Finite-sample inference on NGARCH models with flexible distributions

Author: Emmanuel Djanga Nsumbu
Student number: 1765701

Supervisor: Dr. Francisco J. Rubio
Module Code: 7CCMMS50

2018/19

Dissertation submitted in partial fulfilment of the requirements for the degree of MSc in Mathematics.

Abstract

This paper aims to investigate the use of the Exponential Power distribution (EPD), a parametric flexible distribution, in the context of autoregressive models (NGARCH-EPD). The EPD represents an instance of the flexible distributions implemented by Zhu and Galbraith [23] in the context of NGARCH models. The EPD allows for capturing departures from the usual assumption of normality in terms of heavier and lighter tails than those of the normal distribution. A simulation study is presented in order to investigate the finite sample properties of the resulting NGARCH-EPD model both in the frequentist and the Bayesian frameworks. The inference in NGARCH-EPD model in the latter approach has not been done before.

Zhu and Galbraith [23] present a practical application of flexible distributions, while this dissertation covers theoretical and empirical aspects of the same distributions. Therefore, this paper is complementary to the work of Zhu and Galbraith [23]. The results presented in this work can be of interest for industry practitioners because they provide some guidelines on the use of these distributions in practice and their benefits to volatility modelling. The R codes required to implement these models as well as the simulation studies presented in this work are original contributions of the author.

Acknowledgements

I would like to express my profound gratitude to my supervisor, Dr. Francisco J. Rubio, for accepting to supervise me, introducing me to flexible distributions, his precious help, and his valuable advice during this project. I would also like to thank the Mathematics department for offering such an enjoyable and intellectually stimulating programme and my employer, UBS, as well as my line manager, Iliana Anagnou, for allowing me to work on this paper on a part-time basis along my job and my colleagues for encouraging me throughout the summer. Last but not least, I would like to express my gratitude to my family who has always been there to support me in good and bad times as well as my friends that have contributed to my postgraduate experience.

Contents

Comments on Notation

In order to ensure a clear comprehension and a fluid read of the concepts developed in the upcoming pages, the notation that has been used throughout this paper is commented on.

Italic minuscule letters, *e.g.* n, correspond to regular scalars. Random variables are denoted by an italic majuscule letter, *e.g.* Y and a supplemental index is adjoined to each term of a sequence of random variables, *i.e.* $Y_1, ..., Y_n$. Estimates are random variables used to estimate particular parameters based on a sample. Any lower case Greek letter, *e.g.* θ, denotes one true parameter value, while $\hat{\theta}$ is its corresponding estimates. Additionally, $\boldsymbol{\theta}$ is interpreted as a column-vector of true parameter values, and $\hat{\boldsymbol{\theta}}$ is considered as a column-vector of estimates. The terminologies "innovations" and "error terms" are used interchangeably to refer to the part not captured by models and is denoted by z_t.

The notation $F_A(y; \beta)$ stands for the cumulative distribution function of the sample y that is described by the vector of parameters β, and the subscript A provides the name of the distribution at hand. The probability density function, density for short, of a sample y is represented by $f_A(y; \beta)$ where each component of the notation is analogous to $F_A(y; \beta)$. Due to its extensive use in this paper, the gamma distribution is denoted differently: $\Gamma(\cdot)$. The gamma distribution takes in as argument a shape parameter and the rate parameter is set to 1 unless otherwise mentioned. Moreover, the gamma function is denoted by $\gamma(\cdot)$.

Lastly, standard Gaussian, standard normal, and $\mathcal{N}(0, 1)$ all describe the same distribution. The same logic applies to standard EPD and EPD(0,1). Note that this notation is valid from now on unless otherwise stated.

1 Introduction

Describing financial phenomena is a non trivial task and it has been investigated from various perspectives in literature over the years. Academics and practitioners have developed an innumerable amount of theories and models in order to explain the underlying dynamics of financial assets and macroeconomic variables as accurately as possible. Time series analysis has been extremely practical and it has allowed to shed light on particular features that some data possess. Indeed, Engel [8] was confronted to disturbance variances that were less stable than usually encountered while he was working on time-series models. His findings supported the occurrence of forecast errors that seemed to appear in clusters. This was the evidence of the presence of heteroscedasticity. Greene [10] states that heteroscedasticity is characterised by evident dependency between the variance of the forecast error and the magnitude of the previous disturbance. Engel [8] proposed the autoregressive conditionally heteroscedastic (ARCH) model in order to take this new attribute displayed by certain data set into account. The inception of ARCH models paved the way for further research on conditional volatility forecasting. Bollerslev [3] introduced the generalised autoregressive conditionally heteroscedastic (GARCH) models that enable to model a more general dynamic of volatility. There is empirical evidence that GARCH model is parsimonious and its simplest form, GARCH(1,1), performs as good as ARCH models with multiple lags. This last characteristic of GARCH(1,1) is the reason why it tends to be preferred over ARCH models by practitioners.

However, these aforementioned versions of ARCH and GARCH models still rely on a strong assumption: innovations are standard Gaussian. While this belief eases the inference on the models, it fails to describe most of phenomena that one can use both ARCH and GARCH models for, especially for modelling conditonal volatility of financial asset returns. The reason lies in the distribution of data modelled. Observing mesokurtic distributions (distributions whose kurtosis is similar to the kurtosis of the normal distribution) is rare. Platykurtic and leptokurtic distributions, smaller kurtosis than the kurtosis of the normal distribution and greater kurtosis respectively, arise more frequently. Academics have attempted to use leptokurtic distributions to model volatility, *e.g.* Bollersev [2] replaced Gaussian error terms with Student-t innovations and they have developed various extensions of GARCH models, *e.g.* exponential GARCH or EGARCH, quadratic GARCH or QGARCH to improve volatility forecasting techniques. All these have lead to promising results.

Kurtosis plays a central role in modelling. This figure is related to the tails of distributions, and subsequently the probability of extreme events to occur. Forecasting volatility of asset returns is an ability sought by industry professionals for at least two reasons: risk management purpose and speculative intentions. Zhu and Galbraith [23] compute the value-at-risk and the expected shortfall using flexible distributions, while the theoretical background nor the finite sample properties are not presented in [23].

The purpose of this paper is to explore the properties of one instance of flexible distributions utilised by Zhu and Galbraith [23] on finite samples and performing the same analysis in the Bayesian framework and is organised as follows. Section 2 gives background material on GARCH model and especially on NGARH model. Flexible distributions are introduced in Section 3. The properties of EPD and as well as its distribution is elaborated on in the section 4. The derivation of the estimates of the NGARCH is performed in two settings: frequentist and Bayesian framework. The mathematics implemented under different assumptions for the innovations, *i.e.* normality and EPD, is presented thanks to simulation studies in Section 5. An overall conclusion closes this discussion on flexible distributions and further research extensions are suggested in section 6.

2 Generalised Autoregressive Conditional Heteroscedasticity

Heteroscedasticity is a typical feature exhibited by financial asset return data. As stated above, Engel [8] came across this phenomenon in his work on time series modelling and he developed the theory of ARCH model to describe it. The functional form of this model, ARCH(p), is given by

$$r_t = \sigma_t z_t, \tag{2.1}$$

$$\sigma_t^2 = b_0 + \sum_{i=1}^{p} b_i r_{t-i}^2. \tag{2.2}$$

where z_t is assumed to be standard Gaussian, b_0 and b_i are constant parameters. As its name suggests, GARCH model is a family of models that generalise the relationship described by ARCH models. This section aims to explain GARCH model, its underlying assumptions, and one of its extensions that has been selected to perform simulations and fit real data in later sections: NGARCH model.

