right\};\mathbf{x}\in\mathbb{R}^d,$$

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

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

$$E\left(\mathbf{X}\right)=\boldsymbol{\eta},$$

and its variance-covariance matrix is given by

$$V\left(\mathbf{X}\right)=\boldsymbol{\Sigma}.$$

Let us also notice that it is well known that the normal distribution is always symmetric and does not allow for heavy tails.

2.2.2 Student's-t distribution

We say that a random vector, $\mathbf{V}$, follows a d-dimensional Student's-t distribution with location parameter $\boldsymbol{\eta}$, scale matrix $\boldsymbol{\Sigma}$, and kurtosis parameter ν (also known as the degrees of freedom), denoted as $\mathbf{V} \sim \mathcal{T}_d\left(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \nu\right)$, if its probability density function is given by

$$t_d\left(\mathbf{v}|\boldsymbol{\eta}, \boldsymbol{\Sigma}, \nu\right) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left(\pi\nu\right)^{d/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\left(1 + \frac{\left(\mathbf{v}-\boldsymbol{\eta}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{v}-\boldsymbol{\eta}\right)}{\nu}\right)^{-\frac{\nu+d}{2}} ; \mathbf{v} \in \mathbb{R}^d,$$

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

The Student's-t distribution admits an alternative stochastic representation, given by

$$V \equiv \boldsymbol{\eta} + \sqrt{\nu}U^{-1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{X},$$

where $U \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{1}{2}\right)$ independent from $\mathbf{X} \sim \mathcal{N}_d\left(\mathbf{0}, \mathbf{I}\right)$.

Analogously, we can write

$$\mathbf{V} \equiv \boldsymbol{\eta} + U^{-1/2}\mathbf{X},$$

where $U \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ independent from $\mathbf{X} \sim \mathcal{N}_d\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$.

Let us notice that, if $\mathbf{V} \sim \mathcal{T}_d\left(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \nu\right)$, then its mean is given by

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

and its variance covariance matrix is given by

$$V\left(V\right) = \frac{\nu}{\nu-2}\boldsymbol{\Sigma},$$

for $\nu > 2$.

Let us also notice that, even though the Student's-t distribution is symmetric like the Gaussian one, unlike the normal distribution, the Student's-t distribution allows for heavier tails. Furthermore, let us acknowledge that the Student's t distribution is a particular member of the Normal/independent family of distributions, that will be detailed in section 2.2.4.

2.2.3 Slash distribution

According to Lange and Sinsheimer (1993), we say that a d-dimensional random vector, $\mathbf{V}$, follows a Slash distribution with location parameter $\boldsymbol{\eta}$, scale matrix $\boldsymbol{\Sigma}$, and kurtosis parameter ν, denoted as $\mathbf{V} \sim \mathcal{SL}_d(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \nu)$, if its probability density function is given by

$$f_{\mathbf{V}}(\mathbf{v}) = \nu \int_0^1 u^{\nu-1} \phi_d\left(\mathbf{v} | \boldsymbol{\eta}, u^{-2}\boldsymbol{\Sigma}\right) du; \mathbf{v} \in \mathbb{R}^d,$$

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

This distribution also admits an alternative stochastic representation, given by

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

where we have that $U \sim \mathcal{B}e(\nu, 1)$ is independent from $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$.

Let us notice that, if $\mathbf{V} \sim \mathcal{SL}_d(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \nu)$, then its mean is given by

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

and its variance covariance matrix is given by

$$V(\mathbf{V}) = \frac{\nu}{\nu - 2}\boldsymbol{\Sigma},$$

for $\nu > 2$.

This is a distribution that, like the normal one, is symmetric, but it allows for heavier tails than the Gaussian and is, in fact, together with the Student's-t distribution, a member of the Normal/independent family of distributions that is exposed below.

2.2.4 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 probability density function is given by

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

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 of a unidimensional positive random variable U, indexed by the parameter ν, and we denote it as $\mathbf{V} \sim \mathcal{NI}_d(\boldsymbol{\eta}, \boldsymbol{\Sigma}; H)$. 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.

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 is comprehended by 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.

2.2.5 Skew-Normal distribution

According to the definition given by 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 $\mathbf{Z} \sim \mathcal{SN}_d(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, if its probability density function is given by

$$f_{\mathbf{Z}}(\mathbf{z}) = 2\phi_d(\mathbf{z}|\boldsymbol{\eta}, \boldsymbol{\Sigma})\, \Phi_1\left(\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\eta})\right); \mathbf{z} \in \mathbb{R}^d, \tag{2.12}$$

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

Let us notice that, when $\boldsymbol{\lambda} = \mathbf{0}$, the expression in (2.12) 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

$$\mathbf{Z} \equiv \boldsymbol{\eta} + \boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{\delta}\,|X_0| + (\mathbf{I} - \boldsymbol{\delta}\boldsymbol{\delta}')^{1/2}\,\mathbf{X}_1\right), \tag{2.13}$$

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

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

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

and its variance-covariance matrix is given by

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

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

2.2.6 Skew-t distribution

According to the definition given by Azzalini and Capitanio (2003), we say that a random vector, $\mathbf{W}$, follows a d-dimensional Skew-t distribution with location parameter $\boldsymbol{\eta}$, scale matrix $\boldsymbol{\Sigma}$, skewness parameter $\boldsymbol{\lambda}$, and kurtosis parameter ν, denoted as $\mathbf{W} \sim \mathcal{ST}_d\left(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu\right)$, if its probability density function is given by

$$f_{\mathbf{W}}\left(\mathbf{w}\right) = 2t_d\left(\mathbf{w}|\boldsymbol{\eta}, \boldsymbol{\Sigma}, \nu\right) T_1\left(\left.\frac{\sqrt{\nu+d}\;\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1/2}\left(\mathbf{w}-\boldsymbol{\eta}\right)}{\sqrt{\left(\mathbf{w}-\boldsymbol{\eta}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{w}-\boldsymbol{\eta}\right)+\nu}}\right| 0, 1, \nu+d\right); \mathbf{w} \in \mathbb{R}^d,$$

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 $\nu > 0$.

This distribution allows for an alternative stochastic representation of the form

$$\mathbf{W} \equiv \boldsymbol{\eta} + U^{-1}\mathbf{Z},$$

where $\mathbf{Z} \sim \mathcal{SN}_d\left(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}\right)$ is independent from U with probability density function

$$f_U\left(u\right) = 2\frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)}u^{\nu-1}\exp\left\{-\frac{\nu}{2}u^2\right\}\mathbf{I}\left(u>0\right).$$

This is, $U^2 \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$.

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

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

for $\nu > 1$, its variance-covariance matrix is given by

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

for $\nu > 2$, and its kurtosis coefficient is given by

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

if $\nu > 4$, where

$$\begin{aligned} a_1 &= d(d+2) + 2(d+2)\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}^{-1}_{\mathbf{W}}\boldsymbol{\mu}_{\mathbf{W}} + 3\left(\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}^{-1}_{\mathbf{W}}\boldsymbol{\mu}_{\mathbf{W}}\right)^2, \\ a_2 &= \left(d + 2\sqrt{\frac{2}{\nu}}\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)}\right)\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}^{-1}_{\mathbf{W}}\boldsymbol{\mu}_{\mathbf{W}} \\ &+ \left(1 + 2\sqrt{\frac{2}{\nu}}\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu-1}{2}\right)} - \frac{\pi}{\nu-2}\frac{\Gamma^2\left(\frac{\nu}{2}\right)}{\Gamma^2\left(\frac{\nu-1}{2}\right)}\right)\left(\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}^{-1}_{\mathbf{W}}\boldsymbol{\mu}_{\mathbf{W}}\right)^2, \\ &\text{and } a_3 = 2(d+2)\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}^{-1}_{\mathbf{W}}\boldsymbol{\mu}_{\mathbf{W}} + 3\left(\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}^{-1}_{\mathbf{W}}\boldsymbol{\mu}_{\mathbf{W}}\right)^2, \\ &\text{with } \boldsymbol{\mu}_{\mathbf{W}} = E(\mathbf{W} - \boldsymbol{\eta}) \text{ and } \boldsymbol{\Sigma}_{\mathbf{W}} = V(\mathbf{W}). \end{aligned}$$

The Skew-t distribution is more flexible than the Gaussian because not only does it allow for asymmetry, but it is also able to capture heavier tails than the normal distribution. Also, the Skew-t distribution is a particular member of the Skew-Normal/independent distribution that will be detailed in section 2.2.8.

2.2.7 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

$$\mathbf{W} \equiv \boldsymbol{\eta} + U^{-1}\boldsymbol{\Sigma}^{1/2}\mathbf{Z}. \tag{2.14}$$

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 probability density function 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 (2.14), and taking into account the alternative stochastic representation for a random vector that follows a Skew-Normal distribution given in (2.13), we get a more elaborate alternative stochastic representation for the Skew-Slash distribution

$$\mathbf{W} \equiv \boldsymbol{\eta} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}X + U^{-1}\left[\boldsymbol{\Sigma}^{1/2}\left(\mathbf{I}-\boldsymbol{\delta}\boldsymbol{\delta}'\right)\boldsymbol{\Sigma}^{1/2}\right]^{1/2}\mathbf{X}_1, \tag{2.15}$$

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 (2.14), according to Wang and Genton (2006), its expectation is given by

$$E(\mathbf{W}) = \boldsymbol{\eta} + \sqrt{\frac{2}{\pi}} \frac{\nu}{\nu - 1} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}, \text{ for } \nu > 1, \tag{2.16}$$

and its variance-covariance matrix is given by

$$V(\mathbf{W}) = \boldsymbol{\Sigma}^{1/2} \left\{ \frac{\nu}{\nu - 2} \mathbf{I} - \frac{2}{\pi} \left(\frac{\nu}{\nu - 1} \right)^2 \boldsymbol{\delta} \boldsymbol{\delta}' \right\} \boldsymbol{\Sigma}^{1/2}, \text{ for } \nu > 2. \tag{2.17}$$

Finally, we can give a closed expression for the kurtosis coefficient using Proposition 7 from Lachos, Labra and Ghosh (2007), given in (2.18), defined as $\gamma_2(W) = E\left[\left\{ (\mathbf{W} - \boldsymbol{\mu}_W)' \boldsymbol{\Sigma}_W^{-1} (\mathbf{W} - \boldsymbol{\mu}_W) \right\}^2 \right]$, as established by Mardia (1974), that represents the extension of the kurtosis coefficient proposed by Azzalini and Capitanio (1999) for the Skew-Normal distribution.

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

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

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

To illustrate the ability of the Skew-Slash distribution to generate high kurtosis, Figure 2.6 shows some values of the kurtosis coefficient for a 2-dimensional Skew-Slash random vector and, analogously, Figure 2.7 does the

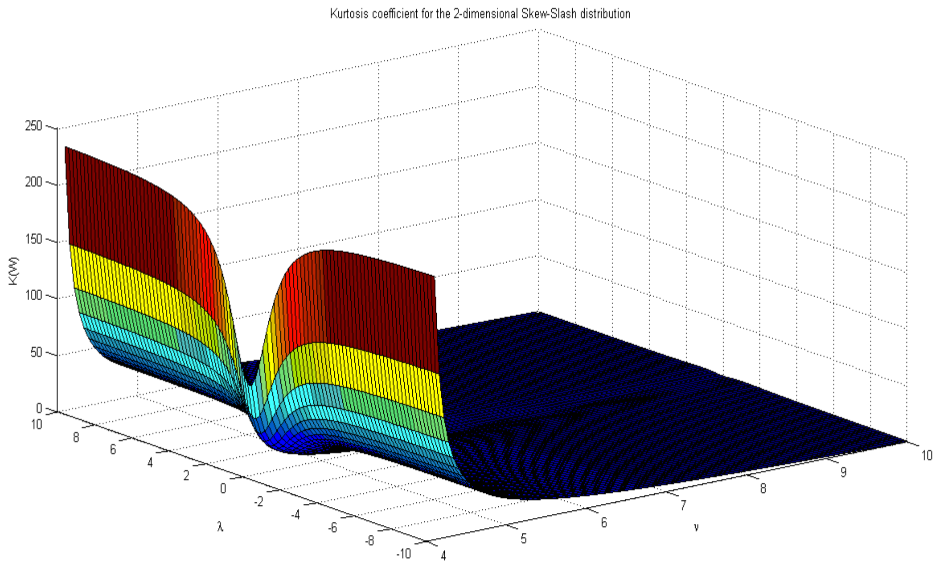


Figure 2.6: Kurtosis coefficient of the 2-dimensional Skew-Slash distribution as a function of λ_i and ν

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. As a consequence, even though we cannot plot the kurtosis coefficient in all cases, we can at least use this illustration to represent an idea of its structure.

It can be seen that the general structure presented for the univariate case is replicated here, and we also are able to notice that, as the dimension gets higher, the kurtosis allowed for the same values also gets higher.

Let us notice that, just like the Skew-t distribution, the Skew-Slash distribution is a particular member of the Skew-Normal/independent family of

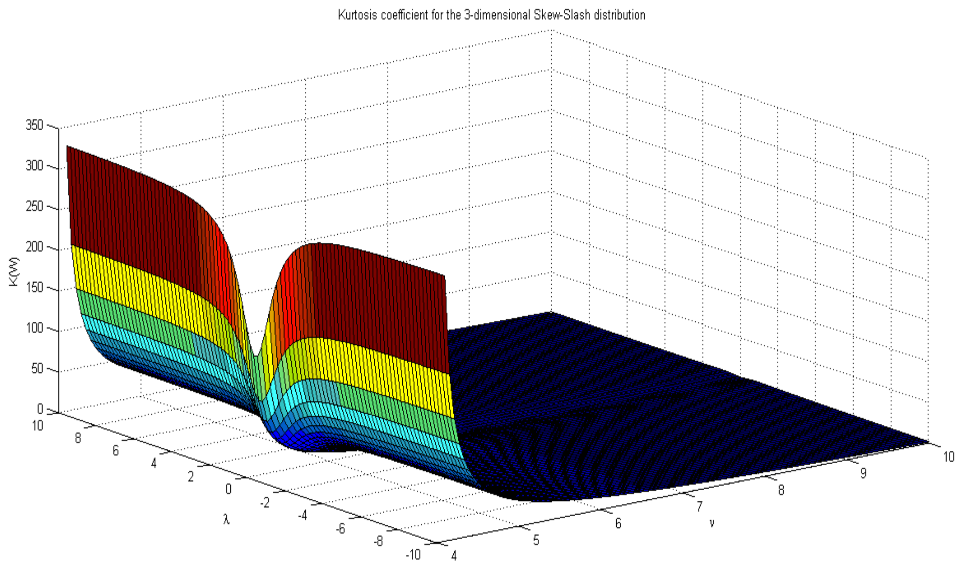


Figure 2.7: Kurtosis coefficient of the 3-dimensional Skew-Slash distribution as a function of λ_i and ν

distributions, that is exposed below.

2.2.8 Skew-Normal/independent family of distributions

According to the definition given by Lachos, Labra, and Ghosh (2007), we say that a d-dimensional random vector, $\mathbf{W}$, follows a Skew-Normal/independent distribution with location parameter $\boldsymbol{\eta}$, dispersion matrix $\boldsymbol{\Sigma}$, and skewness parameter $\boldsymbol{\lambda}$ if its probability density function takes the form

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

where U is a positive random variable with cumulative distribution function $H(u|\nu)$, and we denote it as $\mathbf{W} \sim \mathcal{SNI}_d(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. Also, $\boldsymbol{\eta} \in \mathbb{R}^d$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is a symmetric positive definite matrix, and $\boldsymbol{\lambda} \in \mathbb{R}^d$.

Its alternative stochastic representation is given by

$$W \equiv \boldsymbol{\eta} + U^{-1}\mathbf{Z},$$

where $\mathbf{Z} \sim \mathcal{SN}_d\left(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}\right)$, and U and $\mathbf{Z}$ are independent.

An even more detailed stochastic representation that incorporates (2.13) leads to the following expression.

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

where $X = U^{-1}\left|X_0\right|$; $X_0 \sim \mathcal{N}_1\left(0,1\right)$, $\mathbf{X}_1 \sim \mathcal{N}_d\left(\mathbf{0}, \mathbf{I}\right)$; X_0 and $\mathbf{X}_1$ are independent, and $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{\lambda}}$. In this case, we have that $\left\|\boldsymbol{\delta}\right\|_2 < 1$.

