HEAVY TAILED DISTRIBUTIONS AND THEIR APPLICATIONS IN FINANCE

A Project Report Submitted

for the Course

MA499 Project II

by

Abhinav R & Joel Raja Singh

(Roll No. 170123003 & 170123063)

to the

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781039, INDIA

April 2021

CERTIFICATE

This is to certify that the work contained in this project report entitled **"HEAVY TAILED DISTRIBUTIONS AND THEIR APPLICATIONS IN FINANCE"** submitted by **Abhinav R (Roll No.: 170123003)** and **Joel Raja Singh (Roll No.: 170123063)** to the Department of Mathematics, Indian Institute of Technology Guwahati towards partial requirement of Bachelor of Technology in Mathematics and Computing has been carried out by them under my supervision.

It is also certified that, along with literature survey, computational implementations have been carried out/simulation studies have been carried out-/empirical analysis has been done by the students under the project.

Turnitin Similarity: 29%

Guwahati - 781 039 (Dr. Subhamay Saha)
April 2021 Project Supervisor

ABSTRACT

Historically, selecting a distribution to model the log-returns of financial assets has been an important and challenging task. It is crucial to capture the possible extreme gains and losses with a model of appropriate tail-behaviour for accurate risk estimation. Many classical financial models assume that the returns follow the Gaussian distribution. However, real market data show that extreme events happen more often than predicted by Gaussian models indicating heavy tails. To this end, in the first part of the project, we study the data from the Indian financial markets to assess the performance of three classes of distributions, Gaussian, Stable and Power Law-tails/Pareto-tails in providing reasonable estimates of the risk measure, Value at Risk (VaR). In the second part, we discuss multivariate models such as the Normal Mixture distributions and assess their goodness of fit to market data as well.

In order to do so, we first study the properties of the class of Infinitely Divisible Distributions in Chapter 1. This is followed by establishing the properties of the class of Stable Distributions in Chapter 2, a special class of Infinitely Divisible Distributions. This chapter also describes Domains of Attraction which are closely related to the Stable distributions. In Chapter 3, we adapt the methodology found in [2] to Indian stock data to assess the performance of the three models in estimating VaR. Further, Chapter 4 establishes the class of Normal Mixture distributions and an important special case, the Generalized Hyperbolic distributions. Chapter 5 explains Copulas, which allow us to study dependence structures between random variables. Finally in Chapter 6, we first use the methodology from [2] again to the marginals of the mixture models and also undertake a multivariate/copula analysis of stocks from the Indian markets.

Contents

List of Figures

Chapter 1

Infinitely divisible distributions

In this chapter we provide some elementary theory of Infinitely Divisible distributions. The results from this chapter will be used in studying Stable distributions in Chapter 2 which can be characterized as the only possible limits of appropriately normalized sums of independent random variables.

1.1 Heavy tails and Fat tails

Definition 1.1.1. The distribution of a real valued random variable X is said to have a *heavy right tail* if the tail probabilities $\mathbb{P}(X > x)$ decay more slowly than those of exponential distributions. That is,

$$\lim_{x\to\infty} \mathbb{P}(X > x)e^{\lambda x} = \infty \quad \forall\lambda > 0$$

Heavy left tail is defined analogously for the left tail probabilities.

Example: Let X be a random variable with the distribution function

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^k} & , x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

In this case X is said to have Weibull distribution with shape parameter $k > 0$ and scale parameter $\beta > 0$. When $k < 1$, X has heavy right tail since

$$\lim_{x \to \infty} \lambda x - \left(\frac{x}{\beta}\right)^k = \infty \;\; \forall \lambda > 0$$

Definition 1.1.2. The distribution of a random variable X is said to have a fat right tail if there exists $\alpha > 0$, called the tail index such that

$$m \leq \liminf_{x \to \infty} x^\alpha \mathbb{P}(X > x) \leq \limsup_{x \to \infty} x^\alpha \mathbb{P}(X > x) \leq M$$

for some $M, m > 0$.

Example: X is said to have a Pareto distribution with shape parameter $\alpha > 0$ and scale parameter $x_m > 0$ if the distribution function of X is given by

$$F(x) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^\alpha & , x \geq x_m \\ 0 & \text{otherwise} \end{cases}$$

Then X is fat tailed with tail-index α.