2.1 GARCH Model

Bollerslev [3] proposed a new form of autoregressive model to describe conditional volatility

$$r_t = \sigma_t z_t, \tag{2.3}$$

$$\sigma_t^2 = b_0 + \sum_{i=1}^{p} b_i r_{t-i}^2 + \sum_{i=1}^{q} d_i \sigma_{t-i}^2. \tag{2.4}$$

where d_i is a constant parameter. The main difference between between equations (2.2) and (2.4) is that the latter allows for lagged terms in order to describe σ_t^2. Indeed, as reported by Fryzlewicz [9], the conditional variance of r_t given the information set available up to time *t-1*, σ_t^2, possesses an autoregressive structure as well as a positive correlation with its lagged values and the squared returns r_t^2. Fryzlewicz [9] declares that the functional form of GARCH model is used to capture the idea of conditional volatility being persistent and the presence of clusters[1]. Fryzlewicz [9] covers four basic properties of GARCH model: uniqueness and stationarity, zero mean, lack of autocorrelation, and unconditional variance. The argument proving the first property involves some technicality that is beyond the scope of this dissertation. Nevertheless, Fryzlewicz [9] mentions that uniqueness and stationarity of GARCH model are ensured when $\sum_{i=1}^{p} b_i + \sum_{i=1}^{q} d_i < 1$. Fryzlewicz [9] proves the zero mean and the lack of autocorrelation by way of an information set. Let us define an information set $\mathcal{F}_t$ such that $\mathcal{F}_t = \sigma\{z_i, -\infty < i \leq t\}$. Fryzlewicz [9] proves that the mean of r_t is zero as follows

Proof.

$$\begin{aligned}
\mathbb{E}[r_t] &= \mathbb{E}[\sigma_t z_t] \quad \text{[Definition]} \\
&= \mathbb{E}\left[\mathbb{E}\left[\sigma_t z_t | \mathcal{F}_{t-1}\right]\right] \quad \text{[Law of iterated expectation]} \\
&= \mathbb{E}\left[\sigma_t \mathbb{E}\left[z_t | \mathcal{F}_{t-1}\right]\right] \quad [\sigma_t \text{ known given } \mathcal{F}_{t-1}] \\
&= \mathbb{E}\left[\sigma_t \mathbb{E}[z_t]\right] = \mathbb{E}[\sigma_t 0] = 0.
\end{aligned}$$

□

Fryzlewicz [9] claims the absence of autocorrelation in the same way

Proof.

$$\begin{aligned}
\mathbb{E}[r_t r_{t+h}] =& \mathbb{E}[r_t \sigma_{t+h} z_{t+h}] \quad \text{[Definition]} \\
=& \mathbb{E}\left[r_t \sigma_{t+h} z_{t+h} | \mathcal{F}_{t+h-1}\right] \quad \text{[Law of iterated expectation]} \\
=& \mathbb{E}\left[r_t \sigma_{t+h} \mathbb{E}\left[z_{t+h} | \mathcal{F}_{t+h-1}\right]\right] = \mathbb{E}[r_t \sigma_{t+h} 0] = 0
\end{aligned}$$

□

The reader is kindly invited to refer to [9] for the steps of the reasoning that proves the unconditional variance feature of GARCH model. Nonetheless, the result is given here as Fryzlewicz [9] describes it: $\mathbb{E}\left[r_t^2\right] = \frac{b_0}{1-\left(\sum_{i=1}^{p} b_i + \sum_{i=1}^{q} d_i\right)}$. The unconditional variance plays a critical role in the optimisation process as it shall be seen in section 4.

[1] clusters = phenomenon where large (small) values in a time series are followed by other large (small) values.

2.2 NGARCH model

Zhu and Galbraith [23] fit an NGARCH model to different financial asset returns. Zhu and Galbraith [23] report the functional form of an NGARCH as follows

$$
\begin{aligned}
r_t &= \sigma_t z_t + m, && (2.5)\\
\sigma_t^2 &= b_0 + b_1\sigma_t^2 + b_2\sigma_t^2\,(z_{t-1} - c)^2 && (2.6)\\
&= b_0 + b_1\sigma_t^2 + b_2(r_{t-1} - m - c\sigma_{t-1})^2. && (2.7)
\end{aligned}
$$

The form of the conditional variance allows for an extra parameter c. It has been empirically demonstrated that negative news has more repercussions on volatility compared to positive news of the same magnitude as discussed by Longmore and Robinson [15]. This phenomenon is called the leverage effect and appears in (2.6) and (2.7) via c. This can be mathematically interpreted by a negative correlation between the conditional volatility of the returns and the error terms in the return dynamics as developed by Zhu and Galbraith [23].

3 Flexible distributions

A multitude of pieces of research have supported the use of distributions with fat tails to model conditional asset returns. Zhu and Galbraith [22] make reference to applications of Student-t, asymmetric Weibull, and other parametric distributions to fulfil this task. Hansen [11] pioneered the usage of a skewed version of Student-t distribution to describe conditional distribution of financial product returns. Subsequently, other skew extensions of Student-t have been proposed in literature for financial purposes. However, the estimation of a skewness parameter had never been suggested before Zhu and Galbraith [22]. That parameter enables one to control the asymmetry of the central part of the distribution and boost the performance of models that fit empirical data with abnormal tail shapes. As it shall be shown in this section, any density can be extended to a generalised asymmetric version as long as a set of certain conditions holds. This can be achieved through different means and the scheme implemented in this paper is introduced.

The distribution to be used for z_t is presented through its respective density, mean, variance, cumulative distribution, and quantile function in the following subsections. Moreover, techniques to sample random numbers from this distribution are presented as well. While the theory covered in this section considers the general case, the focus will be narrowed down to a particular member of the family of distributions for the simulation study in later sections.

3.1 Double-Two Piece distribution

Rubio and Steel [18] introduce the family of univariate Double Two-Piece (DTP) distributions. They state that these distributions can be obtained through a transformation of the density of any unimodal symmetric continuous distribution. The new density resulting from this transformation will possess some parameters describing solely the left hand-side of the density and other parameters for the right part as well. Rubio and Steel [18] establish a distinction between 3 types of parameters: location, scale and shape parameters. They classify any parameter that is neither a location nor a scale parameter as a shape parameter.

For any two densities $f(x;\mu,\sigma_1,\delta_1)$ and $f(x;\mu,\sigma_2,\delta_2)$, where $\mu,\sigma_i,$ and δ_i (with $\mathbf{i}=1,2$) are the location, the scale and the shape parameters for each density, the corresponding DTP density is given by

$$f_{DTP}(x;\mu,\sigma_1,\sigma_2,\delta_1,\delta_2)=\frac{2\epsilon}{\sigma_1}f\left(\frac{x-\mu}{\sigma_1};\delta_1\right)I_{\{x<\mu\}}+\frac{2(1-\epsilon)}{\sigma_2}f\left(\frac{x-\mu}{\sigma_2};\delta_2\right)I_{\{x\geq\mu\}}. \quad (3.1)$$

In order to ensure the continuity of equation (3.1), the following must hold:

$$\epsilon=\frac{\sigma_1 f(0;\delta_2)}{\sigma_1 f(0;\delta_2)+\sigma_2 f(0;\delta_1)}. \quad (3.2)$$

Furthermore, Rubio and Steel [18] describe the cumulative distribution function of the DTP family as well as a manner of deriving its quantile function. The DTP distribution is given by

$$F_{DTP}(x;\mu,\sigma_1,\sigma_2,\delta_1,\delta_2)=2\epsilon F\left(\tfrac{x-\mu}{\sigma_1};\delta_1\right)I_{x<\mu}+\left\{\epsilon+(1-\epsilon)\left[2F\left(\tfrac{x-\mu}{\sigma_2};\delta_2\right)-1\right]\right\}I_{x\geq\mu}. \quad (3.3)$$

One can derive the quantile function by inverting equation (3.3). DTP family does not alter the continuity of the initial densities and the density obtained after applying the transformation preserves the unimodal feature at μ. Not only does ϵ guarantees the continuity of $f_{DTP}(x;\mu,\sigma_1,\sigma_2,\delta_1,\delta_2)$, but it does also allocate the quantity of mass to the left of its mode. Moreover, one can control the mass on either side of the mode by changing the ratio σ_1/σ_2. DTP family is flexible since it allows one to control the tails of the density independently from each other. In other words, the scale and the shape parameters can differ on either side of the location parameter.

The flexibility of DTP family is suitable for modelling conditional volatility of financial instrument returns. The greater control of the shape of the tails in the case of DTP densities allows one to take the asymmetry of the impact of the news into consideration. This feature makes DTP family appropriate for modelling the conditional variance as expressed in (2.6) and (2.7). Next, we describe the Asymmetric Exponential Power distribution, which represents a particular case of the family of DTP distributions.

3.2 Asymmetric Exponential Power distribution

3.2.1 Density function

Zhu and Galbraith [23] describe the form of the standard asymmetric exponential power distribution (AEPD) by the expression below:

$$f_{AEP}(y;\beta) = \begin{cases} \left(\frac{\alpha}{\alpha^*}\right) K_{AEPD}\left(p_1\right) \exp\left(-\frac{1}{p_1}\left|\frac{y}{2\alpha^*}\right|^{p_1}\right), & \text{if } y \leq 0; \\ \left(\frac{1-\alpha}{1-\alpha^*}\right) K_{AEPD}\left(p_2\right) \exp\left(-\frac{1}{p_2}\left|\frac{y}{2(1-\alpha^*)}\right|^{p_2}\right), & \text{if } y > 0; \end{cases} \tag{3.4}$$

where the parameter vector $\beta = (\alpha, p_1, p_2)$ is composed of three shape parameters, $\alpha \in (0,1)$ corresponding to the skewness parameter, $p_1 > 0$ and $p_2 > 0$ control the form of the left and the right-hand side of the density respectively, $K_{AEP}(p) \equiv 1/\left[2p^{1/p}\Gamma\left(1+1/p\right)\right]$ represent the normalising constant of the GED distribution[2], and α^* is defined as $\alpha^* = \alpha K_{AEP}\left(p_1\right)/\left[\alpha K_{AEP}\left(p_1\right) + (1-\alpha) K_{AEP}\left(p_2\right)\right]$. The AEPD density function does have desirable properties. As a matter of fact, it is continuous over its domain as well as unimodal with mode at $y=0$. However, it is not differentiable at its mode. The parameter α^* ensures continuity under changes to the parameter vector β by providing scale adjustments to both sides of the mode.