Proposition 7 in the paper by Lachos, Labra, and Ghosh (2007) gives straightforward expressions for the mean vector, variance-covariance matrix, and the multidimensional kurtosis coefficient, defined by Mardia (1974), that is an extension of Azzalini and Capitanio's(1999) kurtosis coefficient for the Skew-Normal distribution, for a Skew-Normal/independent random vector by explaining us that, if $\mathbf{W} \sim \mathcal{SNI}_d\left(\boldsymbol{\eta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H\right)$, then

- If $E\left(U^{-1}\right) < \infty$, then

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

- If $E\left(U^{-2}\right) < \infty$, then

$$V\left(\mathbf{W}\right) = \boldsymbol{\Sigma}_{\mathbf{W}} = E\left(U^{-2}\right)\boldsymbol{\Sigma} - \frac{2}{\pi}E^2\left(U^{-1}\right)\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{1/2}, \text{ and}$$

- If $E\left(U^{-4}\right) < \infty$, then the multidimensional kurtosis coefficient is

$$\gamma_2\left(\mathbf{W}\right) = \frac{E\left(U^{-4}\right)}{E^2\left(U^{-2}\right)}a_1 - 4\frac{E\left(U^{-3}\right)}{E^2\left(U^{-2}\right)}a_2 + a_3 - d\left(d+2\right), \tag{2.18}$$

where

$$\begin{aligned}
a_1 &= d\,(d+2) + 2\,(d+2)\,\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}_{\mathbf{W}}^{-1}\boldsymbol{\mu}_{\mathbf{W}} + 3\left(\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}_{\mathbf{W}}^{-1}\boldsymbol{\mu}_{\mathbf{W}}\right)^2, \\
a_2 &= \left(d + \frac{2}{E\left(U^{-1}\right)}\right)\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}_{\mathbf{W}}^{-1}\boldsymbol{\mu}_{\mathbf{W}} \\
&\quad + \left(1 + \frac{2}{E\left(U^{-1}\right)} - \frac{\pi}{2}\frac{E\left(U^{-2}\right)}{E^2\left(U^{-1}\right)}\right)\left(\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}_{\mathbf{W}}^{-1}\boldsymbol{\mu}_{\mathbf{W}}\right)^2, \\
\text{and } a_3 &= 2\,(d+2)\,\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}_{\mathbf{W}}^{-1}\boldsymbol{\mu}_{\mathbf{W}} + 3\left(\boldsymbol{\mu}'_{\mathbf{W}}\boldsymbol{\Sigma}_{\mathbf{W}}^{-1}\boldsymbol{\mu}_{\mathbf{W}}\right)^2, \\
\text{with } \boldsymbol{\mu}_{\mathbf{W}} &= E\left(\mathbf{W} - \boldsymbol{\eta}\right) = \sqrt{\frac{2}{\pi}}E\left(U^{-1}\right)\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}.
\end{aligned}$$

Some of the distributions comprised in the Skew-Normal/independent family as special cases are the Skew-t, the multivariate Skew-Slash, and the Contaminated Skew-Normal, among others.

Chapter 3

GARCH process with Skew-Slash innovations

In this chapter, we present our proposal to model univariate time series of financial returns.

First of all, we must acknowledge the fact that the use of a conditional heteroskedastic process to model financial time series of returns, together with a flexible model for the innovations, is usually appropriate to capture the structure of this kind of data. To do so, we propose the use of a Generalized Autoregressive Conditional Heteroskedastic (GARCH) process, as defined by Bollerslev (1986), with Skew-Slash innovations, but we could also use the Skew-Slash distribution to model the innovations in other conditional heteroskedastic models, such as the GJR-GARCH by Glosten, Jagannathan, and Runkle (1993), the EGARCH by Nelson (1991) or the TGARCH by Zakoian (1994).

We start by explaining the model that we want to work with. Then, we develop methodologies that allow us to estimate our model, supposing that we have a data set available, and we do it from a Maximum Likelihood point of view, as well as a Bayesian point of view.

Once we have defined all the theoretical part, we can move forward to the implementation of our procedure, and we do it by generating some simulated data sets, comparing the performances of our two methodologies, and, finally, we use our method to fit the corresponding models to help us explain the behavior of a real data set, and we exemplify by using the log-returns of the Standard & Poor's index from January 3$^{\text{rd}}$, 2000 to December 28$^{\text{th}}$, 2013.

Although the proposed approach of modeling the innovations using the Skew-Slash distribution can be adopted in any conditional heteroskedastic model, for simplicity and because it is the most widely used model by practitioners for estimating the dynamics of financial returns, for illustration purposes, we use the GARCH(1,1), defined as:

$$y_t = \mu + \sqrt{h_t}\varepsilon_t; \tag{3.1}$$

$$h_t = \omega + \alpha \left(y_{t-1} - \mu\right)^2 + \beta h_{t-1}, \tag{3.2}$$

for $t = 1, \dots, T$, where h_t is the conditional variance of y_t given $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$, the information set available until time $t-1$, and the innovations, ε_t, are independent and identically distributed random variables such that $E\left(\varepsilon_t\right) = 0$ and $V\left(\varepsilon_t\right) = 1$, for $t = 1, \dots, T$. We also assume that $\omega \geq 0$, $\alpha \geq 0$, and $\beta \geq 0$, to ensure non-negativity of h_t and $\alpha + \beta < 1$, to ensure covariance stationarity.

Here, we model the innovations as Skew-Slash distributed; this is,

$$\varepsilon_t \sim \mathcal{SSL}_1\left(\eta, \sigma, \lambda, \nu\right),$$

for all $t \in \{1, \dots, T\}$, and we assume that the innovations are all independent and identically distributed random variables. Note that it is well known that the absolute value of the unconditional skewness of y_t is larger than the absolute value of the skewness of the innovations, while the unconditional kurtosis of y_t is larger than the kurtosis of the innovations. Therefore, it is expected that the GARCH model with Skew-Slash innovations is able to capture the skewness and kurtosis of financial returns.

Note also that the restrictions on the mean and variance of the innovations imply restrictions on the parameters of the Skew-Slash distribution. In particular, the zero mean restriction implies, from (2.5), that

$$\eta = -\sigma\sqrt{\frac{2}{\pi}}\delta\frac{\nu}{\nu-1}, \tag{3.3}$$

for $\nu > 1$, and the unit variance condition implies, from (2.9), that

$$\sigma^2 = \left\{\frac{\nu}{\nu-2} - \frac{2}{\pi}\delta^2\left(\frac{\nu}{\nu-1}\right)^2\right\}^{-1}, \tag{3.4}$$

for $\nu > 2$. Therefore, η and σ will be functions of λ (or, equivalently, δ) and ν.

It is important to note that usual stationarity conditions of the GARCH process are directly applicable using the results in Carrasco and Chen (2002). In particular, from their Proposition 12, if $\nu > 2$, then, $E\left(h_t\right) < \infty$ and $E\left(y_t^2\right) < \infty$. Moreover, if h_0 is a constant, then $\{(y_t, \varepsilon_t)\}_{t=1}^{T}$ is strictly stationary and β-mixing with exponential decay.

3.1 Inference

In this section, we outline schemes for both Maximum Likelihood and Bayesian inference for the univariate Skew-Slash GARCH model that incorporates the parameter restrictions in (3.3) and (3.4).

We assume throughout that a series of returns $\mathbf{y} = (y_1, \ldots, y_T)$ is observed and that the initial value, y_0, of the series is known, which is not a restrictive assumption from a practical point of view, because financial data sets usually present elevated sample sizes. Also, from now on we will consider $\boldsymbol{\vartheta}$ as the set of parameters.

3.1.1 Maximum Likelihood inference

In this case, we assume also that the initial volatility, h_0, is known. Under the GARCH model with Skew-Slash innovations, the likelihood function of a GARCH(1,1) model is given by

$$f(\mathbf{y}|\boldsymbol{\vartheta}) = f(y_T|\mathcal{F}_{T-1})\, f(y_{T-1}|\mathcal{F}_{T-2}) \cdots f(y_1|\mathcal{F}_0)$$
$$= \prod_{t=1}^{T} \left\{ h_t^{-1/2} \int_0^1 2\nu_0 u^{\nu-1} \phi\left(\frac{y_t - \mu}{\sqrt{h_t}} \middle| \eta, u^{-2}\sigma^2 \right) \Phi\left(\frac{\lambda u \left(y_t - \mu - \eta\sqrt{h_t} \right)}{\sigma^2 \sqrt{h_t}} \right) du \right\},$$

where, $\boldsymbol{\vartheta} = (\mu, \alpha, \beta, \lambda, \nu)'$ is a vector that contains the collection of all the parameters of the model, $f(y_1, \ldots, y_m|\boldsymbol{\vartheta})$ is the joint probability density function of $y_1, \ldots, y_m$, and η and σ^2 are given in (3.3) and (3.4), respectively.

The Maximum Likelihood (ML) estimator can then be obtained by maximizing the log-conditional likelihood function

$$\mathcal{L}(\boldsymbol{\vartheta}|y_1, \ldots, y_T) = \sum_{t=1}^{T} \ell_t(\boldsymbol{\vartheta})$$

where

$$\ell_t(\boldsymbol{\vartheta}) = -\frac{1}{2}\ln(h_t)$$
$$+ \ln\left\{ \int_0^1 2\nu u^{\nu-1} \phi_1\left(\frac{y_t - \mu}{\sqrt{h_t}} \middle| \eta, u^{-2}\sigma^2 \right) \Phi_1\left(\frac{\lambda u \left(y_t - \mu - \eta\sqrt{h_t} \right)}{\sigma^2 \sqrt{h_t}} \right) du \right\}$$

The maximization of the log-conditional likelihood is a highly nonlinear problem, but it can be carried out by standard numerical algorithms. In particular, the implementations that come later in the present thesis are carried out using the `constrOptim` function in the free software `R` (http://www.r-project.org/). Given the stationarity and mixing properties of the processes y_t and h_t previously mentioned, it is reasonable to apply usual large sample results of Maximum Likelihood estimation. Therefore, in the following, it can be assumed that the Maximum Likelihood estimator of $\boldsymbol{\vartheta}$, denoted by $\widehat{\boldsymbol{\vartheta}}$, is asymptotically Gaussian distributed with mean $\boldsymbol{\vartheta}$ and covariance matrix $-E\left(\partial^2\mathcal{L}\left(\boldsymbol{\vartheta}|y_1,\ldots,y_T\right)/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'\right)^{-1}$. Then, approximated standard innovations of the parameters can be obtained by taking the square roots of the diagonal elements of $\partial^2\mathcal{L}\left(\widehat{\boldsymbol{\vartheta}}|y_1,\ldots,y_T\right)/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'$.

3.1.2 Bayesian inference

Here, we shall assume that y_0 is known as earlier, but let the initial volatility, h_0, be unknown. Then, in order to undertake Bayesian inference, it is first necessary to define prior distributions for the model parameters, i.e., $\boldsymbol{\vartheta}$ and h_0. One of the advantages of the Bayesian approach in this context is that, for many of these parameters, real prior information in the form of expert knowledge or based on economic theory will be available and this information can be incorporated into the analysis. Thus, for example, economic theory suggests that the drift parameter, μ, in the GARCH model should be (very close to) zero. Therefore a reasonable prior distribution that incorporates this knowledge is a normal distribution with small variance and centered at 0. Secondly, analysts with experience in GARCH models will often be able to provide good prior estimates of the volatility parameters ω, α, and β. Thirdly, the parameter ν determines the number of finite moments of the error distribution and analysts will often be able to give good estimates of this parameter. We have also seen in Figure 2.3 that the skewness of the Skew-Slash distribution is largely determined by λ or, equivalently, δ. As financial returns usually only exhibit light skewness, this suggests that a distribution for λ (or δ) centered at 0 with moderate variance could be considered. A relatively disperse prior distribution for the initial volatility can be assumed. Let us denote the collection of hyperparameters as $\mathcal{H}$. Here we shall assume the following prior distributions:

1. $\ln(h_0)|\mathcal{H} \sim \mathcal{G}(a_h, b_h)$, a gamma distribution with mean a_h/b_h.
2. $\mu|\mathcal{H} \sim \mathcal{N}_1\left(0, \frac{1}{c_m}\right)$, where $c_m \gg 1$.
3. $\omega|\mathcal{H} \sim \mathcal{G}(a_\omega, b_\omega)$, where $b_\omega \gg a_\omega$.
4. $f(\alpha, \beta|\mathcal{H}) = \frac{\Gamma(c)}{\Gamma(cp_a)\Gamma(cp_b)\Gamma(c(1-p_a-p_b))}\alpha^{cp_a-1}\beta^{cp_b-1}(1-\alpha-\beta)^{c(1-p_a-p_b)}$, where $c, p_a, p_b > 0$ and $p_a + p_b < 1$.
5. $\frac{\delta+1}{2}|\mathcal{H} \sim \mathcal{B}e(e, e)$, a scaled, shifted beta distribution centered at $\delta = 0$.
6. $\nu|\mathcal{H} \sim \mathcal{G}(a_n, b_n)$, a gamma distribution with mean a_n/b_n.

Note, in particular, that the prior for (α, β) is a Dirichlet density with mean $E\{(\alpha, \beta)\} = (p_a, p_b)$. All the prior parameters will be specified later on.

Given this prior structure, exact inference is impossible. However, in a similar way to Lachos, Dey, and Cancho (2009), we can use the stochastic representation of the Skew-Slash distribution to allow us to develop a Markov chain Monte Carlo algorithm. Consider the t-th innovation, ε_t, that follows a Skew-Slash distribution as outlined earlier. Then, from (2.4), introducing the latent variables $u_t|\nu \sim \mathcal{B}e\left(\nu, 1\right)$ and $X_{0t}, X_{1t} \sim \mathcal{N}_1\left(0, 1\right)$, we have that

$$\varepsilon_t \equiv \eta + \sigma u_t^{-1}\left(\delta\left|X_{0t}\right| + \sqrt{1-\delta^2}X_{1t}\right)$$

and, defining $r_t = X_{0t}/u_t$ so that $r_t|u_t \sim \mathcal{N}_1\left(0, u_t^{-1}\right)\mathbf{I}\left(\mathbb{R}^+\right)$, we have that

$$\varepsilon_t \equiv \eta + \sigma\left(\delta\left|r_t\right| + \sqrt{1-\delta^2}\frac{X_{1t}}{u_t}\right).$$

Therefore, conditional on the parameters, δ and ν, and the latent variables, u_t and r_t, we have

$$\varepsilon_t|\delta, \nu, u_t, r_t \sim \mathcal{N}_1\left(\eta + \sigma\delta|r_t|, \frac{\sigma^2}{u_t^2}\right),$$

where we note, to incorporate the restrictions on the innovations, that $\eta = \eta(\delta, \nu)$ and $\sigma = \sigma(\delta, \nu)$ are specified in (3.3) and (3.4), respectively. Finally, (3.1) implies that

$$y_t|\boldsymbol{\vartheta}, h_0, u_t, r_t \sim \mathcal{N}_1\left(\mu + \sqrt{h_t}\left(\eta + \sigma\delta|r_t|\right), \frac{h_t\sigma^2}{u_t^2}\right),$$

where $\boldsymbol{\vartheta} = (\mu, \alpha, \beta, \delta, \nu)'$ and $h_t = h_t(\mu, \omega, \alpha, \beta, h_{t-1})$, are as in (3.2).

It is now straightforward to derive the following conditional posterior densities for the model parameters and latent variables. Firstly, consider the latent variables and parameters associated with the innovation distribution. Then:

$$u_t^2|\varepsilon_t, \delta, \nu, r_t \sim \mathcal{G}\left(\frac{\nu+2}{2}, \frac{\left(\varepsilon_t - \eta - \sigma\delta r_t\right)^2 + \sigma^2 r_t^2\left(1-\delta^2\right)}{2\sigma^2\left(1-\delta^2\right)}\right)\mathbf{I}\left(0, 1\right). \tag{3.5}$$

$$r_t|\varepsilon_t, \delta, \nu, u_t \sim \mathcal{N}_1\left(\frac{\delta\left(\varepsilon_t - \eta\right)}{\sigma}, \frac{1-\delta^2}{u_t^2}\right)\mathbf{I}\left(\mathbb{R}^+\right) \tag{3.6}$$

$$f\left(\delta|\boldsymbol{\varepsilon}, \nu, \mathbf{u}, \mathbf{r}, \mathcal{H}\right) \propto f\left(\delta|\mathcal{H}\right)\frac{1}{\sigma^n}\prod_{t=1}^{T}\phi_1\left(\frac{u_t\left(\boldsymbol{\varepsilon}_t - \eta - \sigma\delta|r_t|\right)}{\sigma}\right) \tag{3.7}$$

$$f\left(\nu|\boldsymbol{\varepsilon}, \delta, \mathbf{u}, \mathbf{r}, \mathcal{H}\right) \propto f\left(\nu|\mathcal{H}\right)\frac{1}{\sigma^n}\prod_{t=1}^{T}\phi_1\left(\frac{u_t\left(\boldsymbol{\varepsilon}_t - \eta - \sigma\delta|r_t|\right)}{\sigma}\right), \tag{3.8}$$

where, in the distributions for δ and ν, we recall that η and σ are functions of δ and ν.