1.2 Infinitely divisible distributions

Definition 1.2.1. The distribution of a random variable X is infinitely divisible if for each $n \in \mathbb{N}$ there exists i.i.d random variables $Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}$

such that

$$X \stackrel{d}{=} Y_1^{(n)} + Y_2^{(n)} + \ldots + Y_n^{(n)} \tag{1.1}$$

In terms of characteristic functions, a random variable X (with characteristic function ϕ) is said to be infinitely divisible if for each $n \in \mathbb{N}$ there exists a characteristic function ϕ_n such that

$$\phi(t) = \phi_n^n(t) \tag{1.2}$$

Definition 1.2.2. Let N be a Poisson(λ) random variable and let X_i be i.i.d with characteristic function ϕ and independent of N. Define

$$Y = \sum_{i=1}^{N} X_i \tag{1.3}$$

Then Y is said to be a Compound Poisson random variable.

Let ψ be the characteristic function of Y. Clearly,

$$\psi(t) = \mathbb{E}\left[e^{itY}\right] = \mathbb{E}\left[e^{it\sum_{i=1}^{N} X_i}\right] = \mathbb{E}\left[\phi^N(t)\right] = \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k\phi^k(t)}{k!} = e^{\lambda(\phi(t)-1)} \tag{1.4}$$

Thus, every compound poisson random variable is infinitely divisible with $\phi_n(t) = e^{\frac{\lambda}{n}(\phi(t)-1)}$, which is the characteristic function of a compound poisson random variable with poisson parameter $\frac{\lambda}{n}$.

We shall now show every infinitely divisible distribution is the limit of a sequence of compound poisson distributions. Hence we state the following theorem.

Theorem 1.2.1. Let ϕ_n be a sequence of characteristic functions. In order that there exist a continous limit

$$\omega(t) = \lim_{n\to\infty} \phi_n^n(t) \tag{1.5}$$

it is necessary and sufficient that

$$n[\phi_n(t) - 1] \to \psi(t) \tag{1.6}$$

with ψ continous. In this case

$$\omega = e^{-\psi}$$

Proof. Refer [5] XVII.1, Theorem 1 $\square$

The convergence condition in this theorem is equivalent to

$$e^{pn[\phi_n(t)-1]} \to e^{p\psi(t)} = \omega^p(t) \tag{1.7}$$

The left side is characteristic function of a compound poisson random variable and thus ω^p is a characteristic function for all $p > 0$. Therefore, ω is infinitely divisible with $\phi_n = \omega^{\frac{1}{n}}$. With that we show the following result:

Theorem 1.2.2. A characteristic function is infinitely divisible if and only if there exists a sequence ϕ_n of characteristic functions such that $\phi_n^n \to \omega$. In this case ω^p is a characteristic function for any $p > 0$ and $\omega(t) \neq 0$ for all t.

Proof. The "if" part of the proof is a direct result of Theorem 1.2.1.

If a characteristic function ω is infinitely divisible, then for each $n \in \mathbb{N}$, there is a characteristic function ϕ_n such that $\omega = \phi_n^n$. Clearly, the constant sequence $\phi_n^n \to \omega$. $\square$

Corollary 1.2.1. A continuous limit ω of a sequence of infinitely divisible distributions $\{\omega_n\}$ is infinitely divisible.

Proof. As ω_n is infinitely divisible, by Theorem 1.2.2, $\phi_n = \omega_n^{1/n}$ is a characteristic function. Because $\omega_n \to \omega$, clearly $\phi_n^n \to \omega$ making ω infinitely divisible by Theorem 1.2.2 □

Theorem 1.2.3. The class of infinitely divisible distributions coincides with the class of limit distributions of compound poisson distributions.