There exists a general expression of the AEPD density and it is specified by $\frac{1}{\sigma} f_{AEP}\left(\frac{y-\mu}{\sigma};\beta\right)$, where μ is the location (mode) and σ represent the scale parameter. Note that when $\mu = 0$ and $\sigma = 1$, the general expression of the AEPD density becomes the standard AEPD density. Zhu and Galbraith [23] argue that the mode ($y=0$) of the AEPD is habitually not equal to its mean, $E\left[Y_{AEP}\right]$, and that expression is given by:

$$\omega_{AEP}\left(\beta\right) \equiv E\left[Y_{AEP}\right] = \frac{1}{B_{AEP}}\left[(1-\alpha)^2 \frac{p_2\Gamma\left(2/p_2\right)}{\Gamma^2\left(1/p_2\right)} - \alpha^2 \frac{p_1\Gamma\left(2/p_1\right)}{\Gamma^2\left(1/p_1\right)}\right], \tag{3.5}$$

where B_{AEP} is a constant (insert use) and is given by

$$B_{AEP} \equiv \frac{\alpha}{\alpha^*} K_{AEPD}\left(p_1\right) = \frac{1-\alpha}{1-\alpha^*} K_{AEPD}\left(p_2\right) = \alpha K_{AEPD}\left(p_1\right) + (1-\alpha) K_{AEPD}\left(p_2\right). \tag{3.6}$$

Its variance $\left[\delta\left(\beta\right)\right]^2 \equiv Var\left[Y_{AEP}\right]$ is defined as

$$\left[\delta_{AEP}\left(\beta\right)\right]^2 = \frac{1}{B_{AEP}^2}\left[(1-\alpha)^3 \frac{p_2^2\Gamma\left(2/p_2\right)}{\Gamma^3\left(1/p_2\right)} - \alpha^3 \frac{p_1^2\Gamma\left(2/p_1\right)}{\Gamma^3\left(1/p_1\right)}\right] - \left[\omega\left(\beta\right)\right]^2. \tag{3.7}$$

Once one relaxes the assumption of normality for the distributions of z_t in (2.6) along with (2.7) and assumes standard AEPD, their density becomes $f\left(z;\beta\right) = \delta\left(\beta\right) f_{AEP}\left(\omega\left(\beta\right) + \delta\left(\beta\right) z;\beta\right)$.

[2]The generalised error distribution is another name of the exponential power distribution.

3.2.2 Cumulative distribution and Quantile function

Assume that Y has a standard AEPD density. Furthermore, the cumulative distribution of the gamma distribution, $F_{\Gamma}(y;\zeta,\eta)$, is reported by Casella and Berger [5] as follows:

$$F_{\Gamma}(y;\zeta) = \frac{\gamma(\zeta,y)}{\Gamma(\zeta)} = \int_0^y \frac{\eta^{\zeta}}{\Gamma(\zeta)} x^{\zeta-1} exp(-y)\ dx, \tag{3.8}$$

where $\gamma(\zeta,y)$ is the lower incomplete gamma function[3]. The inverse cumulative distribution is given by $F_{\Gamma}^{-1}(y;\zeta)$.
Zhu and Zinde-Walsh [24] introduce the cumulative distribution of the AEPD(0,1) by the means of the expression

$$F_{AEP}(y;\beta) = \begin{cases} \alpha\left[1 - F_{\Gamma}\left(\frac{1}{p_1}\left(\frac{|y|}{2\alpha^*}\right);\frac{1}{p_1}\right)^{p_1}\right], & \text{if } y \leq 0; \\ \alpha + (1-\alpha)\,F_{\Gamma}\left(\frac{1}{p_2}\left(\frac{|y|}{2(1-\alpha^*)}\right)^{p_2};\frac{1}{p_2}\right), & \text{if } y > 0; \end{cases} \tag{3.9}$$

while its quantile function is given by

$$F_{AEP}^{-1}(y;\beta) = \begin{cases} -2\alpha^*\left[p_1 F_{\Gamma}^{-1}\left(1 - \frac{U}{\alpha};\frac{1}{p1}\right)\right]^{1/p_1}, & \text{if } U \leq \alpha; \\ -2(1-\alpha^*)\left[p_2 F_{\Gamma}^{-1}\left(1 - \frac{1-U}{1-\alpha};\frac{1}{p2}\right)\right]^{1/p_2} & \text{if } U < \alpha. \end{cases} \tag{3.10}$$

3.2.3 Sampling from standard AEPD

One can make use of various schemes to generate random variables from a specific distribution. The inversion principle described by Devroye [7] is certainly the first technique one would think of in order to complete this task and Komunjer [13] supports this claim for asymmetric power distributions. However, Li [14] discusses another method for drawing random numbers from the standard AEPD using its stochastic representation. This modus operandi has been applied in simulation studies and inference on the NGARCH model with $z_t \sim$ EPD(0,1) presented in section 4 and section 5.
Li [14] claims that one is able to generate random numbers from a standard AEPD by the means of 3 other random variables:

$$U \sim U(0,1)\ [\text{Uniform }(0,1)]$$

$$X_1 \sim \Gamma\left(\frac{1}{p_1}\right)$$

$$X_2 \sim \Gamma\left(\frac{1}{p_2}\right)$$

[3] $\gamma(\zeta,y) = \int_0^y x^{\zeta-1} exp(-x)\ dx$

Li [14] claims that one can generate random number from $W \sim$ AEPD$(0,1)$ using the following equation:

$$W = \mu + \frac{\alpha\sigma[\text{sign}\,(U-\alpha)-1]}{2B_{AEP}\gamma\left(1+1/p_1\right)}Y_1^{1/p_1} + \frac{(1-\alpha)\,\sigma[\text{sign}\,(U-\alpha)+1]}{2B_{AEP}\gamma\left(1+1/p_2\right)}Y_2^{1/p_2}. \tag{3.11}$$

The reader is referred to [14] to gain some insights on the arguments made by Li justifying this particular stochastic representation.

4 Exponential Power Distribution

The EPD is a member of the AEPD family and corresponds to the AEPD with $p_1 = p_2 = p$ and $\alpha = 0.5$. This is the distribution that has been chosen for simulation case study presented in this paper.

Different standard EPD densities are shown in Figure 1 and are compared to the standard Gaussian. A particular attention is given to the left tails of each density and their rate of decay by displaying the left ends in Figure 2, Figure 3 and Figure 4 respectively.