Finally, in the case of the GARCH parameters, we have

$$f(\mu|\mathbf{y},\delta,\nu,h_0,\omega,\alpha,\beta,\mathbf{u},\mathbf{r},\mathcal{H}) \propto f(\mu|\mathcal{H}) \prod_{t=1}^{T} \frac{1}{\sqrt{h_t}}\phi_1\left(\frac{u_t\left(y_t-\mu-\sqrt{h_t}(\eta+\sigma\delta|r_t|)\right)}{\sigma h_t}\right),$$

where we note that h_t is here a function of μ. Similar expressions are available for the remaining GARCH parameters with posterior distributions of the form.

$$f(h_0|\mathbf{y},\delta,\nu,\mu,\omega,\alpha,\beta,\mathbf{u},\mathbf{r},\mathcal{H}) \propto f(h_0|\mathcal{H})\, f(\mathbf{y},\mathbf{r},\mathbf{u}|\vartheta),$$

$$f(\alpha,\beta|\mathbf{y},\delta,\nu,\mu,h_0,\omega,\mathbf{u},\mathbf{r},\mathcal{H}) \propto f(\alpha,\beta|\mathcal{H})\, f(\mathbf{y},\mathbf{r},\mathbf{u}|\vartheta),$$

$$f(\omega|\mathbf{y},\delta,\nu,\mu,h_0,\alpha,\beta,\mathbf{u},\mathbf{r},\mathcal{H}) \propto f(\omega|\mathcal{H})\, f(\mathbf{y},\mathbf{r},\mathbf{u}|\vartheta).$$

Then, we can set up the following Metropolis within Gibbs sampling algorithm to sample from the posterior parameter distributions:

1. Set initial values $\delta,\nu,h_0,\omega,\alpha,\beta,u_t$ for $t=1,\ldots,n$.

2. Repeat:

   (a) For $t=1,\ldots,n$, calculate h_t from (3.2).

   (b) For $t=1,\ldots,n$, calculate $\varepsilon_t=\frac{y_t-\mu}{\sqrt{h_t}}$.

   (c) Calculate $\sigma=\sigma(\delta,\nu)$ from (3.4).

   (d) Calculate $\eta=\eta(\sigma,\delta,\nu)$ from (3.3).

   (e) For $t=1,\ldots,n$ generate r_t from $f(r_t|\varepsilon_t,\delta,\nu,u_t)$.

   (f) For $t=1,\ldots,n$ generate u_t from $f(u_t|\varepsilon_t,\delta,\nu,r_t)$.

   (g) Generate δ from $f(\delta|\varepsilon,\nu,\mathbf{u},\mathbf{r})$.

(h) Generate ν from $f(\nu|\varepsilon, \nu, \mathbf{u}, \mathbf{r})$.

(i) Generate h_0 from $f(h_0|\mathbf{y}, \mu, \omega, \alpha, \beta, \delta, \nu, \mathbf{u}, \mathbf{r})$. For $t = 1, \ldots, n$, calculate h_t from (3.2).

(j) Generate μ from $f(\mu|\mathbf{y}, h_0, \omega, \alpha, \beta, \delta, \nu, \mathbf{u}, \mathbf{r})$. For $t = 1, \ldots, n$, calculate h_t from (3.2).

(k) Generate ω from $f(\omega|\mathbf{y}, h_0, \mu, \alpha, \beta, \delta, \nu, \mathbf{u}, \mathbf{r})$. For $t = 1, \ldots, n$, calculate h_t from (3.2).

(l) Generate α, β from $f(\alpha, \beta|\mathbf{y}, h_0, \mu, \omega, \delta, \nu, \mathbf{u}, \mathbf{r})$.

Steps 2e and 2f can be carried out using Gibbs passes whereas for steps 2g to 2k, random walk algorithms are applied. In particular, in step 2g we generate candidates $\tilde{\delta}$ such that $\ln\left(\frac{1+\tilde{\delta}}{1-\tilde{\delta}}\right)$ is generated from a normal distribution centered at $\ln\left(\frac{1+\delta}{1-\delta}\right)$. In step 2h, we generate $\ln(\tilde{\nu} - 2)$ from a normal distribution centered at $\ln(\nu - 2)$. In 2i and 2j candidates are generated from normal distributions centered at their current values and in 2k, a candidate is generated from a lognormal distribution with location parameter equal to $\ln(\omega)$. Finally, in step 2l, although it might appear natural to generate candidates from a Dirichlet distribution centered at the current values, it is well known that a sampler of this type can have problems in sticking at values very close to the boundaries. Therefore, we instead use an alternative Metropolis Hastings proposal. Writing $\gamma = 1 - \alpha - \beta$, we first generate a candidate $\tilde{\gamma}$ such that logit $\tilde{\gamma}$ comes from a normal distribution centered at logit γ. Then, we generate a candidate value $\tilde{\beta}$ so that $\ln\left(\frac{\tilde{\beta}}{1-\tilde{\gamma}}\right)$ comes from a normal distribution centered at $\ln\left(\frac{\beta}{1-\gamma}\right)$. Finally we set $\tilde{\alpha} = 1 - \tilde{\beta} - \tilde{\gamma}$. In steps 2g to 2l, the generated candidates are accepted according to the appropriate Metropolis Hastings acceptance probability and, if candidates are rejected, the current values are returned. Note finally that the variances of the candidate generators can be set to achieve reasonable acceptance rates of around 20% in each case.

3.2 Examples

In this section, we illustrate our approach with both simulated data and a real data study. In all cases, when the Bayesian approach is used, the prior parameters are specified as $a_h = b_h = 0.1$, $c_m = 4$, $a_w = 0.001$, $b_w = 0.1$, $c = 1$, $(p_a, p_b) = (0.05, 0.9)$, $e = 5$, and $a_n = b_n = 0.5$. Initial values are set to be equal to the Maximum Likelihood Estimates and the sampler is run for 5000 iterations to burn in and 10000 iterations in equilibrium.

3.2.1 Simulated examples

Here, we consider the univariate GARCH with Skew-Slash innovations model with $(y_0 = 0,\ h_0 = 0.2)$, $\mu = 0$, $w = 0.01$, $\alpha = 0.1$ and $\beta = 0.85$ with the Skew-Slash parameters set as $\lambda = -1$ and $\nu = 5$, which implies that, to maintain the mean and variance restrictions of (3.3) and (3.4), $\eta = 0.652$ and $\sigma^2 = 0.855$, respectively.

Regarding the Maximum Likelihood Estimation approach, we focus here in parameter estimation of the Skew-Slash GARCH model. In order to assess the proposed Maximum Likelihood algorithm, we generated three sets of 2500 time series from a Skew-Slash GARCH model with the previous parameters. Each set corresponds to the sample sizes $T = 1000$, $T = 2000$, and $T = 3000$, respectively, that are usual sample sizes of time series of returns. The series in each set is estimated by Maximum Likelihood. Table 3.1 shows the mean and standard error of the model parameters over the 2500 time series estimated through maximum likelihood. Observe that the parameter estimation results are apparently very good, even for the smallest sample size. Also, as expected, the larger the sample size, the smaller the standard innovations and the better the mean estimates. Therefore, we conclude that the Maximum Likelihood procedure is apparently doing a good job in estimating the Skew-Slash GARCH model.

Table 3.1: Mean and Standard deviation for the ML estimators, with T simulated observations.

True parameters	$T = 1000$	$T = 2000$	$T = 3000$
$\mu = 0$	-0.00136 (0.01237)	-0.00058 (0.00861)	-0.00085 (0.00699)
$\omega = 0.01$	0.01161 (0.00505)	0.01068 (0.00286)	0.01025 (0.00237)
$\alpha = 0.1$	0.10115 (0.02610)	0.10114 (0.01906)	0.09981 (0.01542)
$\beta = 0.85$	0.84053 (0.04301)	0.84519 (0.02736)	0.84854 (0.02190)
$\lambda = -1$	-1.02685 (0.27714)	-1.01937 (0.19226)	-1.02641 (0.15933)
$\nu = 5$	5.43052 (1.72854)	5.22406 (0.93940)	5.16790 (0.56823)

To illustrate the proposed Bayesian inference scheme, we focus on a single generated series of length $T = 2000$. We run the MCMC algorithm for 5000 burn-in iterations plus 10000 iterations in equilibrium. All parameters were well estimated, with the true parameter values always inside the 90% credible intervals.

As examples, Figure 3.1 shows a kernel density estimate of the posterior density of h_0 and Figure 3.2 shows a kernel density estimate of the posterior density of ν. Although the density for h_0 is quite long tailed, the posterior

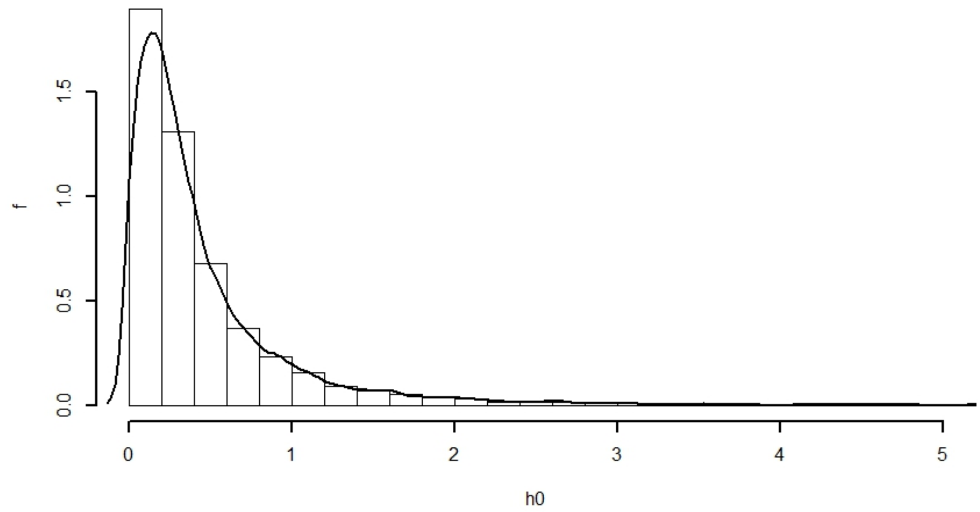


Figure 3.1: Estimated posterior density of h_0.

median of 0.27 is close to the true value of 0.2. The posterior density of ν concentrates on values around the true value of 5.

We also obtained fitted volatilities as the mean value of the fitted volatilities for the 10000 parameter values of the MCMC sampler. Figure 3.3 shows the true (solid line) and fitted (dotted line) volatilities that are very close to each other. Finally, Figure 3.4 shows the true (solid line) and predictive (dotted line) error density function. Again, both curves are very close indicating that the Bayesian estimation method does a remarkable job in estimating the Skew-Slash GARCH model.

3.2.2 Real data example

As an illustration of the usefulness of our approach, here we analyze real financial time series data from the Standard & Poor's 500 Index (S&P 500).

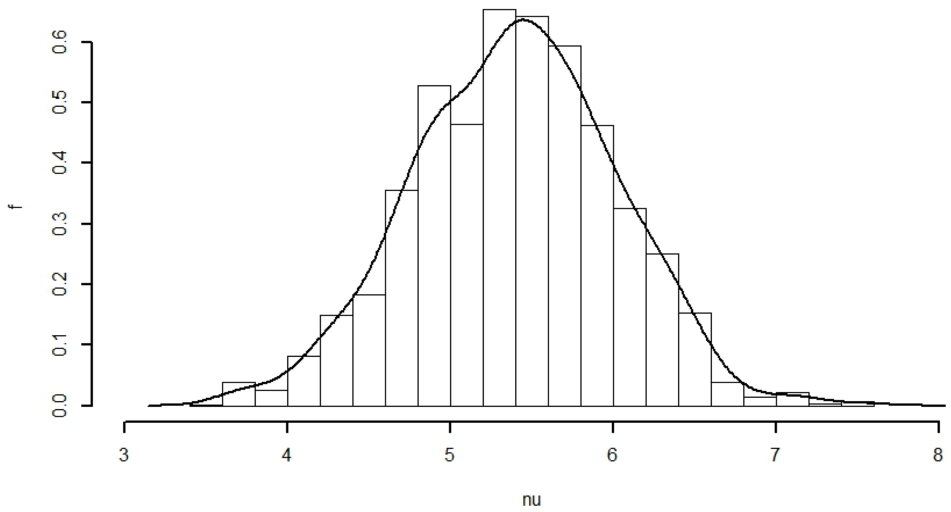


Figure 3.2: Estimated posterior density of ν.

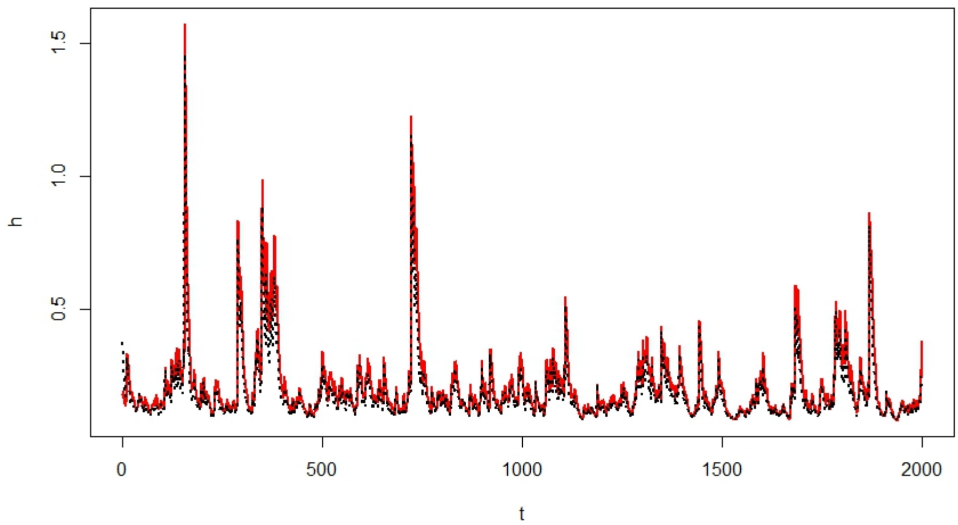


Figure 3.3: True (solid line) and fitted (dotted line) volatilities for the Bayesian fitting of the series with $T = 2000$.

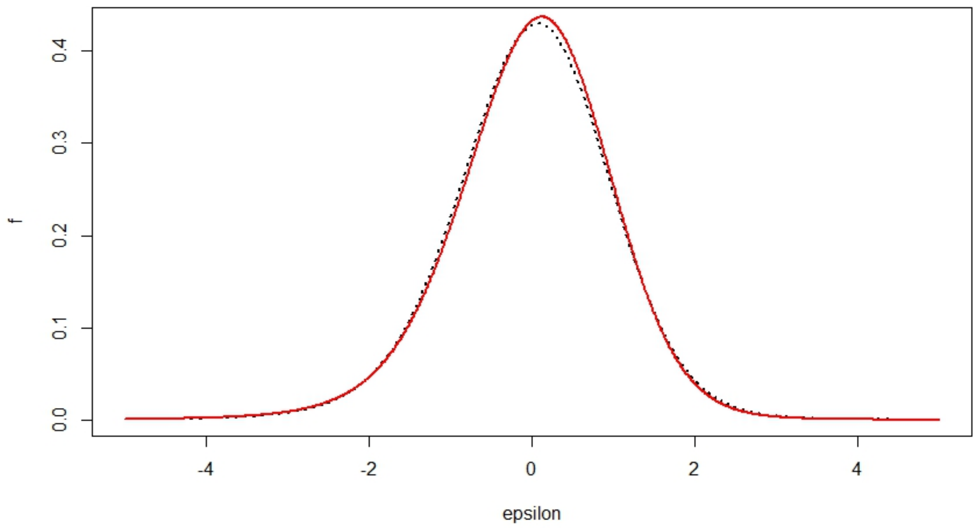


Figure 3.4: True (solid line) and fitted (dotted line) predictive error densities.

The S&P 500 index is a free-float capitalization-weighted index of the prices of 500 of the main companies in leading industries of the U.S. economy. Figure 3.5 shows the time plot of the log-returns of the daily closing prices of the index for the period from January 3$^{\text{rd}}$, 2000 until December 28$^{\text{th}}$, 2013, leading to 3035 index returns. Observe that the returns appear to vary more around mid 2008 till the end of 2009, which corresponds to the start of the financial crisis. The sample mean, standard deviation, skewness, and kurtosis of the return series are -1.85×10^{-5}, 0.0129, 0.1933, and 10.0481, respectively. Observe that the return series is slightly skewed and the kurtosis is large, indicating that the return distribution has higher peaks and heavier tails than a normal distribution with the same variance. Thus, it seems that the proposed Skew-Slash specification can be adequate to address these issues.