Proof. Since compound poisson distributions are infinitely divisible, a continuous limit of compound poisson distributions is also infinitely divisible.
Let ω be a characteristic function of an infinitely divisible distribution. From Theorem 1.2.2, there exists sequence of characteristic functions ϕ_n such that $\phi_n^n \to \omega$. From (1.7) with $p = 1$,

$$e^{n[\phi_n(t)-1]} \to \omega(t)$$

The left side corresponds to a characteristic function of a compound poisson distribution for all $n \geq 1$. Thus ω can be expressed as a limit of compound poisson distributions. □

1.3 Canonical and Lévy measures

We now try to find a generalised expression for characteristic functions of infinitely divisible distributions. Since we know that any infinitely divisible distribution can be expressed as a continuous limit of compound poisson distributions, We seek to find the general form of limits of sequence of characteristic functions $e^{c_n(\phi_n - 1)}$. For various applications it would be useful to allow arbitrary centering, and hence we seek the possible limit of characteristic functions of the form $\omega_n = e^{\psi_n}$ where,

$$\psi_n(t) = c_n[\phi_n(t) - 1 - i\beta_n t] \tag{1.8}$$

where ϕ_n is the characteristic function of a probability distribution F_n, the c_n are positive constants and the centering constants β_n are real.

Naturally we would choose expectations to be the centering constants. But since we work with fat tailed distribution which may not have expectations we look for a more universally applicable centering constant. One such constant would be such that for $t = 1$, $\psi_n(t)$ is real. If $u_n = Re(\phi_n)$ and $v_n = Im(\phi_n)$, then this choice is,

$$\beta_n = v_n(1) = \int_{-\infty}^{\infty} \sin(x) F_n\{dx\} \tag{1.9}$$

Hence (1.8) can be written as

$$\psi_n(t) = c_n \int_{-\infty}^{\infty} [e^{itx} - 1 - it \sin x] F_n\{dx\} \tag{1.10}$$

Lemma 1.3.1. Let $\{c_n\}$ and $\{\phi_n\}$ be given. If there exists centering constants β_n such that ψ_n tends to a continuous limit ψ, then even with β_n given by (1.9), ψ_n (given by (1.10)) would converge to a continuous limit.

Proof. Define ψ_n by (1.8) wth arbitrary β_n such that $\psi_n \to \psi$, with ψ being continuous. If b denotes the imaginary part of $\psi(1)$, we conclude that

$$c_n(v_n(1) - \beta_n) \to b$$

Multiplying by it and subtracting from $\psi_n \to \psi$ we see that

$$c_n[\phi_n(t) - 1 - iv_n(1)t] \to \psi(t) - ibt$$

which proves our assertion. $\square$

This shows that we can use (1.9) as centering constants and only need to study the limit of characteristic functions given by (1.10) to study the general characteristic function of infnitely divisible distributions.

We begin this problem by taking a simple case in which ψ_n is twice continuously differentiable (that is, the distributions have finite variance) i.e., not only $\psi_n \to \psi$ but also $\psi_n'' \to \psi''$. This means

$$c_n \int_{-\infty}^{+\infty} e^{itx} x^2 F_n\{dx\} \to -\psi''(t) \tag{1.11}$$

By assumption $c_n x^2 F_n\{dx\}$ defines a finite measure, let it's total mass be μ_n. We see from (1.11) that $\mu_n \to -\psi''(0)$. On dividing (1.11) by μ_n we get on the left side a characteristic function of a probability distribution converging to $\psi''(t)/\psi''(0)$. We know that if a sequence of characteristic function converges to a continuous limit, then the limit of the sequence is also a characteristic function. Therefore $\psi''(t)/\psi''(0)$ is the characteristic function of a probability distribution. Hence

$$-\psi''(t) = \int_{-\infty}^{+\infty} e^{itx} M\{dx\} \tag{1.12}$$

where M is a finite measure. From this we obtain ψ by repeated integration. Using that $\psi(0) = 0$ and that with our centering condition $\psi(1)$ is real, we get

$$\psi(t) = \int_{-\infty}^{+\infty} \frac{e^{itx} - 1 - it\sin x}{x^2} M\{dx\} \tag{1.13}$$

We had made the assumption that M is a finite measure. However for the integral in (1.13) to make sense, it is not necessary for M to have a finite mass. It suffices that M be a canonical measure as defined below.