Figure 1: Platykurtic EPD density

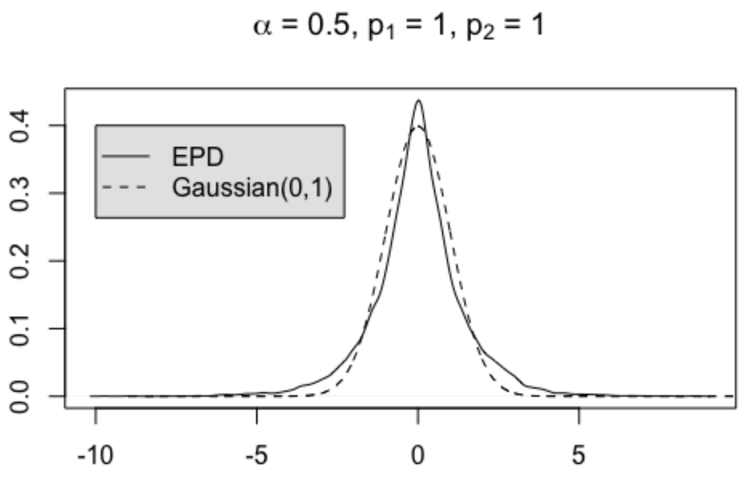


Figure 2: Platykurtic EPD density - Focus on left tail

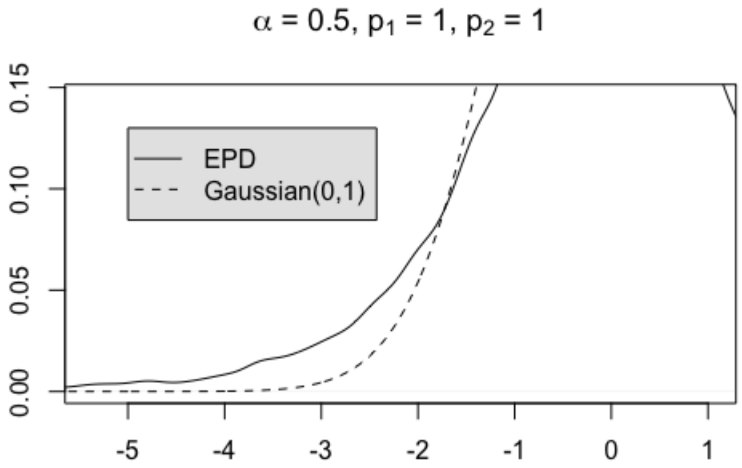


Figure 3: Mesokurtic EPD density

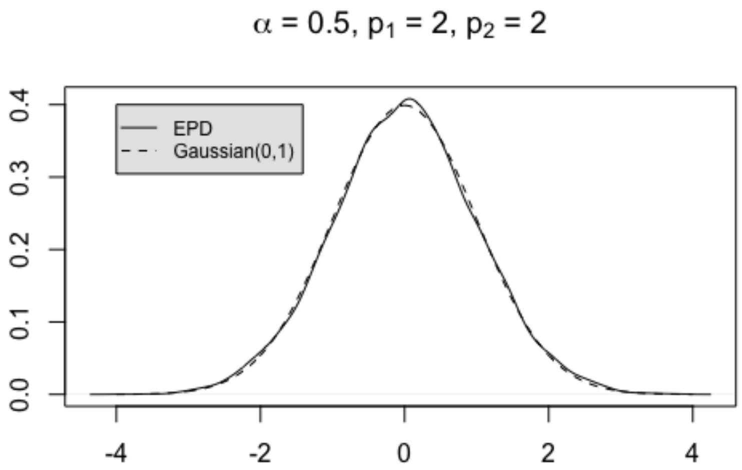


Figure 4: Mesokurtic EPD - Focus on left tail

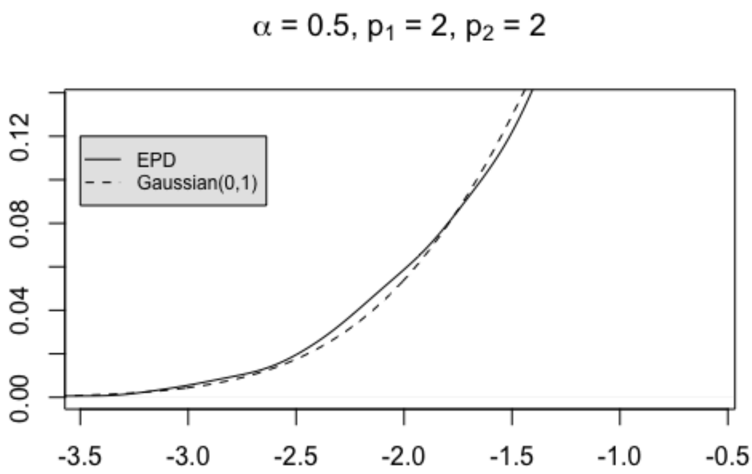


Figure 5: Leptokurtic EPD density

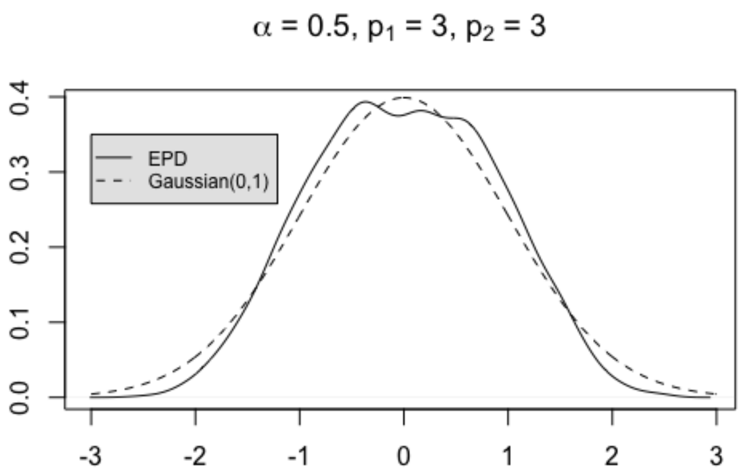
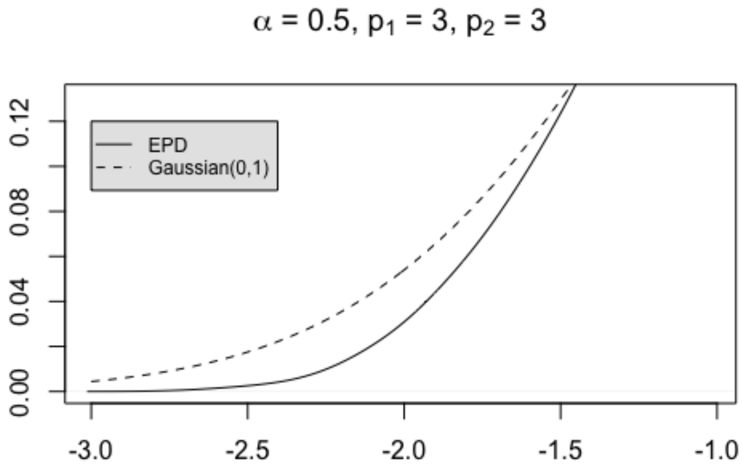


Figure 6: Leptokurtic EPD - Focus on left tail



It is worth noting that for EPD, the following components have an simpler form than the AEPD case:

-$\alpha^* = \alpha K_{AEP}(p) / \left[\alpha K_{AEP}(p) + (1-\alpha) K_{AEP}(p)\right] = \alpha$

-$B_{EP} \equiv \frac{\alpha}{\alpha^*} K_{AEP}(p) = K_{AEP}(p)$

- $\omega_{EP}(\beta) = \frac{1}{B_{AEP}} \left[(1-\alpha)^2 \frac{p\Gamma\left(2/p\right)}{\Gamma^2\left(1/p\right)} - \alpha^2 \frac{p\Gamma\left(2/p\right)}{\Gamma^2\left(1/p\right)}\right] = \frac{1}{B_{EP}} \left[(1-2\alpha) \frac{p\Gamma\left(2/p\right)}{\Gamma^2\left(1/p\right)}\right]$

- $\delta_{EP}(\beta) = \frac{1}{B_{AEP}^2} \left[(1-\alpha)^3 \frac{p^2\Gamma\left(2/p\right)}{\Gamma^3\left(1/p\right)} - \alpha^3 \frac{p^2\Gamma\left(2/p\right)}{\Gamma^3\left(1/p\right)}\right] - \left[\mathbb{E}\left[Y_{EP}\right]\right]^2 = \frac{1}{B_{EP}} \left[\left(1-\alpha\left(3-3\alpha+2\alpha^2\right)\right) \frac{p^2\Gamma\left(2/p\right)}{\Gamma^3\left(1/p\right)}\right] - \left[\mathbb{E}\left[Y_{EP}\right]\right]^2$

5 Simulations and model fitting

There are a couple of techniques available in order to find estimates. Casella and Berger [5] cover four of these methods: the method of moments, the Expectation-Maximisation algorithm, the method of maximum likelihood (MML), and Bayes Estimates. The former procedures are not developed in this paper and the reader is kindly referred to Casella and Berger [see 5, pages 312-313 and page 326]. The MML is described in the subsequent subsection while the remaining method, Bayes Estimates, will be the central object of the second subsection. The performance of each model is presented and assessed through several metrics.

5.1 Method of maximum likelihood

The MML consists of deriving the estimates that maximise the probability of observing x sampled from a certain density or mass function. Let $X_1, ..., X_n$ be an i.i.d. sample from any distribution $f\left(x|\boldsymbol{\theta}\right)$. One defines the likelihood function by

$$L\left(\boldsymbol{\theta}|\mathbf{x}\right) = \Pi_{i=1}^{n} f\left(x_i|\theta_1, ..., \theta_k\right) \tag{5.1}$$

It follows that the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}$ is then $\hat{\boldsymbol{\theta}} = argmax_{\boldsymbol{\theta}} L\left(\boldsymbol{\theta}|\mathbf{x}\right)$. Note that in practice the optimisation is executed on the loglikelihood instead and this does not alter the results because monotonic transformations conserve the order of the original set as mentioned by Toh [21]. Not only are the properties of MLE sought-after, namely unbiasedness, consistent ($\hat{\theta} \sim \theta$ in probability), efficient (standard error tends to zero as n goes to ∞), and asymptotically standard Gaussian, but computing it is rather simple and that procedure works for non-linear functions as well. As a consequence of its optimisation nature, the MLE problem can be solved in all packages and software that possess an optimiser routine[4].