As with the simulated examples, we first fit the model using Maximum

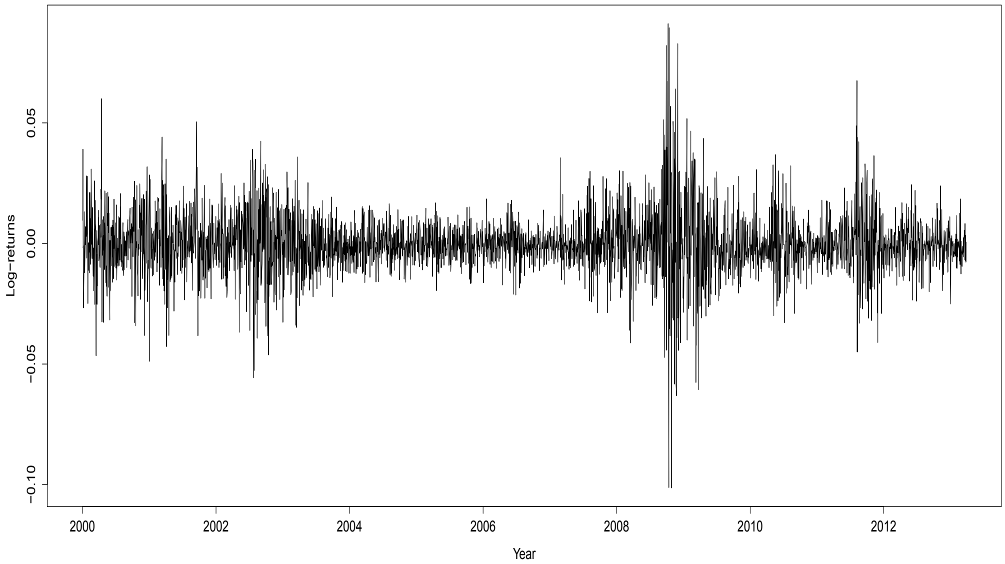


Figure 3.5: Standard & Poor's 500 log-returns between January 3rd, 2000 and December 28th, 2013.

Likelihood Estimation. Table 3.2 shows the Maximum Likelihood estimates of the parameters of the GARCH(1,1) model, assuming five different error distribution formulations, that is: Skew-Slash, Gaussian, Student's-t, Skew-Normal, and Skew-t distributions. Also included are the Maximum Likelihood estimates of the parameters of the Skew-Slash distribution. It can be seen that all estimates of the common parameters in the GARCH(1,1) model are very close, independent of the error distribution proposed.

Also, in Table 3.3, we present the information criterion statistics for the Gaussian, Student's-t, Skew-Normal, Skew-t, and Skew-Slash distributions, for two criteria. On one hand, we can find the Akaike Information Criterion (AIC), defined by Akaike (1976), given by

$$AIC\left(M\right) = 2\ln\left\{\mathcal{L}\left(\vartheta^{*}; M\right)\right\} - 2\,\text{card}\left(M\right),$$

where $\text{card}\,(M)$ denotes the number of parameters in the model, M, and $\mathcal{L}\,(\vartheta^*; M)$ is the maximized value of the likelihood function, $\mathcal{L}$, for the model in question. On the other hand, we can find the Bayesian Information Criterion (BIC), proposed by Schwarz (1978), given by

$$BIC\,(M) = -2\ln\{\mathcal{L}\,(\vartheta^*; M)\} + \text{card}\,(M)\ln(T),$$

where T is the number of observations. In this table (3.3), we can find that the Skew-Slash is the distribution that performs best, among the others, which means that the performance of our model is quite acceptable. The curious thing is that our best competitor is the Student's-t distribution, and not the Skew-t, but it might obey to the penalization of highly parameterized models, along with the fact that the skewness is only slight.

Figure 3.6 shows the histograms of the residual series after fitting with estimated densities for the five previous distributions. Note that the normal and the Student's-t distributions provide with symmetric estimated densities while the Skew-Normal, Skew-t, and Skew-Slash provide with asymmetric fits. In particular, note that the estimated density of the residuals corresponding to the Skew-Slash distribution fits the estimated residual histogram very well as compared to the other models.

Finally, we fitted the model using the Bayesian procedure. Just like we did with the simulations, we ran the MCMC algorithm for 5000 burn-in iterations plus 10000 iterations in equilibrium. Again, we focus on the estimated volatilities and the estimated error density. In particular, we obtain fitted volatilities as the mean value of the fitted volatilities for the 10000 parameter values of the MCMC sampler. Then, 3.7 shows the estimated volatilities using both the Maximum Likelihood Estimation (continuous line) and Bayesian (dotted line) approaches. The volatilities are very similar as expected. Finally, Figure 3.8

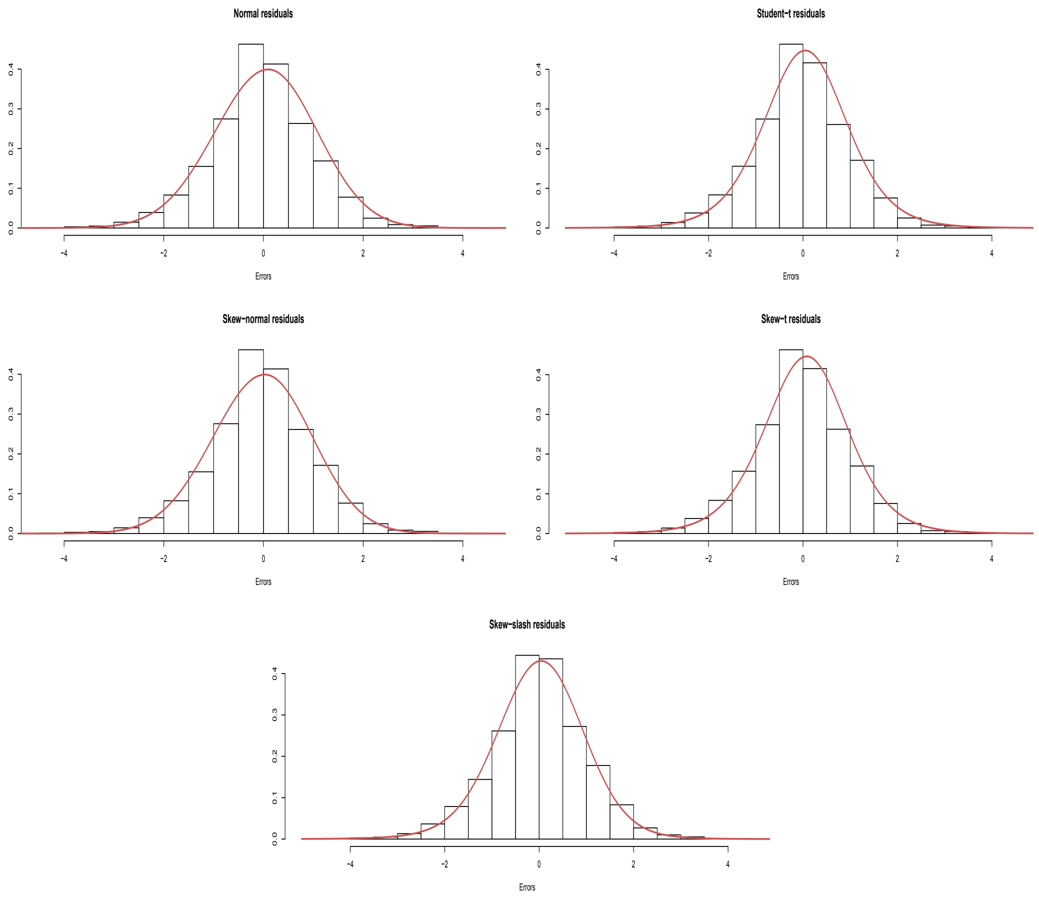


Figure 3.6: Histograms of the residuals after fitting joint with estimated densities for the Standard & Poor's 500 log-returns.

Table 3.2: ML estimates with their standard errors of the GARCH(1,1) model with several distributions.

ϑ	μ	ω	α	β	λ	ν
$\mathcal{SSL}_1$	-1.85×10^{-4} (1.54×10^{-4})	1.23×10^{-6} (3.24×10^{-7})	0.0936 (0.0103)	0.8997 (0.0103)	−0.4853 (0.1883)	5.0113 (0.5225)
$\mathcal{N}_1$	-1.85×10^{-4} (1.53×10^{-4})	1.53×10^{-6} (3.44×10^{-7})	0.0933 (0.0092)	0.8980 (0.0096)	–	–
$\mathcal{T}_1$	-1.85×10^{-4} (1.48×10^{-4})	1.30×10^{-6} (3.69×10^{-7})	0.0953 (0.0105)	0.8997 (0.0103)	–	–
$\mathcal{SN}_1$	-1.85×10^{-4} (1.53×10^{-4})	1.47×10^{-6} (3.41×10^{-7})	0.0947 (0.0093)	0.8972 (0.0096)	–	–
$\mathcal{ST}_1$	-1.85×10^{-4} (1.54×10^{-4})	1.28×10^{-6} (3.68×10^{-7})	0.0959 (0.0106)	0.8993 (0.0103)	–	–

Table 3.3: AIC and BIC for the GARCH(1,1) model with several distributions for the S&P data.

	AIC	BIC
Normal	−6.283566	−6.276222
Student's-t	**−6.303842**	**−6.294662**
Skew-Normal	−6.284018	−6.274838
Skew-t	−6.303420	−6.292404
Skew-Slash	**−6.305842**	**−6.294826**

shows the estimated error densities under both the Maximum Likelihood Estimation (continuous line) and Bayesian (dotted line) approaches. The densities are very similar and they both exhibit slight skewness as expected.

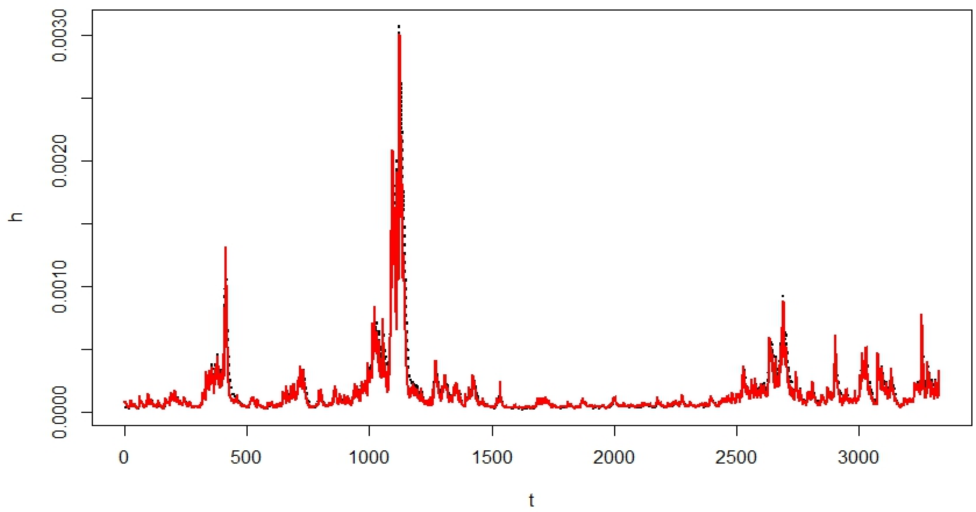


Figure 3.7: MLE (solid line) and Bayesian (dotted line) volatility estimates for the Standard & Poor's 500 log-returns.

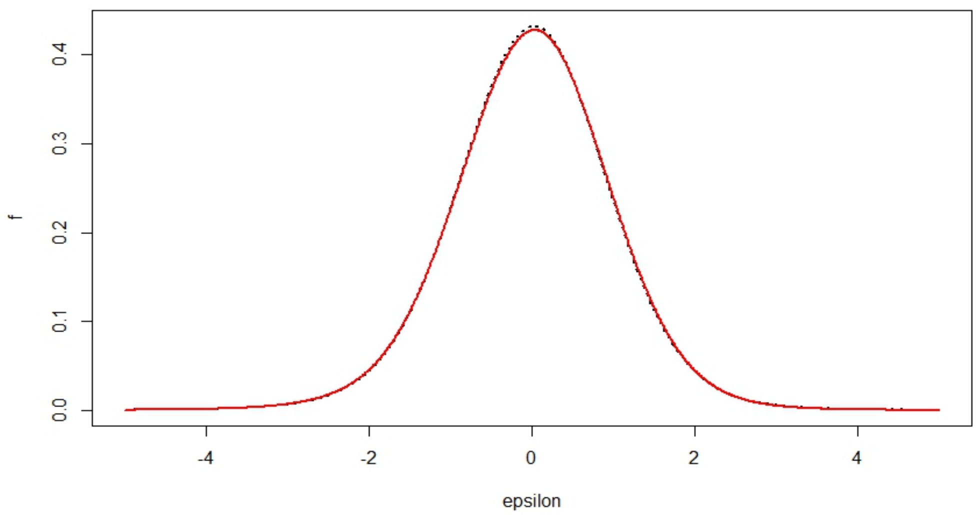


Figure 3.8: MLE (solid line) and Bayesian (dotted line) estimated error densities for the Standard & Poor's 500 log-returns.

Chapter 4

Dynamic Conditional Correlation model with Skew-Slash innovations

In this chapter, we present our proposal to model multivariate time series of financial returns.

First of all, we must acknowledge the fact that the use of a conditional heteroskedastic process to model the series of returns, together with a flexible model for the residuals, is usually appropriate to capture the behavior of financial returns. With this in mind, we propose the use of a Dynamic Conditional Correlation model with multivariate Skew-Slash innovations.

We begin explaining our model and, afterwards, we get to the development of a methodology that permits the estimation of our model under the assumption that we have an available data set, and we take a Bayesian approach.

With the theoretical part elaborated, we move forward towards the assessment of our methodology. For the implementation, we start by estimating the parameters of 2-dimensional and 3-dimensional simulated data sets. Afterwards, we use our method to fit a model that helps us explain the behavior of the returns of the Dow Jones and NASDAQ data sets from January 2$^{\text{nd}}$, 1996 to December 29$^{\text{th}}$, 2006 on one hand and, on the other hand, we try to explain the behavior of the log-returns of the DAX, CAC40, and Nikkei stock market indices between October 10$^{\text{th}}$, 1991 and December 30$^{\text{th}}$, 1997.

As we mentioned before, one of the most well known models used to describe the behavior of financial returns is the GARCH process. In this case, we want to develop a model in a multidimensional framework, which makes an important issue of the use of an appropriate generalization of the univariate GARCH process proposed by Bollerslev (1986) to the multivariate case that is able to estimate the correlation between financial actives in an apt manner.

One of the most popular approaches in the class of Dynamic Conditional Correlation models are the ones proposed by Tse and Tsui (2002) – that will be used later in this thesis –; Bollerslev (1990); Jeantheau (1998); Engle and Sheppard (2001); Engle (2002); Cappiello, Engle, and Sheppard (2006); Billio, Caporin, and Gobbo (2006), and Aielli (2013).

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 presented 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 used to work with financial returns by Galeano and Ausín (2010) using a finite mixture of normal distributions for the innovations.

Just like in the univariate case, multivariate innovations usually present a slight skewness and high kurtosis in this framework, which makes the multidimensional Skew-Slash distribution very suitable to perform the task of modeling these innovations because, thanks to its flexibility, it is able to capture both, skewness and heavy tails, simultaneously. In the present thesis, we propose the infinite mixture that is the d-dimensional Skew-Slash distribution.

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 is crucial in order to be able to deal properly with the kind of data we want to model because it allows us to establish the restrictions that the Dynamic Conditional Correlation model subjects us to.

Taking into account the mentioned features presented by financial data sets, as well as the power offered by the Skew-Slash distribution, we propose a Dynamic Conditional Correlation (DCC) model with d-dimensional Skew-Slash innovations, defined as

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

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 $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$, the information set available until time $t-1$, and $\boldsymbol{\varepsilon}_t$ is the innovation at time t.

We specify

$$\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t, \tag{4.1}$$

where

$$\mathbf{D}_t = diag\left(\left\{h_{it}^{1/2}\right\}_{i=1}^{d}\right) \in \mathbb{R}^{d\times d} \tag{4.2}$$

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

$$h_{it} = \omega_i + \alpha_i \left(y_{t-1,i} - \mu_i\right)^2 + \beta_i h_{t-1,i}, \tag{4.3}$$

$$\mathbf{R}_t = \left(1 - \theta_1 - \theta_2\right)\mathbf{R} + \theta_1 \mathbf{R}_{t-1} + \theta_2 \mathbf{I}, \tag{4.4}$$

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. 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.

4.1 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 $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $V(\boldsymbol{\varepsilon}_t) = \mathbf{I}$. Incorporating this restriction to (2.16) and (2.17), 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,

$$\boldsymbol{\eta} = -\sqrt{\frac{2}{\pi}\frac{\nu}{\nu-1}}\left\{\frac{\nu}{\nu-2}\mathbf{I} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}\boldsymbol{\delta}'\right\}^{-1/2}\boldsymbol{\delta} \tag{4.5}$$

and

$$\boldsymbol{\Sigma} = \left\{\frac{\nu}{\nu-2}\mathbf{I} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2\boldsymbol{\delta}\boldsymbol{\delta}'\right\}^{-1}. \tag{4.6}$$

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

We define a new matrix $\boldsymbol{\Gamma}$ that will help us to specify an explicit expression for $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$, and will allow for a more compact notation later on. Let us make

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

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

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

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

Introducing the parameter restrictions found in (4.5) and (4.6) into the alternative stochastic representation for the Skew-Slash distribution presented in (2.15), we can express the innovations alternatively as

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

where

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

It is useful to take into account the fact that, because $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$ is a symmetric positive definite matrix, we can define $\mathbf{H}_t^{1/2}$, such that $\mathbf{H}_t = \mathbf{H}_t^{1/2}\mathbf{H}_t^{1/2}$. Also, from now on we consider $a_{xt} = x_t - \sqrt{\frac{2}{\pi}\frac{\nu}{\nu-1}}$, for compacting purposes.