Definition 1.3.1. A measure M is said to be a canonical measure if $M(I)$ is finite for all finite intervals I, and $M^+(x) = \int_x^\infty y^{-2}M\{dy\}$ and $M^-(-x) = \int_{-\infty}^{-x} y^{-2}M\{dy\}$ converge for all $x > 0$. That is, $M\{-x, x\}$ grows slow enough that the two integrals converge.

Lemma 1.3.2. If M is a canonical measure and $\psi(t)$ be defined as in (1.13), then $e^{\psi(t)}$ is an infinitely divisible characteristic function.

Proof. We consider two special cases.

(a) Suppose M is concentrated at the origin and attributes mass $m > 0$ to it. Observe that as $x \to 0$, the integrand in (1.13) tends to $-\frac{t^2}{2}$. Therefore $\psi(t) = \frac{-mt^2}{2}$, and so e^ψ is a normal characteristic function with variance m which is infinitely divisible. (This is because, for any $X \sim N(\mu, \sigma^2)$ and $n \in \mathbb{N}$, we know that $X \overset{d}{=} X_1 + ... + X_n$ where $\{X_i\}$ are i.i.d and each $X_i \sim N(\frac{\mu}{n}, \frac{\sigma^2}{n})$. So, normal variables are infinitely divisible by definition.)

(b) Suppose M is concentrated on $|x| > \eta$ where $\eta > 0$. Then $x^{-2}M\{dx\}$ will be a finite measure since $\int_{-\infty}^{\infty} x^{-2}M\{dx\} = M^+(\eta) + M^-(-\eta) < \infty$ by definition of canonical measure. Let the total mass of M be μ. Then $F\{dy\} = \frac{y^{-2}M\{dy\}}{\mu}$ would be a probability measure with characteristic function ϕ. Then (1.13) can be written as $\psi(t) = \mu[\phi(t) - 1 - ibt]$, where b is a real constant. Thus in this case e^ψ is the characteristic function of a compound poisson distribution and hence is infinitely divisible.

(c) In the general case, let $m \geq 0$ be the mass assigned by M at $x = 0$, and let

$$\psi_\eta(t) = \int_{|x|>\eta} \frac{e^{itx} - 1 - it\sin x}{x^2} M\{dx\} \tag{1.14}$$

Then

$$\psi(t) = \frac{-m}{2}t^2 + \lim_{\eta \to 0^+} \psi_\eta(t) \tag{1.15}$$

We saw that e^{ψ_η} is infinitely divisible. If $m > 0$, e^ψ would represent the characteristic function obtained by adding an infinitely divisible random variable corresponding to e^{ψ_η} to a normal random variable which is also infinitely divisible. Thus e^ψ can be represented as a limit of a sequence of infinitely divisible characteristic functions, and hence e^ψ is also infinitely divisible as asserted. $\square$

Definition 1.3.2. A measure Λ defined on $\mathbb{R}\backslash\{0\}$ is said to be a Lévy measure if it satisfies the following:

$$\int_{-\infty}^{\infty} (|x|^2 \wedge 1) d\Lambda(x) < \infty \tag{1.16}$$

Theorem 1.3.1. Let $h(x)$ be a fixed bounded measurable real valued function on $\mathbb{R}$ such that $h(x) = x + O(x^2)$ as $x \to 0$. Then the following are equivalent.

(i) $\phi(t)$ is the characteristic function of an infinitely divisible distribution.

(ii) There exists a canonical measure M in $\mathbb{R}$ and a real constant b such that

$$\phi(t) = \exp\left(ibt + \int_{-\infty}^{\infty} \frac{e^{itx} - 1 - ith(x)}{x^2} dM(x)\right) \tag{1.17}$$

where the integrand is interpreted as $-t^2/2$ at $x = 0$.

(iii) There exists a Lévy measure Λ in $\mathbb{R}\backslash\{0\}$ and real constants $a \geq 0$ and b such that

$$\phi(t) = \exp\left(ibt - \frac{1}{2}at^2 + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - ith(x)\right) d\Lambda(x)\right) \tag{1.18}$$

(iv) There exists a bounded measure K on $\mathbb{R}$ and a real constant b such that

$$\phi(t) = \exp\left(ibt + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dK(x)\right) \tag{1.19}$$

where the integrand is interpreted as $-t^2/2$ at $x = 0$. The constants and measures are determined uniquely by ϕ.