One common modus operandi utilised when one desires to compute MLE is to proceed to the reparametrisation of the estimates. Rubio [17] provides a formal definition of this concept:

Definition 5.1 (Reparametrisation)**.** Let $f\left(\mathbf{x};\boldsymbol{\theta}\right)$ be a pdf with parameters $\boldsymbol{\theta} = \left(\theta_1, ..., \theta_p\right)'$ $\in \Theta \subset \mathbb{R}^p,\ \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^n$. A reparameterisation $\boldsymbol{\eta} = \phi\left(\boldsymbol{\theta}\right)$ is the change of variables $\theta_j \mapsto \eta_j,\ j = 1, ..., p,$
via a one-to-one function ϕ such that, for each $\theta \in \Theta,$ there exists $\boldsymbol{\eta} \in \phi\left(\Theta\right)$ such that $f\left(x;\phi\left(\boldsymbol{\theta}\right)\right) = f\left(x;\phi^{-1}\left(\boldsymbol{\eta}\right)\right)$. Analogously, for each $\eta \in \phi\left(\Theta\right)$ there exists $\boldsymbol{\theta} \in \Theta$ such that $f\left(x;\boldsymbol{\eta}\right) = f\left(x;\phi\left(\boldsymbol{\theta}\right)\right)$.

This manoeuvre is popular in statistics because it prevents singularities in the domain of the likelihood function and ensures that the likelihood surface is smooth[5]. Furthermore, the choice of parametrisation does not affect the computation of the MLE. This is due to the invariance property of MLE which is stated in the next lines. Casella and Berger enunciate this theorem as follows:

Theorem 5.1 (Invariance property of MLE)**.** If $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$, then for any function $\tau\left(\boldsymbol{\theta}\right)$, the MLE of $\tau\left(\boldsymbol{\theta}\right)$, is $\tau\left(\hat{\boldsymbol{\theta}}\right)$.

Note that no further requirement on the function τ is needed. Although one-to-one functions have been chosen to implement the reparametrisation technique in this paper,

[4]The results presented in this paper were derived thanks to the library "ucminf" in RStudio Version 1.2.1335

[5]Surjective likelihood function, *i.e.* every point in the domain is mapped onto the likelihood surface.

nothing prevents one to opt for non-bijective functions instead. Nevertheless, this would require an extra piece of machinery, *i.e.* the notion of induced likelihood which is touched upon by Scott and Nowak [19]. Theorem (5.1) shall not be proven because it involves this very concept, however the proof is provided by Casella and Berger[5].
The reparametrisations that have been used to solve the MLE problem in this paper are listed below:

1. Any θ_j which has a strict positivity constraint imposed on have been exponentiated, *i.e.* $\theta_j \mapsto \exp\left(\theta_j\right)$.
2. For all $\theta_j \in (0,1)$, the sigmoid function has been applied, *i.e.* $\theta_j \mapsto S\left(\theta_j\right)$.

Although MLE is not impacted by any parametrisation, its variance is. Ergo, an approximation of this quantity is necessary since exact calculations are relatively laborious. Campbell et al. [4] assert that the Delta Method relies on Taylor expansion and allows one to perform inferences in that more complicated settings.
Casella and Berger present this result:

Theorem 5.2 (Delta Method)**.** Assume that X_n satisfy $\sqrt{n}\left(X_n - \theta\right) \sim \mathcal{N}\left(0, \sigma^2\right)$ in distribution. For a given function τ and a specific value of θ, suppose that $\tau'\left(\theta\right)$ exists and is not 0. Then

$$\sqrt{n}\left[\tau\left(X_n\right) - \tau\left(\theta\right)\right] \sim \mathcal{N}\left(0, \sigma^2\left[\tau'\left(\theta\right)\right]^2\right) \text{ in distribution.}$$

Before proving Theorem 2, another theorem shall be introduced. It provides for addition and multiplication rules for random variables converging to either a distribution or a constant as reported by Casella and Berger [5]:

Theorem 5.3 (Slutsky's Theorem)**.** If $X_n \sim X$ in distribution and $X_n \sim a$, then

1. $Y_n X_n \sim aX$ in distribution.
2. $X_n + Y_n \sim X + a$ in distribution.

Given that the proof of Slutsky's theorem is beyond the scope of this paper, it shall not be proved. The proof of Theorem 2 can be provided by the means of Slutsky's theorem and is proceeded to in the upcoming lines.

Proof.

-The Taylor expansion of $\tau\left(Y_n\right)$ about the θ is given by

$$\tau\left(X_n\right) = \tau\left(\theta\right) + \tau'\left(\theta\right)\left(X_n - \theta\right) + \sum_{n=2}^{\infty} \frac{\tau^{(n)}\left(\theta\right)}{n!}\left(X_n - \theta\right)^n \tag{5.2}$$

- One can invoke the Central limit theorem to justify that $\sqrt{n}\left(X_n-\theta\right) \sim \mathcal{N}\left(0,\sigma^2\right)$ in distribution as n $\to \infty$.

- This implies that $X_n - \theta \sim 0$, and by extension $X_n \sim \theta$.

- Using this last fact, one can demonstrate that the last term in (5.2), the remainder from the approximation, converges to 0 in probability [5].

- Substituting the approximation of $\tau\left(X_n\right)$ in $\sqrt{n}\left[\tau\left(X_n\right)-\tau\left(\theta\right)\right]$ leads to

$$\sqrt{n}\left[\tau\left(X_n\right)-\tau\left(\theta\right)\right] = \tau'\left(\theta\right)\sqrt{n}\left(X_n-\theta\right) \sim \tau'\left(\theta\right)\sqrt{n}\,\mathcal{N}\left(0,\sigma^2\right) \text{ in distribution.} \quad (5.3)$$

where Slutsky's theorem has been employed in the right hand-side of (5.3).

- Note that the last term in (5.3) is equivalent to $\sqrt{n}\,\mathcal{N}\left(0,\sigma^2\left[\tau'\left(\theta\right)\right]^2\right)$ in distribution. $\square$

The two loglikelihood functions used in this paper are given by

$$l_{Gaussian} = -\frac{T}{2}log\ \left(\sigma_t^2\right) - \frac{1}{\sigma_t^2}\sum_{t=1}^{T}\left(r_t-m\right)^2 \quad (5.4)$$

and

$$l_{EPD} = \sum_{t=1}^{T} log\ \delta_{EPD}\left(\beta\right) - log\ \sigma_t + log\ f_{EPD}\left(\omega_{EPD}\left(\beta\right)+\delta_{EPD}\frac{r_t-m}{\sigma_t};\beta\right) \quad (5.5)$$

as reported by Zhu and Galbraith [23]. The theory presented represents the asymptotic behaviour of MLE. However, finite sample properties are studied through Monte Carlo simulations in the next subsection.

5.2 Simulation Study: Maximum Likelihood Estimation

The finite sample properties of the MLEs are now explored by the use of a Monte Carlo simulation. For each assumption of the error distribution, MLE for equation (2.5) has been computed. Tables 1 and 2 report the output of the fitting procedure where the innovations follow a standard normal and a standard EPD respectively. Note that there are 200 time points in each scenario and two sample sizes have been chosen for the total of simulations: 100 and 200. The real parameters to be estimated are $\boldsymbol{\theta}_1 = (b_0 = 1, b_1 = 0.4, b_2 = 0.3, c = 0.2, m = 0.7)$ for the standard normal case and $\boldsymbol{\theta}_2 = (b_0 = 1, b_1 = 0.5, b_2 = 0.3, c = 0.1, \alpha = 0.5, p = 1.75, m = 0.7)$ for the platykurtic EPD(0,1) innovations and $\boldsymbol{\theta}_3 = (b_0 = 1, b_1 = 0.5, b_2 = 0.3, c = 0.1, \alpha = 0.5, p = 3, m = 0.7)$ for the leptokurtic EPD(0,1) case. Zhu and Galbraith [23] argue that since returns series is modelled as a stationary EPD(0,1) process, its unconditional mean $\mathbb{E}\left(r_t\right)$ and unconditional variance $var\left(r_t\right)$ should be set equal to m and $\frac{b_0}{1-b_1-b_2(1+c)^2}$ respectively. Furthermore, $var\left(r_t\right)$ has been chosen as the starting point for the volatility process as Zhu and Galbraith [23] recommend. This way, one can purely focus on the effect of different distribution specifications of the error terms as Zhu and Galbraith mention [23]. Therefore, the quantities reported in the tables below correspond to the arithmetic mean of MLEs, the medians, the mean standard errors (MSE), the standard errors (SE), and the length of the confidence interval (length C.I.). The latter is calculated as the difference between the upper bounds and the lower bounds of each scenario. The bounds are given according to the following formula:

$$LB_i = MLE_i - z_{\alpha/2} * SE_i \qquad \text{[Lower bound]}$$
$$UB_i = MLE_i + z_{\alpha/2} * SE_i \qquad \text{[Upper bound]}$$

where i is the i-th simulation and $z_{\alpha/2}$ is the z-score for 100*(1-α)% confidence level. Following common practice, a 95% confidence level has been selected.