Let us assume that we have observed a series of returns $\{\mathbf{y}_t\}_{t=1}^T$, and let us define the set of parameters as $\vartheta = \left\{\boldsymbol{\delta},\nu,\boldsymbol{\mu},\{\omega_i,\alpha_i,\beta_i\}_{i=1}^d,\theta_1,\theta_2,\mathbf{R}\right\}$. Under the proposed model, the likelihood function is given by

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

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 thesis) 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

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

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

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

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

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

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 for (α_i, β_i) by means of a Dirichlet distribution for $(\alpha_i, \beta_i, 1 - \alpha_i - \beta_i)$, for $i \in \{1, \ldots, d\}$, and an analogous prior for (θ_1, θ_2) by means of a Dirichlet distribution for $(\theta_1, \theta_2, 1 - \theta_1 - \theta_2)$. 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:

$$
\begin{aligned}
f & \left(\boldsymbol{\delta} | \nu, \{\boldsymbol{\varepsilon}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T, \mathcal{H}\right) \\
& \propto \mathbf{I}\left(\|\boldsymbol{\delta}\|_2^2 < 1\right) \left[\frac{\frac{\nu}{\nu-2} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2 \boldsymbol{\delta}'\boldsymbol{\delta}}{1-\boldsymbol{\delta}'\boldsymbol{\delta}}\right]^{T/2} \\
& \exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2 \left\{\boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t - 2a_{xt}\boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta} + a_{xt}^2\boldsymbol{\delta}'\mathbf{M}_D^{-1}\boldsymbol{\delta}\right\}\right\}
\end{aligned} \tag{4.7}
$$

$$
\begin{aligned}
f & \left(\nu | \boldsymbol{\delta}, \{\boldsymbol{\varepsilon}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T, \mathcal{H}\right) \\
\propto & \ (\nu-2)^{a_\nu - 1} \exp\left\{-\frac{\nu}{b_\nu}\right\} \mathbf{I}(\nu > 2)\, \nu^T \left[\frac{\nu}{\nu-2} - \frac{2}{\pi}\left(\frac{\nu}{\nu-1}\right)^2 \boldsymbol{\delta}'\boldsymbol{\delta}\right]^{T/2} \left[\prod_{t=1}^T u_t^\nu\right] \\
& \exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2 \left\{\begin{array}{c} \boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\boldsymbol{\varepsilon}_t - 2a_{xt}\boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta} \\ +2\left(\frac{\nu}{\nu-1}\right)\left[\frac{1}{\pi}\left(\frac{\nu}{\nu-1}\right) - \sqrt{\frac{2}{\pi}}x_t\right]\boldsymbol{\delta}'\mathbf{M}_D^{-1}\boldsymbol{\delta} \end{array}\right\}\right\}
\end{aligned} \tag{4.8}
$$

$$
\begin{aligned}
& f\left(\boldsymbol{\mu} | \boldsymbol{\delta}, \nu, \{\omega_i, \alpha_i, \beta_i\}_{i=1}^d, \theta_1, \theta_2, \mathbf{R}, \{\mathbf{y}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T, \mathcal{H}\right) \\
& \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \mathbf{m}_\mu)' \mathbf{S}_\mu (\boldsymbol{\mu} - \mathbf{m}_\mu)\right\} \left[\prod_{t=1}^T \prod_{i=1}^d \frac{1}{\sqrt{h_{it}}}\right] \\
& \exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2 \left[\begin{array}{l} \mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\mathbf{y}_t \\ +\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu} \\ -2\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu} \\ -2a_{xt}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta} \\ +2a_{xt}\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta} \end{array}\right]\right\}
\end{aligned} \tag{4.9}
$$

$$
\begin{aligned}
&f\left(\{\omega_i\}_{i=1}^d|\boldsymbol{\delta},\nu,\boldsymbol{\mu},\{\alpha_i,\beta_i\}_{i=1}^d,\theta_1,\theta_2,\mathbf{R},\{\mathbf{Y}_t\}_{t=1}^T,\{X_t\}_{t=1}^T,\{U_t\}_{t=1}^T,\mathcal{H}\right)\\
&\propto\left\{\prod_{i=1}^d\omega_i^{1-a_\omega}\right\}\exp\left\{-\sum_{i=1}^d\frac{b_\omega}{\omega_i}\right\}\mathbf{I}\left(\{\omega_i\}_{i=1}^d>0\right)\left[\prod_{t=1}^T\prod_{i=1}^d\frac{1}{\sqrt{h_{it}}}\right]\\
&\qquad\exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2\left[\begin{array}{l}
\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\mathbf{y}_t\\
+\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\
-2\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\
-2a_{xt}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\\
+2a_{xt}\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}
\end{array}\right]\right\}
\end{aligned}
\tag{4.10}
$$

$$
\begin{aligned}
&f\left(\alpha_{i^*},\beta_{i^*}|\boldsymbol{\delta},\nu,\boldsymbol{\mu},\{\omega_i\}_{i=1}^d,\{\alpha_i,\beta_i\}_{\substack{i=1\\i\neq i^*}}^d,\theta_1,\theta_2,\mathbf{R},\{\mathbf{Y}_t,X_t,U_t\}_{t=1}^T,\mathcal{H}\right)\\
&\propto\alpha_{i^*}^{c_{\alpha\beta}p_\alpha-1}\beta_{i^*}^{c_{\alpha\beta}p_\beta-1}\left(1-\alpha_{i^*}-\beta_{i^*}\right)^{c_{\alpha\beta}\left(1-p_\alpha-p_\beta\right)-1}\\
&\quad\mathbf{I}\left(\alpha_{i^*}+\beta_{i^*}<1;\alpha_{i^*},\beta_{i^*}>0\right)\left[\prod_{t=1}^T\frac{1}{\sqrt{h_{i^*t}}}\right]\\
&\qquad\exp\left\{-\frac{1}{2}\sum_{t=1}^T u_t^2\left[\begin{array}{l}
\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\mathbf{y}_t\\
+\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\
-2\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\Gamma}^{1/2}\mathbf{H}_t^{-1/2}\boldsymbol{\mu}\\
-2a_{xt}\mathbf{y}_t'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}\\
+2a_{xt}\boldsymbol{\mu}'\mathbf{H}_t^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{M}_D^{-1}\boldsymbol{\delta}
\end{array}\right]\right\}
\end{aligned}
\tag{4.11}
$$

$$
\begin{aligned}
& f\left(\theta_1, \theta_2 | \boldsymbol{\delta}, \nu, \boldsymbol{\mu}, \{\omega_i, \alpha_i, \beta_i\}_{i=1}^d, \mathbf{R}, \{\mathbf{Y}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T, \mathcal{H}\right) \\
& \propto \theta_1^{c_\theta p_1 - 1} \theta_2^{c_\theta p_2 - 1} (1 - \theta_1 - \theta_2)^{c_\theta (1 - p_1 - p_2) - 1} \\
& \quad \mathbf{I}(\theta_1 + \theta_2 < 1; \theta_1, \theta_2 > 0) \left[\prod_{t=1}^T \frac{1}{\sqrt{\det(\mathbf{R}_t)}} \right] \\
& \quad \exp \left\{ -\frac{1}{2} \sum_{t=1}^T u_t^2 \left[\begin{array}{l}
\mathbf{y}_t' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\Gamma}^{1/2} \mathbf{H}_t^{-1/2} \mathbf{y}_t \\
+ \boldsymbol{\mu}' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\Gamma}^{1/2} \mathbf{H}_t^{-1/2} \boldsymbol{\mu} \\
- 2 \mathbf{y}_t' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\Gamma}^{1/2} \mathbf{H}_t^{-1/2} \boldsymbol{\mu} \\
- 2 a_{xt} \mathbf{y}_t' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\delta} \\
+ 2 a_{xt} \boldsymbol{\mu}' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\delta}
\end{array} \right] \right\}
\end{aligned}
\tag{4.12}
$$

$$
\begin{aligned}
& f\left(\mathbf{R} | \boldsymbol{\delta}, \nu, \boldsymbol{\mu}, \{\omega_i, \alpha_i, \beta_i\}_{i=1}^d, \theta_1, \theta_2, \{\mathbf{Y}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_t\}_{t=1}^T, \mathcal{H}\right) \\
& \propto \mathbf{I}(\mathbf{R} \ s.p.d. \text{ with unit diagonal}) \left[\prod_{t=1}^T \frac{1}{\sqrt{\det(\mathbf{R}_t)}} \right] \\
& \quad \exp \left\{ -\frac{1}{2} \sum_{t=1}^T u_t^2 \left[\begin{array}{l}
\mathbf{y}_t' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\Gamma}^{1/2} \mathbf{H}_t^{-1/2} \mathbf{y}_t \\
+ \boldsymbol{\mu}' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\Gamma}^{1/2} \mathbf{H}_t^{-1/2} \boldsymbol{\mu} \\
- 2 \mathbf{y}_t' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\Gamma}^{1/2} \mathbf{H}_t^{-1/2} \boldsymbol{\mu} \\
- 2 a_{xt} \mathbf{y}_t' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\delta} \\
+ 2 a_{xt} \boldsymbol{\mu}' \mathbf{H}_t^{-1/2} \boldsymbol{\Gamma}^{1/2} \mathbf{M}_D^{-1} \boldsymbol{\delta}
\end{array} \right] \right\}
\end{aligned}
\tag{4.13}
$$

$$
X_t | \boldsymbol{\delta}, \nu, \{\boldsymbol{\varepsilon}_t\}_{t=1}^T, \{X_1, \ldots, X_{t-1}, X_{t+1}, \ldots, X_T\}, \{U_t\}_{t=1}^T, \mathcal{H} \sim \mathcal{N}_1(\mu_x, \sigma_x) \mathbf{I}\left(\mathbb{R}^+\right),
\tag{4.14}
$$

where

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

and

$$
\sigma_x = u_t^{-1} \sqrt{1 - \boldsymbol{\delta}' \boldsymbol{\delta}}.
$$

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

$$U_t^2|\boldsymbol{\delta}, \nu, \{\boldsymbol{\varepsilon}_t\}_{t=1}^T, \{X_t\}_{t=1}^T, \{U_1, \ldots, U_{t-1}, U_{t+1}, \ldots, U_T\}, \mathcal{H} \sim \mathcal{G}\left(a_u, b_u\right) \mathbf{I}\left(0,1\right), \tag{4.15}$$

where

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

and

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

By writing $\mathbf{I}\left(0,1\right)$, we denote that the marginal posterior distribution for u_t^2 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 $x_0^{(m)} \sim \mathcal{N}_1\left(0,1\right)$, 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 (4.3), $\left\{\mathbf{D}_t^{(m)}\right\}_{t=1}^{T}$ as in (4.2), $\left\{\mathbf{R}_t^{(m)}\right\}_{t=1}^{T}$ as in (4.4), and $\left\{\mathbf{H}_t^{(m)}\right\}_{t=1}^{T}$ as in (4.1).

5. Compute $\boldsymbol{\eta}^{(m)}$ and $\boldsymbol{\Sigma}^{(m)}$ to satisfy the null mean and unit variance restrictions as specified in (4.5) and (4.6), 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 (4.7). 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 (4.8).

9. Obtain a sample $\boldsymbol{\mu}^{(m+1)}$ from the distribution in (4.9). Execute step 4 alone.

10. Obtain a sample $\boldsymbol{\omega}^{(m+1)}$ from the distribution in (4.10). Execute step 4 alone.

11. Obtain samples $\alpha_{i^*}^{(m+1)}, \beta_{i^*}^{(m+1)}$ from the distribution in (4.11) 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 (4.12). Execute step 4 alone.

13. Obtain a sample $\mathbf{R}^{(m+1)}$ from the distribution in (4.13). Execute step 4 alone.

14. Obtain a sample $x_t^{(m+1)}$ from the distribution in (4.14).

15. Obtain a sample $u_t^{(m+1)}$ from the distribution in (4.15).

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 candidates for $\boldsymbol{\delta}$ through a transformation of a normal random vector to ensure that the restriction is satisfied. The candidates for ν, $\boldsymbol{\omega}$, (α_i, β_i) for $i \in \{1, \dots, d\}$, and $\boldsymbol{\theta}$ 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.

4.2 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 couple of real data sets. The first one contains information about the joint behavior of the Dow Jones and the NASDAQ indices; the second one is formed by the daily log-returns of the DAX, CAC40, and Nikkei indices. In every case, 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 for 20000 iterations to burn in and 25000 in equilibrium for the real data.

4.2.1 Simulated examples

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

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 both diagonal elements equal to 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 null mean and unit variance of the innovations.

Figure 4.1 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 4.2 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 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 4.3 shows the joint theoretical density of the innovations in 4.3(a) next to their joint predictive density in 4.3(b), while 4.4 shows the contour plot of the joint theoretical density of the innovations in 4.4(a) together with their contour plot

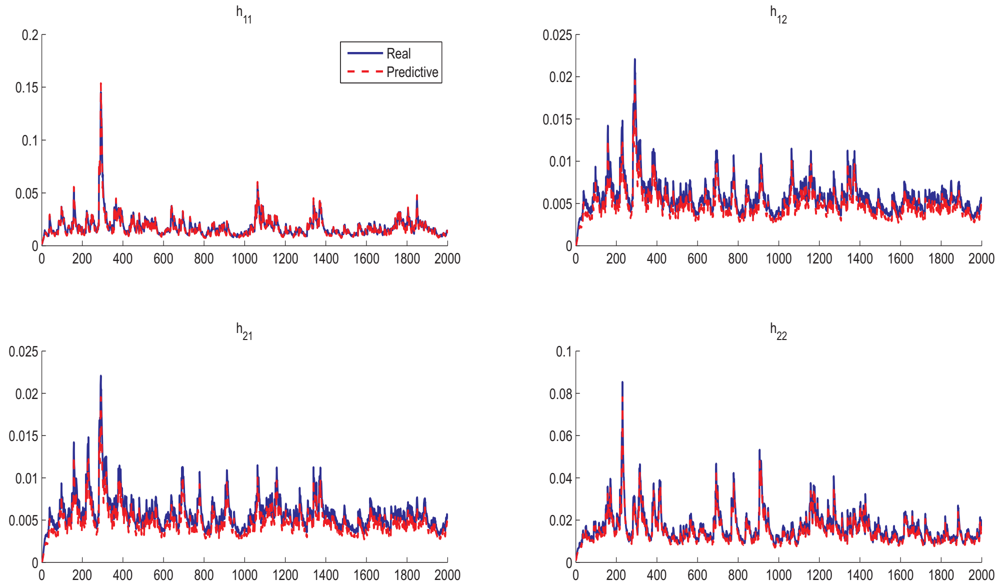


Figure 4.1: True (solid line) and fitted (dotted line) volatilities for the 2-dimensional simulated data.

of the joint predictive density in 4.4(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, also keep in mind that $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ are set to verify the null mean and unit variance of the innovations.