Proof. (i) $\iff$ (ii)

(i) $\Longleftarrow$ (ii): This part of the proof can be implied from Lemma 1.3.2 where $h(x) = \sin x$. It can be seen in Remark 1.3.1 that the canonical measure M is independent of h as long as it satisfies the condition stated in the theorem.

(i) $\Longrightarrow$ (ii): Refer [5] XVII.2, Theorem 1.

(ii) $\iff$ (iii): Given M in (ii), we let $a := M\{0\}$ and $d\Lambda(x) = x^{-2}dM(x)$, $x \neq 0$. It can be seen that this substitution satisfies (1.18).

Since $M\{-1,1\}$ is finite $\int_{-1}^{1} x^2 d\Lambda(x) < \infty$. And by definition of canonical measure $\int_1^\infty d\Lambda(x) < \infty$ and $\int_{-\infty}^{-1} d\Lambda(x) < \infty$. Hence $\int_{-\infty}^{\infty}(|x|^2 \wedge 1)d\Lambda(x) < \infty$ and Λ is a Lévy measure.

Conversely, if Λ is a Lévy measure and (1.18) holds, let $dM(x) = a\delta_0 + x^2 d\Lambda(x)$ where δ_0 is a Dirac measure with unit mass at 0. It can be seen that this substitution satisfies (1.17).

(1.16) implies that for any finite interval I, $M(I) < \infty$. Now we need to show that $M^+(x)$ and $M^-(-x)$ (as defined in Definition 1.3.1) are finite for all $x > 0$. Note that $M^+(x) = \int_x^\infty d\Lambda(x)$ and it can be trivially seen that it is finite for $x \geq 1$. For $0 < x < 1$

$$\int_x^\infty d\Lambda(t) = \int_x^1 d\Lambda(t) + \int_1^\infty d\Lambda(t)$$

$$\int_x^1 d\Lambda(t) \leq x^{-2}\int_x^1 t^2 d\Lambda(t) < \infty$$

Therefore $M^+(x) = \int_x^\infty y^{-2} dM(y) < \infty$. Using similar arguments it can be shown that $M^-(-x) = \int_{-\infty}^{-x} y^{-2} dM(y) < \infty$. Hence M is a canonical measure.

(ii) $\Longleftrightarrow$ (iv): Choose $h(x) = x/(1+x^2)$ and define

$$dK(x) := \frac{1}{1+x^2} dM(x) \tag{1.20}$$

and conversely, $dM(x) = (1+x^2)dK(x)$. Then (1.17) is equivalent to (1.19).

The proof for the uniqueness of representation can be found in [5], XVII.2, Lemma 3. $\square$

Remark 1.3.1. Different choices of $h(x)$ yield the same measures M and Λ in (ii) and (iii) but different constants b; changing h to $\tilde{h}$ corresponds changing b to

$$\tilde{b} := b + \int_{-\infty}^{\infty} \frac{\tilde{h}(x) - h(x)}{x^2} dM(x) = b + \int_{-\infty}^{\infty} (\tilde{h}(x) - h(x)) d\Lambda(x) \tag{1.21}$$

We denote an infinitely divisible random variable with parameters (a, b, Λ) as per equation (1.18) as $\mathrm{ID}(a, b, \Lambda)$.

Remark 1.3.2. If X_1 and X_2 are independent infinitely divisible random variables with parameters (a_1, b_1, Λ_1) and (a_2, b_2, Λ_2), then $X_1 + X_2$ is infinitely divisible with parameters $(a_1 + a_2, b_1 + b_2, \Lambda_1 + \Lambda_2)$. In particular, if $X \sim \mathrm{ID}(a, b, \Lambda)$, then

$$X \stackrel{\mathrm{d}}{=} X_1 + Y \quad \text{with} \quad X_1 \sim \mathrm{ID}(0, 0, \Lambda), Y \sim \mathrm{ID}(a, b, 0) = N(b, a) \tag{1.22}$$

with X_1 and Y independent. Moreover for any finite partition $\mathbb{R} = \bigcup A_i$, we can split X as a sum of independent infinitely divisible random variables X_i with Lévy measure of X_i having supports in A_i.