Table 1: Inference on NGARCH model with standard Gaussian innovations (sample size = 100)

	Mean	Median	SE	MSE	Length C.I.
$\hat{b_0}$	1.0801	1.0283	0.1041	0.2684	1.0522
$\hat{b_1}$	0.4792	0.4838	0.0089	0.0892	0.3498
$\hat{b_2}$	0.2973	0.2973	0.0031	0.0601	0.2356
$\hat{c}$	0.1000	0.0961	0.0060	0.0851	0.3335
$\hat{m}$	0.6989	0.6929	0.0047	0.0702	0.2750

Table 2: Inference on NGARCH model with standard Gaussian innovations (sample size =200)

	Mean	Median	SE	MSE	Length C.I.
$\hat{b_0}$	1.0461	1.0108	0.0806	0.2741	1.0744
$\hat{b_1}$	0.4851	0.4838	0.0077	0.0901	0.3534
$\hat{b_2}$	0.2992	0.2965	0.0032	0.0592	0.2321
$\hat{c}$	0.1007	0.0864	0.0085	0.0814	0.3192
$\hat{m}$	0.7027	0.6982	0.0051	0.0692	0.2714

Table 3: Inference on NGARCH model with platykurtic standard EPD innovations (sample size = 100)

	Mean	Median	SE	MSE	Length C.I.
$\hat{b}_0$	1.1404	1.0982	0.1339	0.3103	1.2163
$\hat{b}_1$	0.4958	0.4986	0.0086	0.0955	0.3743
$\hat{b}_2$	0.3267	0.3211	0.0051	0.0748	0.2933
$\hat{c}$	0.0877	0.0647	0.0095	0.0631	0.2474
$\hat{\alpha}$	0.5011	0.5010	0.0007	0.0323	0.1267
$\hat{p}$	1.8336	1.7785	0.3427	0.1820	0.7134
$\hat{m}$	0.7003	0.6990	0.0074	0.1080	0.4234

Table 4: Inference on NGARCH model with platykurtic standard EPD innovations (sample size = 200)

	Mean	Median	SE	MSE	Length C.I.
$\hat{b}_0$	1.1247	1.0975	0.1274	0.3269	1.2816
$\hat{b}_1$	0.4996	0.5020	0.0088	0.0953	0.3735
$\hat{b}_2$	0.3242	0.3225	0.0052	0.0781	0.3061
$\hat{c}$	0.0847	0.0523	0.0094	0.0635	0.2491
$\hat{\alpha}$	0.5002	0.5005	0.0007	0.0334	0.1310
$\hat{p}$	1.8282	1.7809	0.2605	0.1720	0.6744
$\hat{m}$	0.7039	0.7048	0.0073	0.0976	0.3828

Table 5: Inference on NGARCH model with leptokurtic standard EPD innovations (sample size = 100)

	Mean	Median	SE	MSE	Length C.I.
$\hat{b}_0$	0.8182	0.8173	0.0911	0.2574	1.0090
$\hat{b}_1$	0.4866	0.4826	0.0131	0.1110	0.4349
$\hat{b}_2$	0.2297	0.2280	0.0078	0.0500	0.1958
$\hat{c}$	0.1255	0.1145	0.0100	0.0979	0.3836
$\hat{\alpha}$	0.5013	0.4975	0.0015	0.0410	0.1606
$\hat{p}$	3.0906	3.0766	0.1026	0.3549	1.3913
$\hat{m}$	0.6918	0.6929	0.0037	0.0676	0.2650

Table 6: Inference on NGARCH model with leptokurtic standard EPD innovations (sample size = 200)

	Mean	Median	SE	MSE	Length C.I.
$\hat{b}_0$	0.8180	0.7768	0.0991	0.2704	1.0598
$\hat{b}_1$	0.4856	0.4924	0.0136	0.1214	0.4759
$\hat{b}_2$	0.2297	0.2280	0.0073	0.0521	0.2044
$\hat{c}$	0.1272	0.1145	0.0103	0.1062	0.4165
$\hat{\alpha}$	0.4981	0.4963	0.0013	0.0403	0.1580
$\hat{p}$	3.0996	3.0917	0.1074	0.3508	1.3752
$\hat{m}$	0.6943	0.6932	0.0038	0.0683	0.2678

Some comments can be made about bias reduction, the relation between the mean, the median, and the variability of the estimates. The general observation is that the bias as well as SE tend to reduce as the sample size increases. As a consequence, the length of C.I. shrinks. Discrepancies in the results are due to outliers in the simulation process[6]. The means and the medians of estimates are relatively close to each other, although the medians are slightly smaller. This is an evidence of a slight negative skew.

5.3 Bayesian inference

Despite the coventional implementation of the MML for the inference on GARCH models, this scheme carries several disadvantages that could potentially obstruct the derivation of estimates. As Ardia [1] declares, parameters are the results of a constrained optimisation in the MML. While reparametrisation is a subterfuge to partially circumvent constraints, the optimisation procedure itself could remain complex. Firstly, convergence may not be reached if the true parameter values are on the verge of the parameter space and the covariance matrix at the maximising arguments, which is used in the derivation of standard errors, can be cumbersome to compute. Secondly, the initial guesses as well as the optimisation routines can have a non-negligible impact on the estimate output.

Bayes estimates bring a solution to the issues raised above as it shall be seen in this subsection.

In the Bayesian framework, parameters are not seen as constant quantities but they are considered as random variables themselves. Their distribution, before data are collected, is called the prior distribution, or prior for short, and is denoted by $\pi(\boldsymbol{\theta})$. They represent the preliminary knowledge that one has on the parameters. Note that this includes restrictions on parameters, *e.g.* positivity or more complex inequalities. In contrast to the frequentist approach, one is interested in investigating the probability of observing estimates given

[6]Histograms are provided in the source codes to support that argument and detect these outliers.

some data, the posterior distribution or posterior (hereafter $\pi\left(\boldsymbol{\theta}|y_1,...,y_2\right)$).
The central piece of Bayesian theory lies in the Bayes' formula. Dekking et al. [6] state it as follows

Theorem 5.4 (Bayes Formula)**.** Suppose the events $C_1, C_2, ..., C_m$ are disjoint and $C_1 \cap C_2 \cap ... \cap C_m = \Omega$, the sample space. The conditional probability of C_i, given an arbitrary event A, can be expressed as:

$$P\left(C_i|A\right) = \frac{P\left(A|C_i\right)P\left(C_i\right)}{P\left(A|C_1\right)P\left(C_1\right) + P\left(A|C_2\right)P\left(C_2\right) + ... + P\left(A|C_m\right)P\left(C_m\right)} \tag{5.6}$$

Dekking et al. claim that theorem (5.4) is the traditional form of the Bayes' rules. The variant of (5.4) used in the Bayesian framework is

$$P\left(C|A\right) = \frac{P\left(A|C\right)P\left(C\right)}{P\left(A\right)} \tag{5.7}$$

Equation (5.6) and equation (5.7) differ in the denominator. The law of total probability has been applied in (5.7) in order to simplify $P\left(A|C_1\right)P\left(C_1\right) + P\left(A|C_2\right)P\left(C_2\right) + ... + P\left(A|C_m\right)P\left(C_m\right)$. A proof of the latter version of the Bayes' rule is provided in the subsequent lines.

Proof.

- Let A and C be two events in a sample space Ω.

- Note that $P\left(A \cap C\right) = P(A)P(A|C)$ [Multiplication rule]

- Also note that $P\left(A \cap C\right) = P(C)P(C|A)$ [Multiplication rule]

- One has $P(A)P(A|C) = P(C)P(C|A)$ and the desired result is obtained by dividing each side by $P(C)$ (provided that $P(C) > 0$). □

If one lets A be $y_1, ..., y_n$ and C be $\boldsymbol{\theta}$, equation (5.7) becomes

$$\pi\left(\boldsymbol{\theta}|y_1,...,y_n\right) = \frac{P\left(y_1,...,y_n|\boldsymbol{\theta}\right)P\left(\boldsymbol{\theta}\right)}{P\left(y_1,...,y_n\right)} \text{[Discrete random varibale]} \tag{5.8}$$

$$\pi\left(\boldsymbol{\theta}|y_1,...,y_n\right) = \frac{f\left(y_1,...,y_n|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right)}{f\left(y_1,...,y_n\right)} \text{[Continuous random varibale]} \tag{5.9}$$

$$\propto L\left(y_1,...,y_n|\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta}\right) \tag{5.10}$$

Hurlin [12] comments on equation (5.8), equation (5.9), and equation (5.10). The denominator in (5.9), $f\left(y_1,...,y_n\right) = \int_\Theta f\left(y_1,...,y_n|\boldsymbol{\theta}\right)d(\boldsymbol{\theta})$, acts as a normalising factor and ensures that that the posterior is a proper density, *i.e.* the integral of $\pi\left(\boldsymbol{\theta}|y_1,...,y_n\right)$ over the parameter space Θ yields 1. Moreover, Hurlin [12] states that Bayes' rule implies that the posterior is proportional to the likelihood and the prior (5.10).