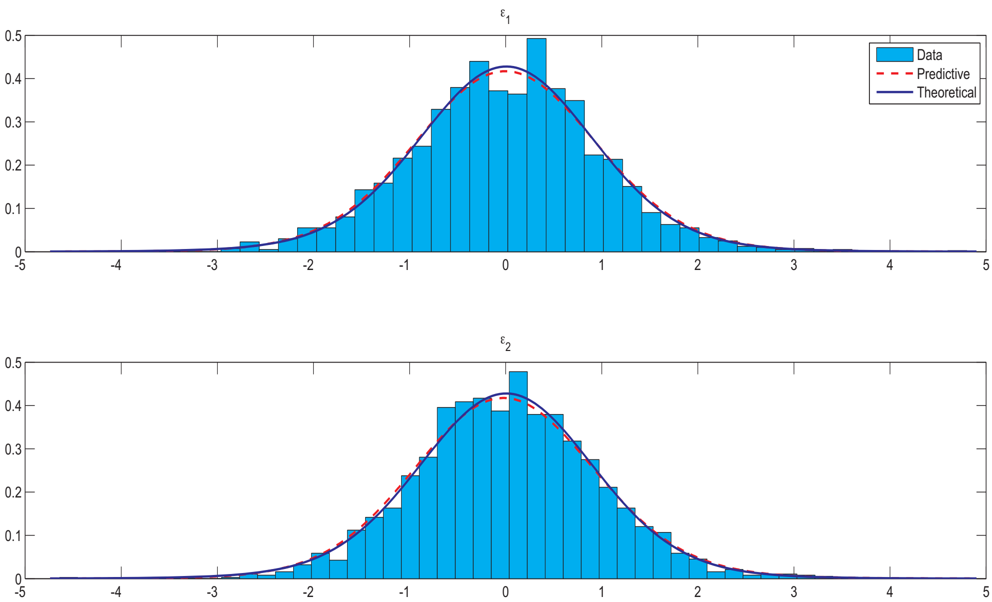
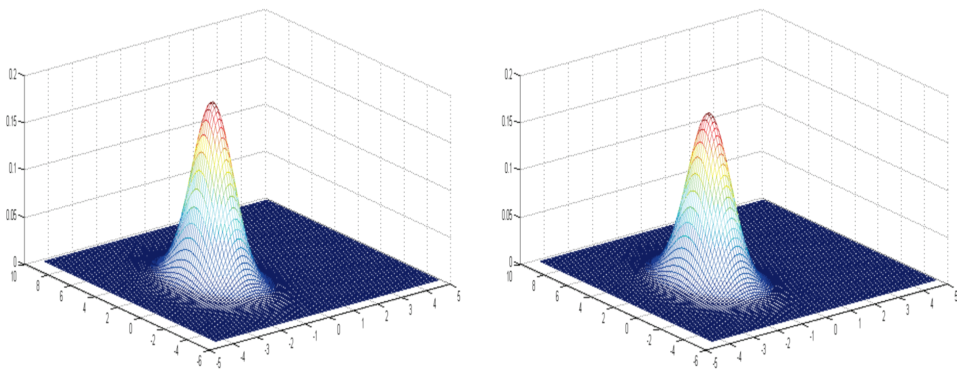
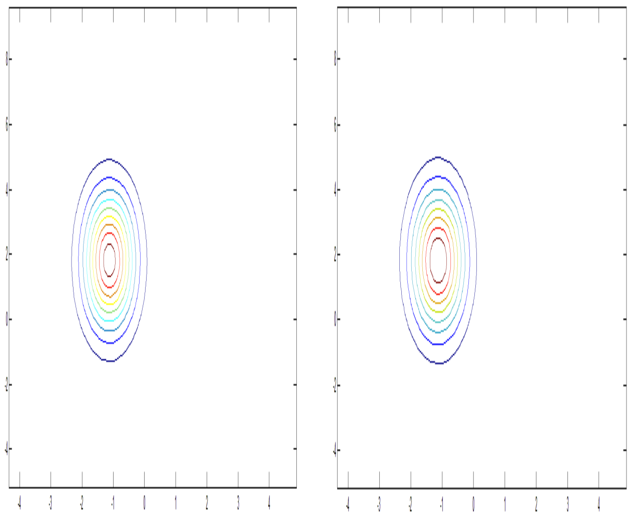


Figure 4.2: True (solid line) and fitted (dotted line) predictive innovation marginal densities compared to their histograms for the 2-dimensional data set.



(a) (b)

Figure 4.3: True (a) and fitted (b) predictive innovation joint densities for the 2-dimensional simulated data.



(a) (b)

Figure 4.4: Contour plots of the true (a) and fitted (b) predictive innovation densities for the 2-dimensional simulated data.

Figure 4.5 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 4.6 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 4.7 shows the 2-variable marginal theoretical densities of the innovations in 4.7(a) for $(\varepsilon_1, \varepsilon_2)$, 4.7(c) for $(\varepsilon_1, \varepsilon_3)$, and 4.7(e) for $(\varepsilon_2, \varepsilon_3)$ next to their 2-variable marginal predictive densities in 4.7(b) for $(\varepsilon_1, \varepsilon_2)$, 4.7(d) for $(\varepsilon_1, \varepsilon_3)$, and 4.7(f) for $(\varepsilon_2, \varepsilon_3)$.

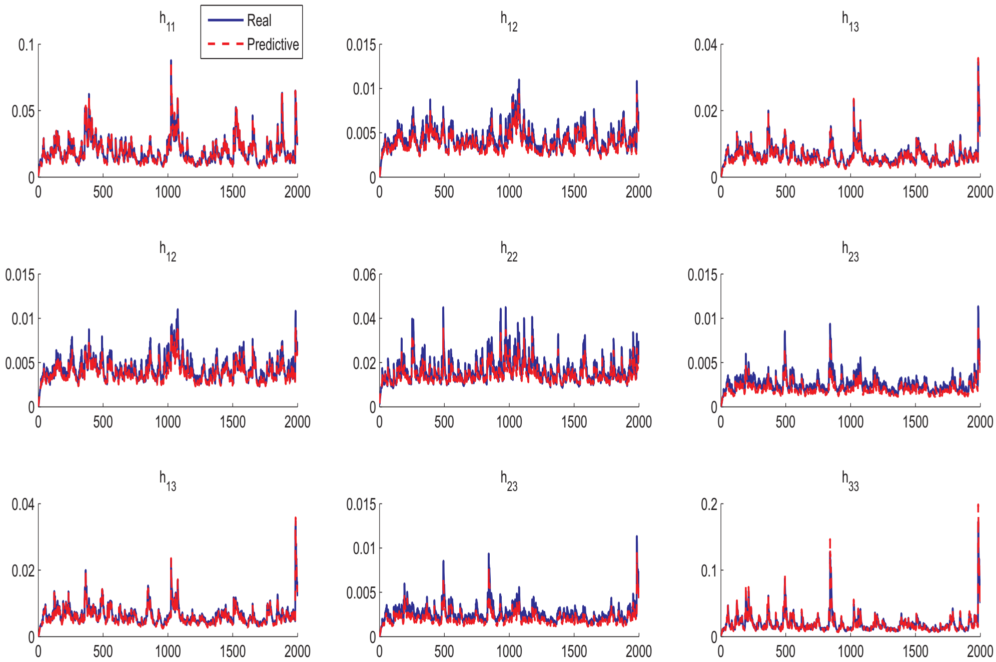


Figure 4.5: True (solid line) and fitted (dotted line) volatilities for the 3-dimensional data set.

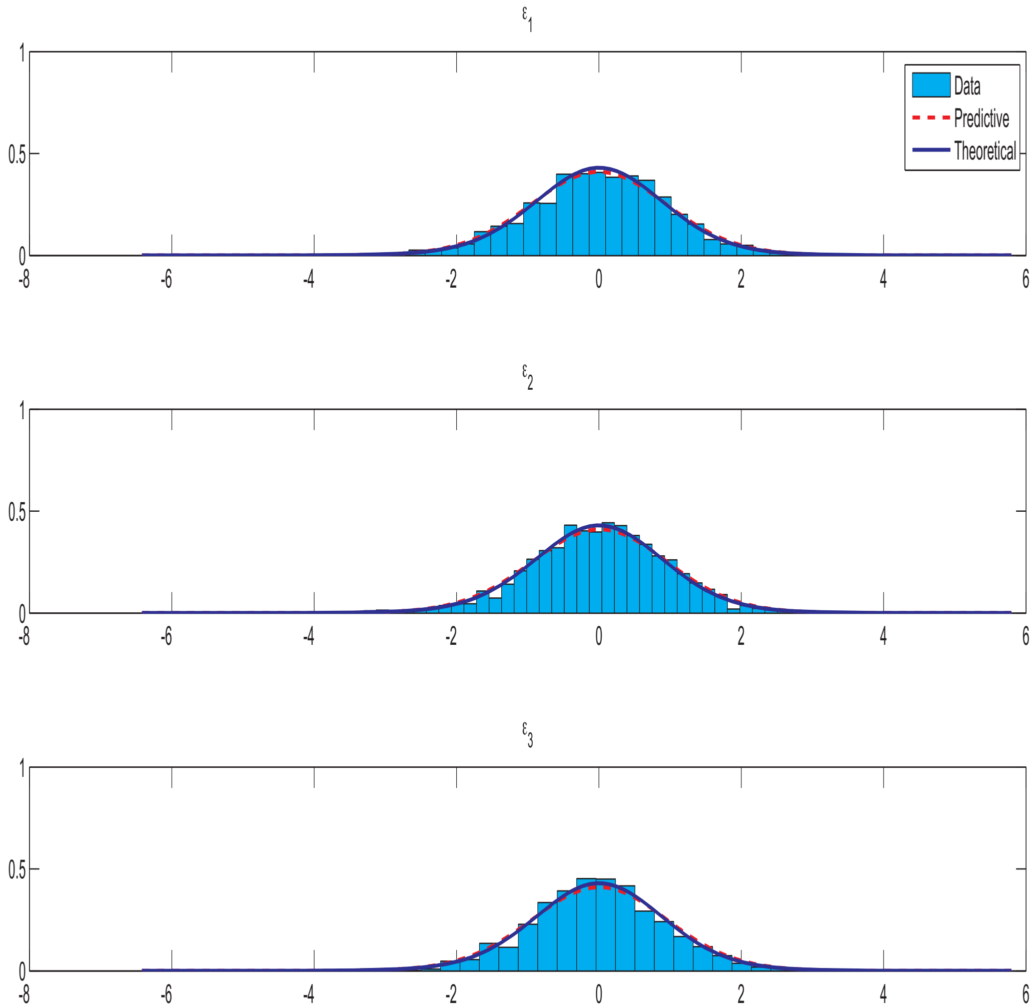
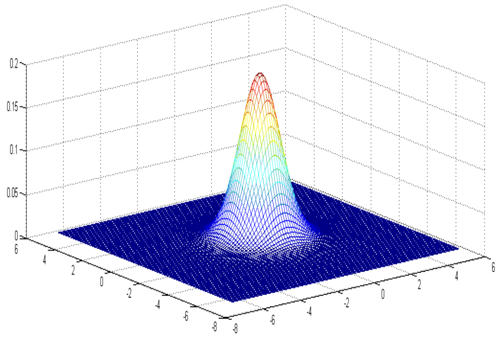
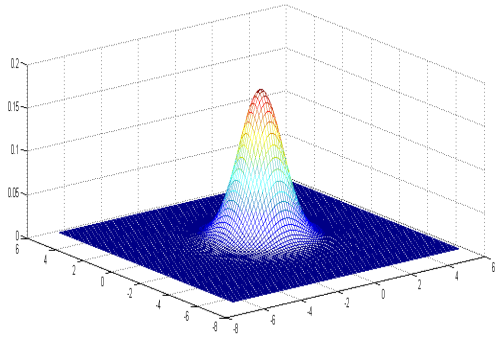


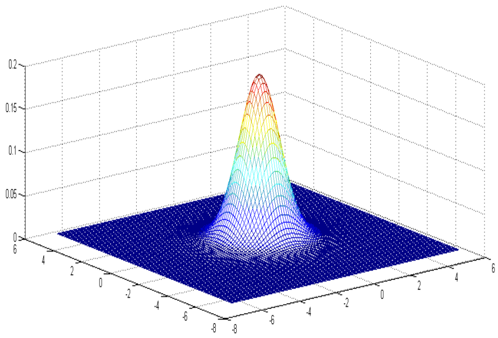
Figure 4.6: True (solid line) and fitted (dotted line) predictive innovation marginal densities compared to their histograms for the 3-dimensional data set.



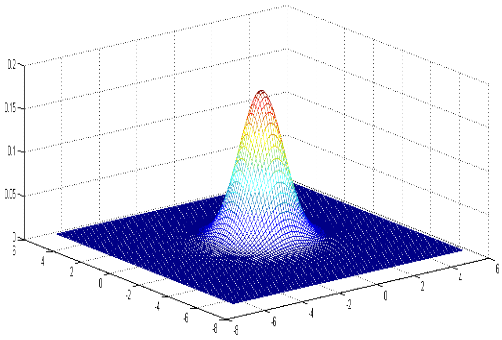
(a)



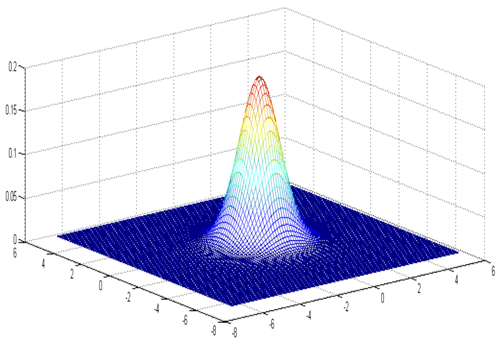
(b)



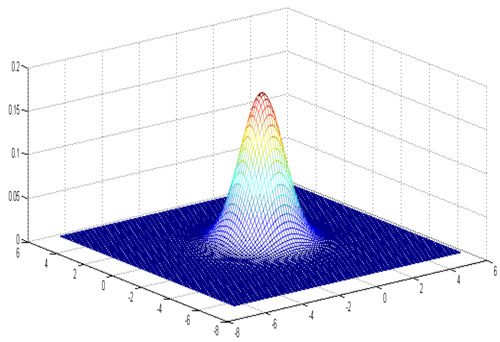
(c)



(d)

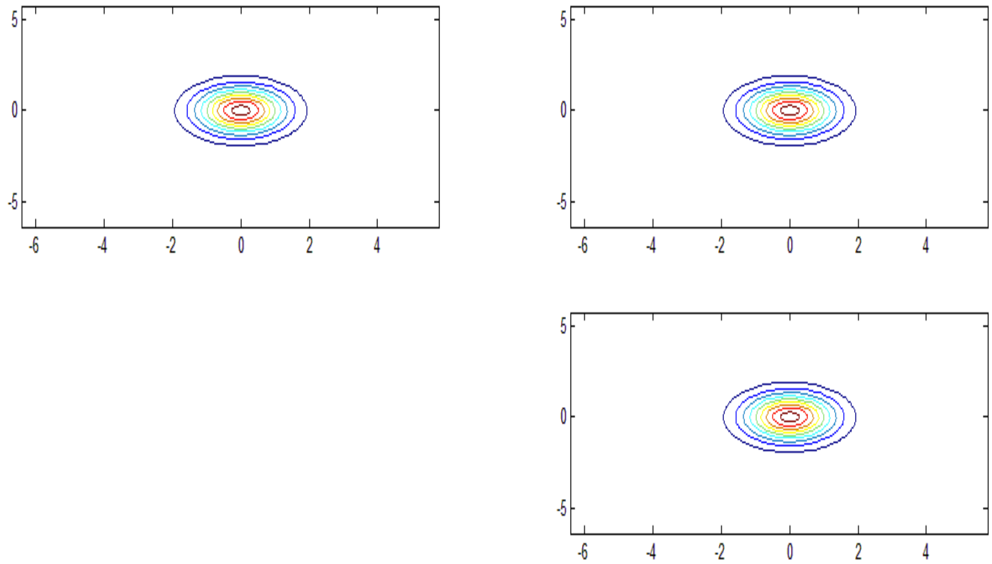


(e)

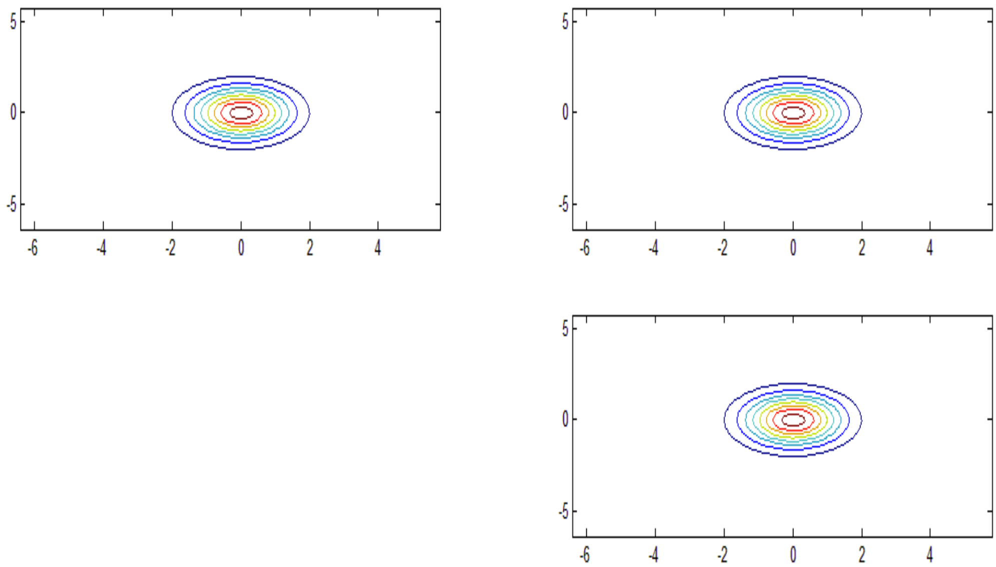


(f)

Figure 4.7: True (left) and fitted (right) predictive innovation 2-variable marginal densities for the 3-dimensional simulated data.



(a)



(b)

Figure 4.8: Contour plots of the true (a) and fitted (b) predictive innovation 2-variable marginal densities for the 3-dimensional simulated data.

Figure 4.8 shows the contour plots of the 2-variable marginal theoretical densities of the innovations in 4.8(a) together with their contour plots of the 2-variable marginal predictive densities in 4.8(b). From both sets of figures, we can see that, even though we are not able to superimpose them, their similarity is clear.

4.2.2 Real data examples

To illustrate the usefulness of the approach for the multivariate modeling of financial returns proposed in the present thesis, in this section we analyze two real data sets.

First of all, we analyze the daily closing prices of the Dow Jones and NASDAQ New Yorker stock market indices. Second of all, we analyze the daily log-returns of the German DAX, the French CAC40, and the Japanese Nikkei stock market indices.

In each case, we obtained fitted volatilities as the mean value of the fitted volatilities of the 25000 parameter values of the MCMC sampler.