Definition 1.3.3. A sequence $\{M_n\}$ of canonical measures is said to *converge properly* to the canonical measure M if $M_n\{I\} \to M\{I\}$ for all finite intervals of continuity for M and $M_n^+(x) \to M^+(x)$, $M_n^-(-x) \to M^-(-x)$ for all points of continuity x.

Theorem 1.3.2. Let $\{\omega_n\}$ be the characteristic functions of infinitely divisible distributions. ω_n converges to a continuous limit ω if and only if their corresponding canonical measures $\{M_n\}$ converge properly to a canonical measure $\{M\}$ which is the canonical measure corresponding to ω.

Proof. Refer [5], XVII,2 Theorem 2 □

In the following chapters, we shall use this theorem only in the special case

$$\psi_n(t) = c_n[\phi_n(t) - 1 - ib_n t] \tag{1.23}$$

where $\omega_n = e^{\psi_n}$. In this case, the proper convergence condition takes on the form

$$c_n x^2 F_n\{dx\} \to M\{dx\}, \quad c_n\left(\beta_n - b_n\right) \to b \tag{1.24}$$

where $\beta_n = \int_{-\infty}^{\infty} \sin x F_n\{dx\}$.

1.4 Examples

In this section we will look into a few examples of infinitely divisible distributions and their canonical and Lévy measures.

Example 1.4.1. The *normal distribution* $N(\mu, \sigma^2)$ with characteristic function $\phi(t) = \exp(-\frac{1}{2}\sigma^2 t^2)$ has $\Lambda = 0$ and $a = \sigma^2$. Thus $M = K = \sigma^2\delta_0$ where δ_0 is a Dirac measure with unit mass at 0 and $b = 0$ for any h. Thus $N(\mu, \sigma^2) = \mathrm{ID}(\sigma^2, \mu, 0)$.

Example 1.4.2. The *Poisson distribution*, $\mathrm{Po}(\lambda)$ with characteristic function $\phi_\lambda(t) = \exp(\lambda(e^{it} - 1))$ has $M = \Lambda = \lambda\delta_1$ and $K = \frac{\lambda}{2}\delta_1$ where δ_1 is a Dirac measure with unit mass at 1. Further $b = \lambda h(1)$ and $a = 0$. Thus $\mathrm{Po}(\lambda) = \mathrm{ID}(0, \lambda h(1), \lambda\delta_1)$.

Example 1.4.3. *Gamma distributions* with density $f_\alpha(t) = \frac{e^{-x}x^{t-1}}{\Gamma(\alpha)}$ for $x > 0$ which has the characteristic function $\phi_\alpha(t) = (1 - it)^{-\alpha}$ is clearly infinitely divisible. To put it in canonical form note that

$$(\log(\phi_\alpha(t)))' = i\alpha(1 - it)^{-1} = i\alpha \int_0^\infty e^{itx-x}dx \tag{1.25}$$

On integration

$$\log(\phi_\alpha(t)) = \alpha \int_0^\infty \frac{e^{itx} - 1}{x} e^{-x}dx \tag{1.26}$$

Thus the canonical measure is defined as follows:

$$dM(x) = \alpha x e^{-x},\ x > 0 \tag{1.27}$$

As described in Theorem 1.3.1, we define the Lévy measure as:

$$d\Lambda(x) = \frac{\alpha e^{-x}}{x} \tag{1.28}$$

Chapter 2

Stable Distributions and Domains of Attraction

In this chapter, we derive some key results regarding Stable Distributions and their Domains of Attraction using the ideas about Infinitely Divisible Distributions developed in the previous chapter.

We start with some useful results from analysis.

2.1 Regularly Varying Functions

Definition 2.1.1. Consider the function $L\colon [0,\infty) \to [0,\infty)$. It is said to be *Slowly Varying* at infinity if:

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1 \quad \forall t > 0$$

Lemma 2.1.1. If a function L varies slowly at infinity, then for any fixed $\varepsilon > 0$,

$$x^{-\varepsilon} < L(x) < x^{\varepsilon}$$

for sufficiently large x.