As mentioned above, one is interested in the estimation of the posterior in Bayesian framework. According to Hurlin [12], the posterior distribution may be expressed analytically when it comes from a standard distribution. However, this rarely occurs in practice, hence one has to rely on numerical simulations to simulate from $\pi\left(\boldsymbol{\theta}|y_1,...,y_n\right)$. This is a crucial step in order to derive Bayesian estimates. Takaishi [20] claims that one can estimate $\boldsymbol{\theta}$ by deriving the expectation values of $\hat{\boldsymbol{\theta}}$ that is given by

$$\mathbb{E}\left[\boldsymbol{\theta}\right] = \int_{\Theta} \boldsymbol{\theta}\pi\left(\boldsymbol{\theta}|y_1,...,y_n\right) d\boldsymbol{\theta}. \tag{5.11}$$

Takaishi [20] argues that MCMC techniques allow one to estimate equation (5.11) numerically and describes the procedure as follows. Firstly, $\boldsymbol{\theta}$ is drawn from $\pi\left(\boldsymbol{\theta}|y_1,...,y_n\right)$. The sampling method produces a Markov chain and once data have been sampled, the expectation value are evaluated as the arithmetic mean over the sample data $\boldsymbol{\theta}^{(i)}$

$$\mathbb{E}\left[\boldsymbol{\theta}\right] = lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\theta}^{(i)}, \tag{5.12}$$

where k is the number of the sampled data [20]. Ardia [1] invokes the law of large numbers to justify that equation (5.11) is an appropriate manner to estimate the posterior expectation that can be used as an estimate of $\boldsymbol{\theta}$. Unfortunately, the data generated through this method are correlated. As a result of this, techniques to assess the performance of the inference should be adapted to factor this feature out.

The sampling method applied in this paper is the principal object of the next subsubsection.

5.3.1 Metropolis-Hastings algorithm

Various sampling methods, or samplers for short, are available to generate data from a posterior. Ardia [1] elaborates on two of the most commonly used samplers in practice: the Gibbs sampler and the the Metropolis-Hastings algorithm (MH algorithm hereafter). As stated by Ardia [1], the Gibbs sampler is a logical choice when the posterior distributions has an easy form. Nonetheless, it can quickly become toilsome in higher dimensions or when the expression of the posterior density is complex (the reader is kindly referred to [1] for further information on Gibbs sampler). In these scenarios, Ardia [1] recommends the use of the MH algorithm given that this sampler demands less knowledge on the densities involved. Ardia [1] breaks down the MH algorithm in four concise steps as reported below

- Initialisation of the counter to $j = 1$ and choice of an initial set of values, $\boldsymbol{\theta}^{[0]}$.
- Move of the chain to a set of new values $\boldsymbol{\theta}^{*}$ sampled from a proposal (or candidate) density $q\left(\cdot|\boldsymbol{\theta}^{[j-1]}\right)$.

- Evaluation of the acceptance probability of the previous move from $\boldsymbol{\theta}^{[j-1]}$ to $\boldsymbol{\theta}^{*}$ given

$$min \left\{ \frac{p\left(\boldsymbol{\theta}^{*}|y_1,...,y_n\right) q\left(\boldsymbol{\theta}^{[j-1]}|\boldsymbol{\theta}^{*}\right)}{p\left(\boldsymbol{\theta}^{[j-1]}|y_1,...,y_n\right) q\left(\boldsymbol{\theta}^{*}|\boldsymbol{\theta}^{[j-1]}\right)}, 1 \right\} \tag{5.13}$$

  The move can either be accepted, then $\boldsymbol{\theta}^{[j]}$ is set to $\boldsymbol{\theta}^{*}$ or be rejected and this will lead to $\boldsymbol{\theta}^{[j]}$ iset to $\boldsymbol{\theta}^{[j-1]}$.

- Change of the counter from j to $j+1$ and repetition of two previous steps until convergence has been reached.

As the number of iterations increases, the chain approaches its equilibrium distribution as supported by Ardia [1]. It is worth mentioning that the posterior and the proposal densities have been derived on log scale in this dissertation which does not alter the outcome of the scheme because of the monotonicity of the log transformation as by claimed by Toh [21].

5.4 Simulation Study: Bayesian Inference

As mentioned above, inference in Bayesian framework allows one to express his believes about parameters before data are collected. There are several facts about the parameters of interest known prior to the collection of data:

- The inequality $b_1 + b_2(1 + c^2) < 1$ should hold given the stationary feature of r_t as argued by Zhu and Galbraith [23].

- The parameter c usually is strictly positive for stocks [23].

- There is a strict non-negativity constraint imposed on $b_0, b_1,$ and b_2.

For the scenario where innovations are standard Gaussian, the following priors have been utilised:

- Uniform distribution $U(0, 10, 000)$ for all parameters b_i with i $= \{0, 1, 2\}$ and c.

- Normal distribution for m with a large standard deviation $[m \sim \mathcal{N}(0, 100)]$

For the standard EPD innovations case, the additional parameters to be estimated, namely α and p, have condition imposed on them

- The parameter α lies between 0 and 1.

- The parameter p has to be strictly positive.

New priors are needed for these two parameters and an uniform $U(0,1)$ and a $\log \mathcal{N}(0.01, 0.5)$ distributions have been chosen for both.

All these priors have been selected for a specific reason. For b_i with i $= \{0,1,2\}$, one can argue that their role is comparable to regression coefficients. Therefore, flat priors were appropriate for them. However, we restricted on the polygon defined by $b_1 + b_2(1+c^2) < 1$ to ensure the stationarity of the process r_t. A flat prior has been picked for c as well given than not much is known about that parameter other than it has to be positive. Regarding m, a Gaussian with a large standard deviation has been employed. Although there is little information about m too, observing extreme values for m is unlikely (given knowledge on equities), ergo assigning the equal probability to extreme values and values that arise more frequently is not pertinent. The choice of the prior for α has made given that the condition imposed on α and EPD is a symmetric distribution. As a result of these two facts, one needs a distribution that is strictly positive and ranging from 0 to 1 with a mean that equals 0.5. Lastly, p describes the shape of the density tails. This value should be fairly given that large number will lead to a uniform density. Therefore, a distribution that covers small range and allocates more mass towards the left-hand side of the density is preferable. The lognormal distribution is appropriate for this purpose.

The library "MHadaptive" has been used to proceed to the Bayesian inference. However, the optimisation method implemented by the function "Metro_Hastings", "optim", has been changed to "ucminf" to keep the consistency of the optimisation method across the frequentist and Bayesian approach. Tables 7,8,9,10,11, and 12 report the outcome of the inference on the standard normal, platykurtic EPD(0,1), and leptokurtic EPD(0,1) error terms respectively. For each simulation, the mean, the median, the SE, the MSE and the length of the Bayesian credible interval (length B.C.I.) are derived their arithmetic mean which is displayed in the subsequent tables. Reporting the same quantities in Bayesian inference as in the frequentist inference ease the comparison between the two approaches.

Table 7: Inference on NGARCH model with standard Gaussian innovations - Bayesian (sample size = 100)

	Mean	Median	MSE	SE	Length B.C.I.
$\hat{b}_0$	1.1278	1.1032	0.0913	0.0947	0.2754
$\hat{b}_1$	0.4898	0.4892	0.0019	0.0135	0.0394
$\hat{b}_2$	0.3038	0.3057	0.0014	0.0117	0.0343
$\hat{c}$	0.1198	0.1149	0.0050	0.0234	0.0649
$\hat{m}$	0.7342	0.7285	0.0130	0.0244	0.0714

Table 8: Inference on NGARCH model with standard Gaussian innovations - Bayesian (sample size = 200)

	Mean	Median	MSE	SE	Length B.C.I.
$\hat{b}_0$	1.1052	1.0821	0.0706	0.0757	0.2182
$\hat{b}_1$	0.4944	0.4940	0.0015	0.0112	0.0324
$\hat{b}_2$	0.3017	0.3036	0.0011	0.0104	0.0304
$\hat{c}$	0.1182	0.1118	0.0049	0.0210	0.0598
$\hat{m}$	0.7295	0.7229	0.0114	0.0238	0.0711

Table 9: Inference on NGARCH model with platykurtic standard EPD innovations - Bayesian (sample size = 100)