Dow Jones - NASDAQ

We begin by analyzing the daily closing prices of the Dow Jones and NASDAQ stock market indices, from January 2$^{\text{nd}}$, 1996 to December 29$^{\text{th}}$, 2006, which leads to 2769 observations. Figure 4.9 shows the plot of the time series generated by the simple returns of both indices, where we rename the Dow Jones index as the first component of our variable, Y_1, and, analogously, we rename the NASDAQ index as Y_2. We can see that this data clearly needs to be modeled in a manner that is able to capture heavy tails.

The univariate sample means, standard deviations, skewness coefficients,

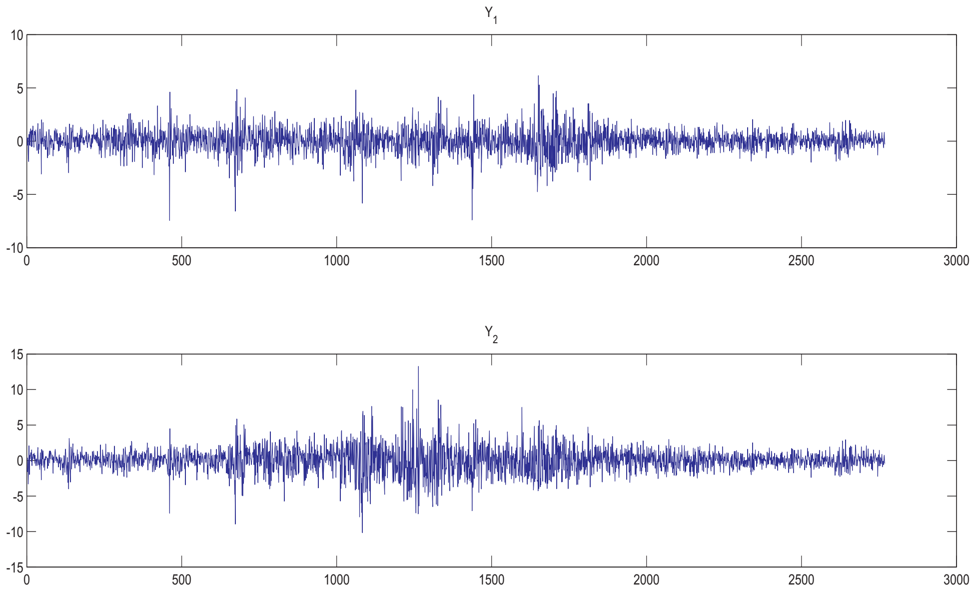


Figure 4.9: Dow Jones and NASDAQ returns between January 2^{nd}, 1996 and December 29^{th}, 2006.

and kurtosis coefficients of the return series are $(0.0317, 0.0298)'$, $(1.088, 1.7614)'$, $(-0.2961, 0.1559)'$, and $(10.0587, 69.0901)'$, respectively, from which we may believe that, even though the skewness is very slight, the kurtosis are clearly very large, especially the one presented by the NASDAQ index. This indicates that the normal distribution could not suffice in the task of modeling this data; hence, the proposal of modeling the data by means of a 2-dimensional Skew-Slash DCC process seems appropriate.

Figure 4.10 shows the fitted volatilities, while Figure 4.11 shows the estimated innovation marginal densities. On the other hand, in Figure 4.12 we can see the joint predictive density of the innovations, and Figure 4.13 shows its corresponding contour plot.

We wanted to go further and be able to compare our model with another

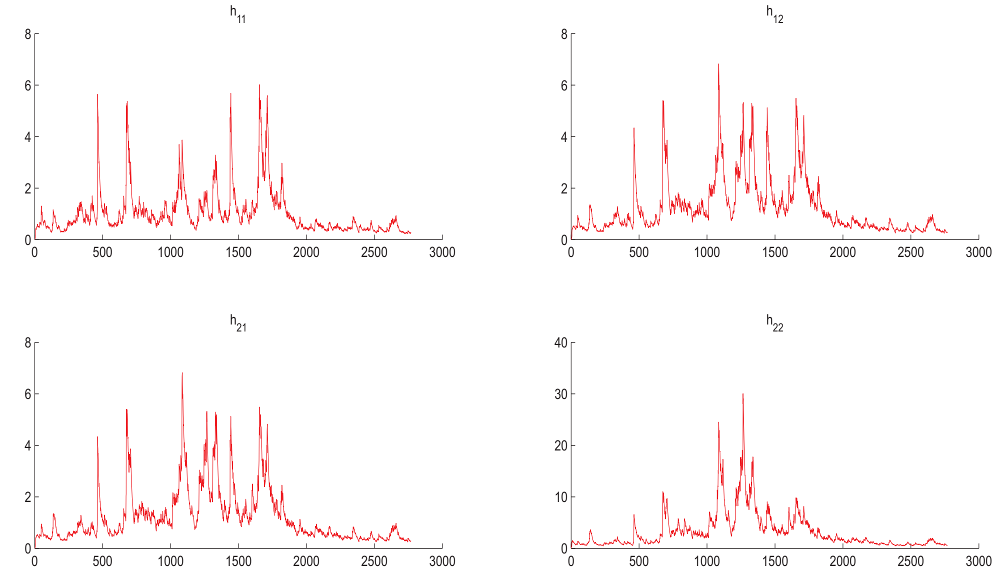


Figure 4.10: Volatility estimates for the Dow Jones and NASDAQ returns.

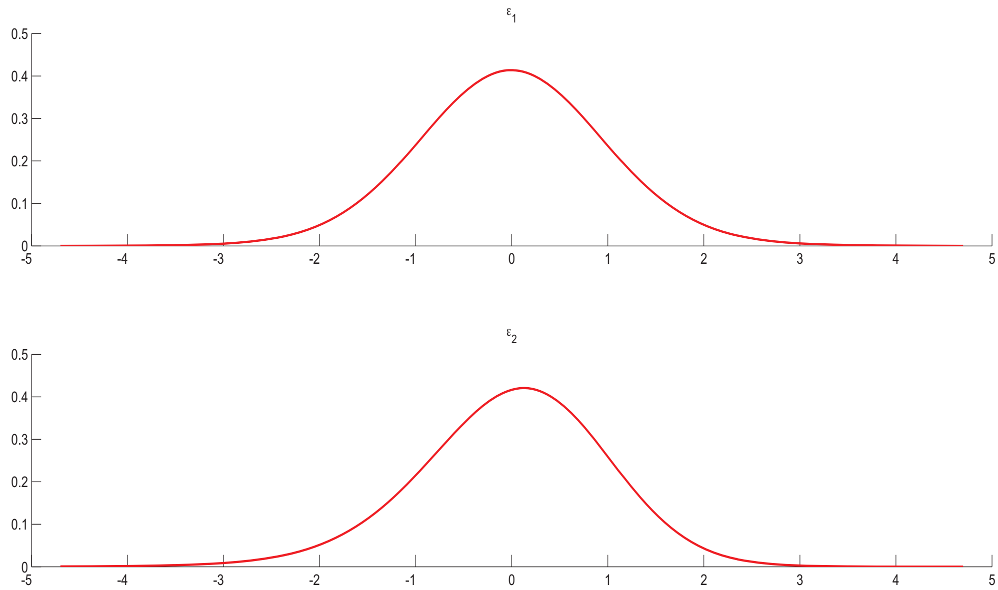


Figure 4.11: Estimated innovation marginal densities for the Dow Jones and NASDAQ returns.

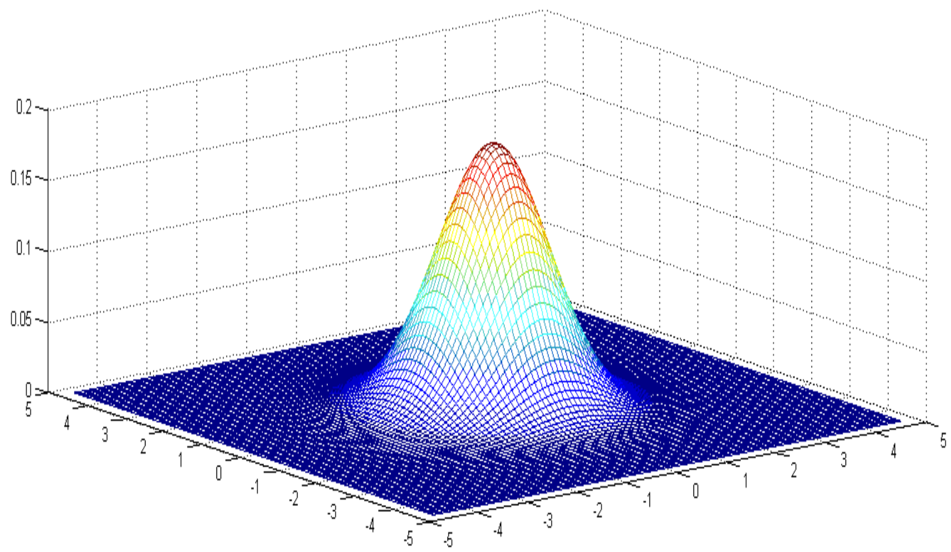


Figure 4.12: Estimated innovation joint density for the Dow Jones and NASDAQ returns.

one and, that way, prove the effectiveness of our methodology. To do this, we decided to take into account the data set presented by Fioruci, Ehlers, and Andrade (2014), and evaluate the performance of our model fitting this information.

DAX - CAC40 - Nikkei

The original information consists on the daily closing prices of the stock market indices in Frankfurt (DAX), Paris (CAC40), and Tokyo (Nikkei) from October 10th, 1991 through December 30th, 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

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

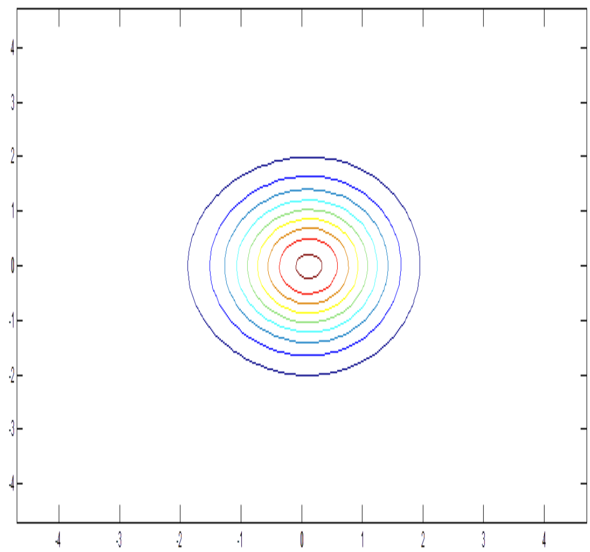


Figure 4.13: Contour plot of the estimated joint density for the innovations of the Dow Jones and NASDAQ returns.

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 4.14 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.

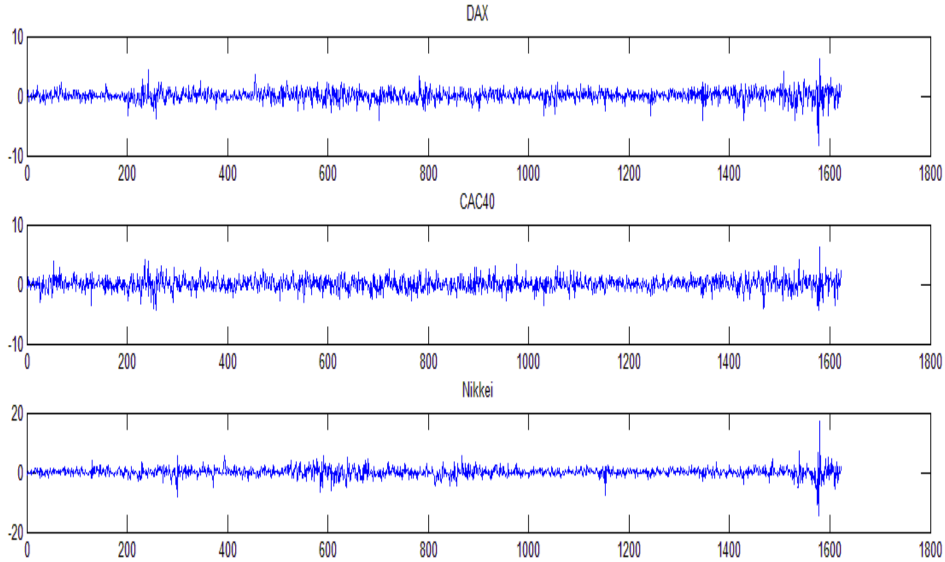


Figure 4.14: DAX, CAC40, and Nikkei log-returns between October 10$^{\text{th}}$, 1991 and December 30$^{\text{th}}$, 1997.

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 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 4.15 shows the fitted volatilities, while Figure 4.16 shows the estimated innovation marginal densities. On the other hand, in Figure 4.17 we

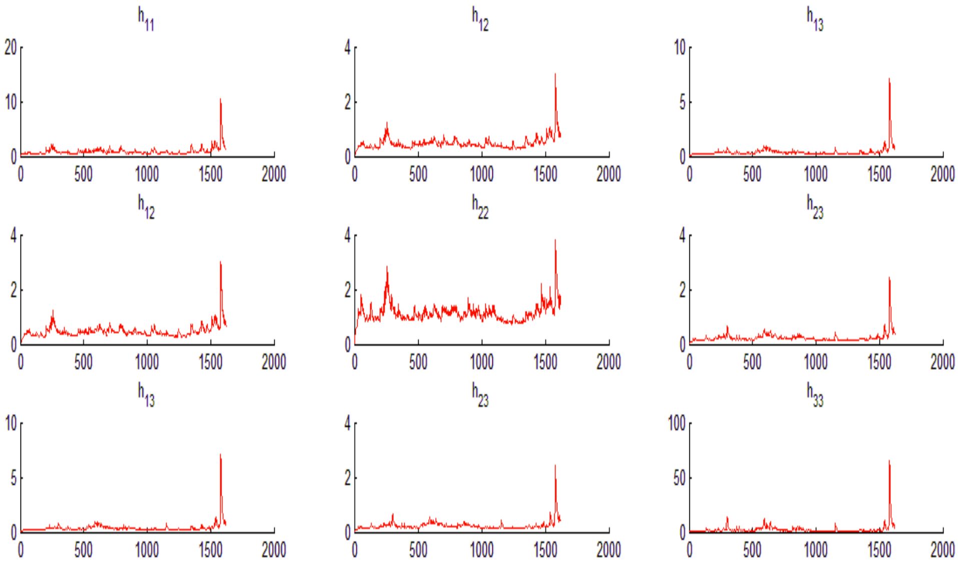


Figure 4.15: Volatility estimates for the DAX, CAC40, and Nikkei log-returns.

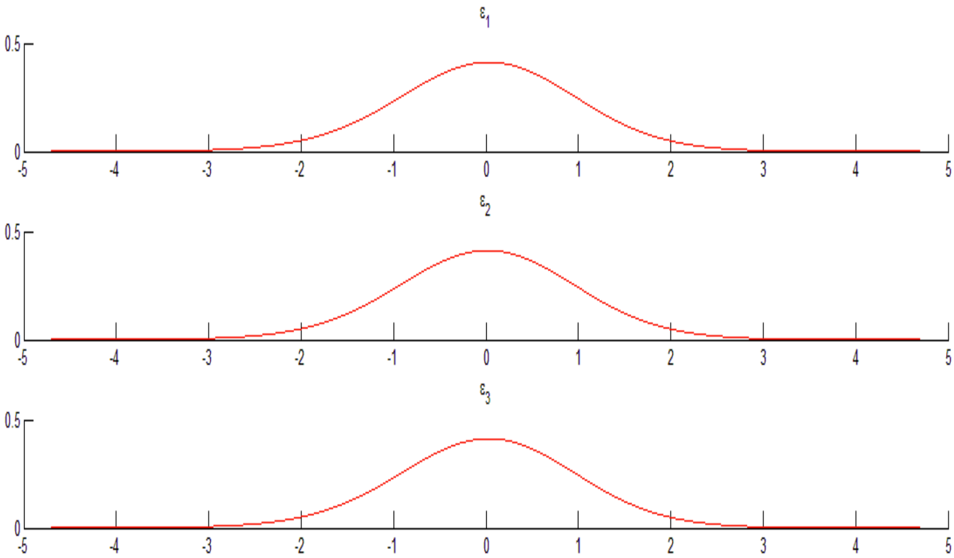
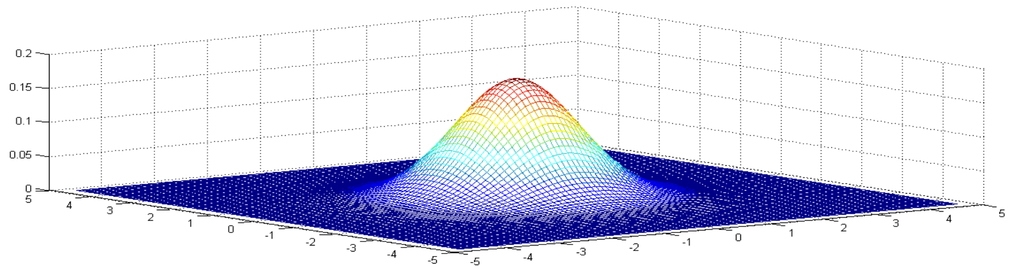


Figure 4.16: Estimated innovation marginal densities for the DAX, CAC40, and Nikkei log-returns.