	Mean	Median	MSE	SE	Length B.C.I.
$\hat{b}_0$	1.0673	1.0999	1.8057	0.7378	2.1712
$\hat{b}_1$	0.5666	0.5542	1.9747	0.7329	2.1184
$\hat{b}_2$	0.0897	0.0690	2.0706	0.8127	2.3664
$\hat{c}$	0.3134	0.3655	2.1851	0.7765	2.2062
$\hat{\alpha}$	0.4413	0.4290	1.7052	0.8290	2.3289
$\hat{p}$	1.8765	1.7862	2.7136	0.7680	2.2528
$\hat{m}$	0.7733	0.8223	1.7562	0.7618	2.1920

Table 10: Inference on NGARCH model with platykurtic standard EPD innovations- Bayesian (sample size = 200)

	Mean	Median	MSE	SE	Length B.C.I.
$\hat{b}_0$	1.0245	1.0180	1.9701	0.7505	2.1651
$\hat{b}_1$	0.6048	0.6033	1.9753	0.7614	2.1673
$\hat{b}_2$	0.0910	0.0695	1.8150	0.7592	2.1855
$\hat{c}$	0.1103	0.1137	1.9710	0.7652	2.2207
$\hat{\alpha}$	0.4339	0.4159	2.0179	0.8334	2.3363
$\hat{p}$	1.8569	1.8364	2.4870	0.7645	2.2372
$\hat{m}$	0.7374	0.7818	1.7091	0.7608	2.1887

Table 11: Inference on NGARCH model with leptokurtic standard EPD innovations - Bayesian (sample size = 100)

	Mean	Median	MSE	SE	Length B.C.I.
$\hat{b}_0$	1.0658	1.0426	2.6684	0.7558	2.2449
$\hat{b}_1$	0.5747	0.6410	2.1234	0.7612	2.2596
$\hat{b}_2$	0.0397	0.0543	2.2551	0.8515	2.5028
$\hat{c}$	0.2298	0.2653	1.8532	0.8688	2.5437
$\hat{\alpha}$	0.2785	0.2645	2.1751	0.7735	2.3194
$\hat{p}$	3.0101	2.9944	1.6840	0.7393	2.2032
$\hat{m}$	0.6503	0.6822	1.6769	0.8303	2.3354

Table 12: Inference on NGARCH model with leptokurtic standard EPD innovations - Bayesian (sample size = 200)

	Mean	Median	MSE	SE	Length B.C.I.
$\hat{b}_0$	1.0555	1.0420	2.2799	0.7427	2.1558
$\hat{b}_1$	0.6067	0.6024	1.8210	0.7705	2.2592
$\hat{b}_2$	0.1816	0.2051	2.0638	0.7964	2.3009
$\hat{c}$	0.0624	0.1025	1.8735	0.8065	2.3298
$\hat{\alpha}$	0.3800	0.4007	2.1780	0.7845	2.3158
$\hat{p}$	3.0015	2.9806	1.6858	0.7316	2.1483
$\hat{m}$	0.7182	0.7274	1.8412	0.7598	2.1802

One can comment on bias reduction, the relation between the mean, the median, and the variability of the estimates. The general observation is that the bias as well as SE tend to reduce as the sample size increases. As a result of this, the length of B.C.I. decreases. Some quantities increases as sample increases while one would expect them to decrease. These counter-intuitive results are due to outliers in simulations.

6 Conclusions

This paper has discussed flexible distributions and more precisely their use in the context of conditional volatility modelling. Several aspects of these distributions have been touched upon like the benefits to the description of light and fat tails as well as the allocation of the mass of data towards a particular side of the density. Their potential use in practice, especially in finance, has been elaborated on and empirical evidence of their additional value to the field has been provided. Indeed, through simulation studies, the asymptotic theoretical results have been assessed on finite samples in the frequentist and Bayesian framework. The outcome supports the evidence of minor discrepancies, *e.g.* some instances where the SE and the bias of certain estimates were greater for some parameter as the sample size increased from 100 to 200 which is not in line the expected asymptotic behaviour.

Possibility for further research is wide. Something as simple as using different priors in the Bayesian setting can lead to different results. As a consequence of the claim of Rubio and Steel [16], other symmetric unimodal densities can be used to get other flexible distributions and model the return of any financial asset afterwards. Additionally, one can proceed to the inference on the NGARCH model proposed by Zhu and Galbraith [23] using alternative techniques such as the Expectation-Maximisation algorithm discussed by Casella and Berger [5] and artificial neural networks based schemes. Lastly, one can investigate the benefits of flexible distributions to trading strategies. These distributions coupled with rigorous risk management rules adapted to different risk appetites of traders can be implemented in order to either limit potential losses or boost prospective return through a more accurate estimation of conditional volatility.

Source codes

Given their length, the R codes required to implement the aforementioned models and inferences are not provided within this dissertation. Nevertheless, they are available on the Rpubs page of the author should the reader be interested in gaining a greater understanding of the theory:

1. Inference on NGARCH model with standard Gaussian innovations - Frequentist approach:
   `http://rpubs.com/emmanueldjanga/528804`.

2. Inference on NGARCH model with standard EPD innovations - Frequentist approach:
   `http://rpubs.com/emmanueldjanga/528823`

3. Inference on NGARCH model with standard Gaussian innovations - Bayesian approach:
   `http://rpubs.com/emmanueldjanga/528879`

4. Inference on NGARCH model with standard Gaussian innovations - Bayesian approach:
   `http://rpubs.com/emmanueldjanga/529811`

List of Figures

List of Tables

List of Theorems

References

Article Sources

[2] Tim Bollerslev et al. "A conditionally heteroskedastic time series model for speculative prices and rates of return". In: *Review of economics and statistics* 69.3 (1987), pp. 542–547.

[3] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity". In: *Journal of econometrics* 31.3 (1986), pp. 307–327.

[8] Robert F Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica: Journal of the Econometric Society* (1982), pp. 987–1007.

[11] Bruce E Hansen. "Autoregressive conditional density estimation". In: *International Economic Review* (1994), pp. 705–730.

[13] Ivana Komunjer. "Asymmetric power distribution: Theory and applications to risk measurement". In: *Journal of applied econometrics* 22.5 (2007), pp. 891–921.

[15] Rohan Longmore and Wayne Robinson. "Modelling and forecasting exchange rate dynamics: an application of asymmetric volatility models". In: *Bank of Jamaica, Working Paper, WP2004* 3 (2004), pp. 191–217.

[16] FJ Rubio, MFJ Steel, et al. "Bayesian modelling of skewness and kurtosis with two-piece scale and shape distributions". In: *Electronic Journal of Statistics* 9.2 (2015), pp. 1884–1912.

[18] Francisco Javier Rubio and Mark FJ Steel. "Bayesian modelling of skewness and kurtosis with two-piece scale and shape transformations". In: (2014).

[19] Clayton Scott and Robert Nowak. "Maximum likelihood estimation". In: *Connexions Module Webpage. Retrieved January* 25 (2004), p. 2006.

[21] Kar-Ann Toh. "Global optimization by monotonic transformation". In: *Computational Optimization and Applications* 23.1 (2002), pp. 77–99.

[22] Dongming Zhu and John W Galbraith. "A generalized asymmetric Student-t distribution with application to financial econometrics". In: *Journal of Econometrics* 157.2 (2010), pp. 297–305.

[23] Dongming Zhu and John W Galbraith. "Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions". In: *Journal of Empirical Finance* 18.4 (2011), pp. 765–778.

[24] Dongming Zhu and Victoria Zinde-Walsh. "Properties and estimation of asymmetric exponential power distribution". In: *Journal of econometrics* 148.1 (2009), pp. 86–99.

Book Sources

[1] David Ardia et al. *Financial risk management with Bayesian estimation of GARCH models*. Springer, 2008.

[4] John Y Campbell et al. *The econometrics of financial markets*. princeton University press, 1997.

[5] George Casella and Roger L Berger. *Statistical inference*. Vol. 2. Duxbury Pacific Grove, CA, 2002.

[6] Frederik Michel Dekking et al. *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media, 2005.

[10] William H Greene. *Econometric analysis*. Pearson Education India, 2003.

Proceedings Sources

[7] Luc Devroye. "Sample-based non-uniform random variate generation". In: *Proceedings of the 18th conference on Winter simulation*. ACM. 1986, pp. 260–265.

[20] Tetsuya Takaishi. "Markov chain Monte Carlo on asymmetric GARCH model using the adaptive construction scheme". In: *International Conference on Intelligent Computing*. Springer. 2009, pp. 1112–1121.

Lecture Sources

[9] Piotr Fryzlewicz. *Lecture notes: Financial time series, ARCH and GARCH models*. 2007.

[12] Christophe Hurlin. *Chapter 7: Bayesian Econometrics*. 2014.

[17] Francisco Javier Rubio. *Statistical Inference*. 2018.