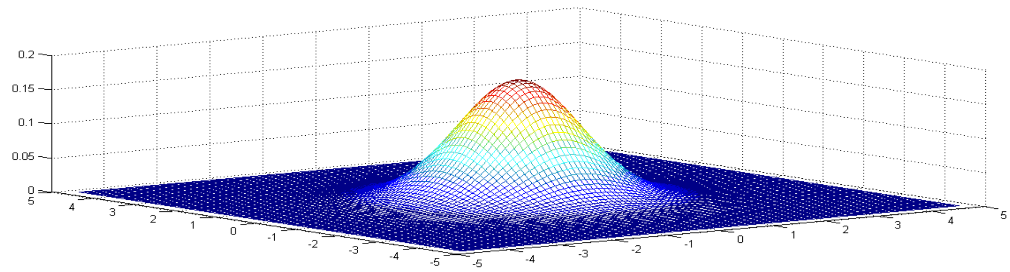
(a)

DAX-CAC40



(b)

DAX-Nikkei



(c)

CAC40-Nikkei

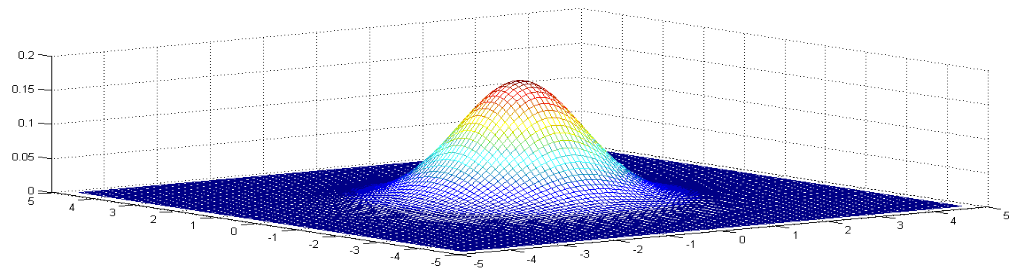


Figure 4.17: Estimated innovation 2-variable marginal densities for the DAX, CAC40, and Nikkei log-returns.

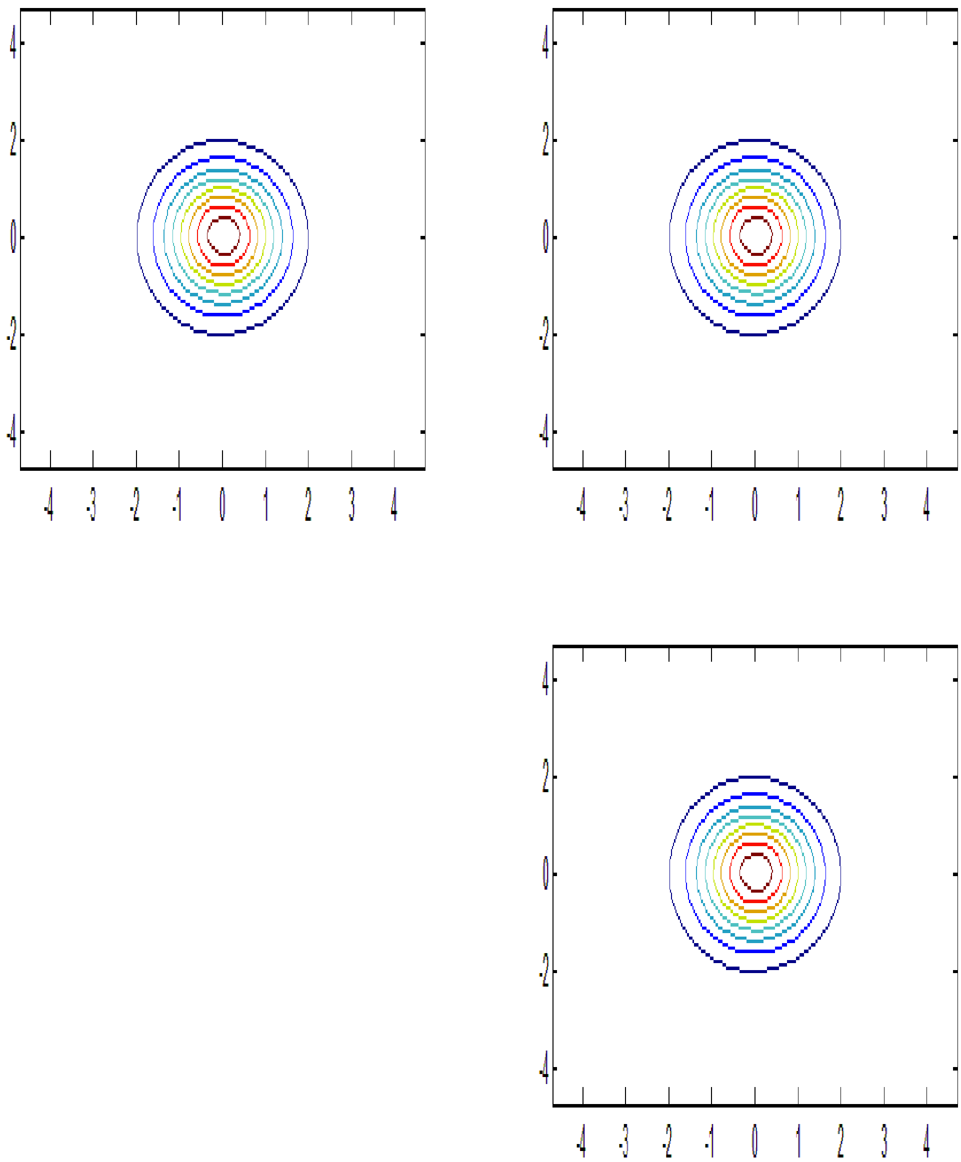


Figure 4.18: Contour plots of the estimated innovation 2-variable marginal densities for the innovations of the DAX, CAC40, and Nikkei log-returns.

can see the 2-variable marginal predictive densities of the innovations, and Figure 4.18 shows their corresponding contour plots.

Now, we proceed to compare the performance of our model in front of the one proposed by Fioruci, Ehlers, and Andrade (2014). We used the same data set that they used, and we transformed it in the way they explained their transformation. 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). 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), a measurement already provided in their paper, for comparison.

The Deviance Information Criterion, as defined by Spiegelhalter, Best, Carlin, and van der Linde (2002), is given by

$$DIC\left(M\right) = 2E\{D\left(\vartheta; M\right)\} - D\left(E\{\vartheta\}; M\right),$$

where $\vartheta; M$ denote, in this case, the set of parameters, ϑ, for the model in question, M, and $D\left(\cdot;\cdot\right)$ denotes the deviance function, defined as

$$D\left(\vartheta; M\right) = -2\ln\left\{\mathcal{L}\left(\vartheta; M\right)\right\}.$$

Let us remember that $\mathcal{L}$ denotes the likelihood function.

In table 4.1 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 4.1: 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.

Chapter 5

Conclusions and future lines of research

5.1 Conclusions

In this thesis, we have studied several probability distributions that generalize in one way or another the normal distribution to incorporate skewness or kurtosis, which are typical features of financial data.

For the univariate case, we proposed to model the returns of a single dimensional financial asset by means of a Generalized Autoregressive Conditional Heteroskedastic process with Skew-Slash innovations. Developing our proposal, we showed a Maximum Likelihood approach as well as a Bayesian method for the estimation of our model.

To illustrate the power of the proposed models and methodologies, we recurred to several data sets. Firstly, we worked with several simulated data sets of different sizes to illustrate the capability of the Maximum Likelihood approach and then we picked one of the simulated data sets to assess the

Bayesian methodology and to compare both proposals. Second, we studied a series of log-returns drawn from the Standard & Poor's index between January 3rd, 2000 and December 28th, 2013 using both methods, and compared them.

For the multivariate case, our proposal was to model the structure of a multidimensional time series of financial returns by means of the Dynamic Conditional Correlation model with Skew-Slash innovations. 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 a number of data sets. We started by fitting a 2-dimensional simulated series of returns, as well as an analogous 3-dimensional data set and compared true features such as the marginal densities, volatilities, and even the joint density in the two-dimensional case, to their estimated pairs, and we found very good results. Later on, we estimated two real data sets.

Fist of all, we fitted a 2-dimensional financial time series that contains the returns of the Dow Jones and NASDAQ stock market indices between January 2nd, 1996 and December 29th, 2006; nevertheless, we realized that, with the information available, we could not find a way to evaluate the performance of our methodology. To respond to this issue, we decided to fit a second data set consistent on a 3-dimensional time series that contains the log-returns of the DAX, CAC40, and Nikkei stock market indices between October 10th, 1991 and December 30th, 1997.

This data has already been studied by Fioruci, Ehlers, and Andrade (2014), except they use other members of the Skew-Normal/independent family of distributions, as well as some Normal/independent distributions, and the Gaussian one, and provide the values of the Deviance Information Criterion (DIC) that they obtained with their estimations; therefore, we decided to compute the DIC inherent to our model estimation, so we could compare the models.

In this case, the model selection criterion manifested a better performance of our model capturing the essence of the data set that we were dealing with, 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 similar performances fitting their features, but maybe each data set has its own better partner, to call it some way.

5.2 Future lines of research

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

First of all, we realize that it is not always realistic to assume that the autocorrelations will only be significant for one lag; thus, we believe that it would be very interesting to drop this restriction, present all along our work. In the univariate case, the idea would be to go from the GARCH$(1,1)$ process to the more general GARCH(p,q), while maintaining the Skew-Slash innovations. For the multivariate case, we would have to remain open to the possibility of a Dynamic Conditional Correlation model with Skew-Slash innovations, but, in this case, basing the structure in a model of the form GARCH$(\{p_i, q_i\}_{i=1}^{d})$, that would allow for a much more flexible structure not only in terms of lags, but in terms of the different individual behaviors.

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.

Third, we have already manifested our belief that, because all members of the Skew-Normal/independent family of distributions share some of their features, it is very likely that, in certain frameworks, they might behave in similar ways, but we think that it is possible that, for every data set, we might be able to find one distribution that suits the information better than the others in the family.

Specifically speaking, we know that the Skew-Normal/independent family is built to be able to capture skewness and kurtosis in some environments, such as financial data sets. In this case, the idea would be to estimate several comparable conditional heteroskedastic models (like the GARCH or the Dynamic Conditional Correlation model) using different elements of the Skew-Normal/independent family of distributions to model the innovations, and use a model selection criterion to decide, in every case, what distribution constitutes a better match for the data under study.

On the other hand, we know that skewness and kurtosis do not only present themselves simultaneously in financial returns, and we believe it could be interesting to apply the distributions we have studied to model them. For example, it might be possible to classify galaxies by studying the structure and distribution of the stars that compose them without having to directly look at them, or maybe even to use this features in some medical field.

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.

Appendix A

Proof of the second part of Proposition 2 First, $W-E(W)$ can be written as

$$W - E(W) = \sigma \left[U^{-1}Z - E\left(U^{-1}\right) E(Z)\right],$$

leading to

$$\begin{aligned}
[W - E(W)]^k &= \sigma^k \left[U^{-1}Z - E\left(U^{-1}\right) E(Z)\right]^k \\
&= \sigma^k \sum_{j=0}^{k} (-1)^{k-j} \tbinom{k}{j} U^{-j} Z^j E\left(U^{-1}\right)^{k-j} E(Z)^{k-j}
\end{aligned}$$

Now, taking expectations in the previous equation,

$$\begin{aligned}
m_k(W) &= \sigma^k E\left[\sum_{j=0}^{k} (-1)^{k-j} \tbinom{k}{j} U^{-j} Z^j E\left(U^{-1}\right)^{k-j} E(Z)^{k-j}\right] \\
&= \sigma^k \sum_{j=0}^{k} (-1)^{k-j} \tbinom{k}{j} E\left(U^{-j}\right) E\left(Z^j\right) E\left(U^{-1}\right)^{k-j} E(Z)^{k-j} \\
&= 2^{(k-2)/2} \sigma^k \left(\tfrac{1}{1+\lambda^2}\right)^{k/2} \sum_{j=0}^{k} \sum_{i=0}^{j} b_{ijk} \lambda^{k-j+i},
\end{aligned}$$

for $\nu > \frac{k}{2}$, where

$$b_{ijk} = (-1)^{k-j} \binom{k}{j} \frac{\nu}{\nu - j} \left(\frac{\nu}{\nu - 1}\right)^{k-j} \pi^{-(k-j+2)/2} a_{ij},$$

with k fixed, $j \in \{0, 1, \ldots, k\}$, and $i \in \{0, 1, \ldots, j\}$, respectively.

Finally, the two sums can be reduced as follows.

$$m_k(W) = 2^{(k-2)/2} \sigma^k \left(\frac{1}{1+\lambda^2}\right)^{k/2} \sum_{l=0}^{k} c_{lk} \lambda^l,$$

where

$$c_{lk} = \sum_{m=0}^{l} b_{m,k-l+m,k},$$

and

$$b_{m,k-l+m,k} =$$
$$(-1)^{l-m}\left[1+(-1)^{k-l}\right]\pi^{-(l-m+2)/2}\frac{\nu}{\nu-(k-l+m)}\left(\frac{\nu}{\nu-1}\right)^{l-m}$$
$$\binom{k}{l}\binom{l}{m}\Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{k-l+1}{2}\right),$$

for fixed k, fixed $l \in \{0,1,\ldots,k\}$, and $m \in \{0,1,\ldots,l\}$, respectively. ■

Bibliography

- G. P. Aielli. Dynamic Conditional Correlation: On Properties and Estimation. *Journal of Business & Economic Statistics*, 31(3):282–299, 2013.
- H. Akaike. Canonical Correlation Analysis of Time Series and the use of an Information Criterion. *Mathematics in Science and Engineering*, 14(153):5–9, 1976.
- C. O. Alexander and A. M. Chibumba. Multivariate Orthogonal Factor GARCH. *University of Sussex Discussion Papers in Mathematics*, 1997.
- D. F. Andrews and C. L. Mallows. Scale Mixtures of Normal Distributions. *Journal of the Royal Statistical Society. Series B*, 36:99–102, 1974.
- 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.

- K.-H. Bae, G. A. Karolyi, and R. M. Stulz. A New Approach to Measuring Financial Contagion. *The Review of Financial Studies*, 16:717–763, 2003.

- 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.

- C. R. Barbosa Cabral, V. H. Lachos, and M. O. Prates. Multivariate Mixture Modeling Using Skew-Normal Independent Distributions. *Computational Statistics & Data Analysis*, 56:126–142, 2012.

- L. Bauwens and S. Laurent. A New Class of Multivariate Skew Densities, With Application to Generalized Autoregressive Conditional Heteroscedasticity Models. *Journal of Business & Economic Satatistics*, 23:346–354, 2005.

- L. Bauwens, S. Laurent, and J. V. K. Rombouts. Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

- 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.

- M. Branco and D. Dey. A General Class of Multivariate Skew-Elliptical Distributions. *Journal of Multivariate Analysis*, 79:93–113, 2001.

- 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.

- M. Carrasco and X. Chen. Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory*, 18(1):17–39, 2002.

- 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.

- P. H. Franses and C. Hafner. A Generalized Dynamic Conditional Correlation Model for Many Asset Returns. *Economic Reviews*, 28:612–631, 2009.

- 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.

- C. G. de la Fuente, P. Galeano, and M. P. Wiper. Bayesian Estimation of Dynamic Conditional Correlation Models With Skew-Slash Innovations. *UC3M Working papers. Statistics and Econometrics*, 2014.

- 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, editor. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality.* Chapman & Hall / CRC, 2004.

- 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.

- M. G. Genton and N. Loperfido. Generalized Skew-Elliptical Distributions and their Quadratic Forms. *Annals of the Institute of Statistical Mathematics*, 57:389–401, 2005.

- L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the Realtion Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48:1779–1801, 1993.

- B. Hansson and P. Hördahl. Testing the Conditional CAPM Using Multivariate GARCH-M. *Applied Financial Economics*, 8(4):377–388, 1998.

- N. Henze. A Probabilistic Representation of the 'Skew-Normal' Distribution. *Scandinavian Journal of Statistics*, 13:271–275, 1986.

- 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, P. Ghosh, and R. B. Arellano Valle. Likelihood Based Inference for Skew-Normal Independent Linear Mixed Models. *Statistica Sinica*, 20:303–322, 2010.

- 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.

- L. Ng. Tests of the CAPM with Time-Varying Covariances: A Multivariate GARCH Approach. *The Journal of Finance*, 46:15071521, 1991.

- A. O'Hagan and T. Leonard. Bayes Estimation Subject to Uncertainty About Parameter Constraints. *Biometrika*, 63:201–203, 1976.

- G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6:461–464, 1978.

- 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.

- R. S. Tsay. *Analysis of Financial Time Series, Third Edition*. John Wiley & Sons, 2010.
- 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.
- J. M. Zakoian. Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control*, 18:931–955, 1994.