Examples: Clearly, any L for which $\lim_{x\to\infty} L(x) = k \in (0,\infty)$ is slowly varying. Functions such as $\tan^{-1}(x)$ and $\frac{1}{1+e^{-x}}$ satisfy this property. This condition is sufficient for slow variation but not necessary as seen from the simple example $\log(x)$ for which $\lim_{x\to\infty} \log(x) = \infty$ but $\lim_{x\to\infty} \frac{\log(tx)}{\log(x)} = \frac{\log(t)+\log(x)}{\log(x)} = 1 \quad \forall t > 0$.

Definition 2.1.2. A function $U\colon [0,\infty) \to [0,\infty)$ is said to be *Regularly Varying* at infinity with exponent $-\infty < \rho < \infty$ if, for some slowly varying L,

$$U(x) = x^{\rho} L(x)$$

Lemma 2.1.2. If U is a monotone positive function on $[0,\infty)$ and $\{\lambda_n\}$ and $\{a_n\}$ are sequences such that:

1. $\frac{\lambda_{n+1}}{\lambda_n} \to 1$
2. $a_n \to \infty$
3. $\lambda_n U(a_n x) \to \chi(x) \leq \infty$ exists on a dense set.
4. χ is finite and positive in some interval.

Then, U varies regularly and $\chi(x) = cx^{\rho}$ for some $\rho \in (-\infty,\infty)$

Proof. Since $a_n \to \infty$, for all $t > 0$ there is a finite n_t defined by:

$$n_t = \min\{m : a_{m+1} > t\}$$

$$\Rightarrow a_{n_t} \leq t < a_{n_t+1}$$

Let us assume U is non-decreasing, WLOG. From monotonicity of U,

$$\frac{U(a_{n_t}x)}{U(a_{n_t+1})} \leq \frac{U(tx)}{U(t)} \leq \frac{U(a_{n_t+1}x)}{U(a_{n_t})}$$

$$\Rightarrow \left(\frac{\lambda_{n_t+1}}{\lambda_{n_t}}\right)\left(\frac{\lambda_{n_t}}{\lambda_{n_t+1}}\frac{U(a_{n_t}x)}{U(a_{n_t+1})}\right) \leq \frac{U(tx)}{U(t)} \leq \left(\frac{\lambda_{n_t}}{\lambda_{n_t+1}}\right)\left(\frac{\lambda_{n_t+1}}{\lambda_{n_t}}\frac{U(a_{n_t+1}x)}{U(a_{n_t})}\right)$$

Now, let $t \to \infty$ and use the conditions 1 and 3 to obtain:

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = \frac{\chi(x)}{\chi(1)}$$

Further for $x, y > 0$,

$$\frac{U(txy)}{U(t)} = \frac{U(txy)}{U(tx)}\frac{U(tx)}{U(t)}$$

Setting $t \to \infty$, we get:

$$\frac{\chi(xy)}{\chi(1)} = \frac{\chi(x)}{\chi(1)}\frac{\chi(y)}{\chi(1)}$$

It can be shown that the solution to this functional equation is necessarily of the form:

$$\frac{\chi(x)}{\chi(1)} = x^{\rho} \quad \rho \in (-\infty, \infty)$$

□

Now we can proceed to connect these results with probability theory.

Definition 2.1.3. Let F be a cumulative distribution function. For $\zeta > 0$ and $-\infty < \eta < \zeta$, define:

$$U_\zeta(x) = \int_{-x}^{x} |y|^\zeta dF(y)$$

$$V_\eta(x) = \int_{(-\infty,x)\cup(x,\infty)} |y|^\eta dF(y)$$

We can note that U_ζ is the ζ^{th} truncated moment on the interval $[-x, x]$ and V_η is the η^{th} tail moment corresponding to this interval. In particular, $U_0(x) = F(x) - F(-x)$ and $V_0(x) = 1 - F(x) + F(-x)$

Theorem 2.1.1. Suppose $U_\zeta(\infty) = \infty$

1. If either U_ζ or V_η are regularly varying, then the following limit exists for some $\alpha \in [\eta, \zeta]$:
$$\lim_{t\to\infty} \frac{t^{\zeta-\eta}V_\eta(t)}{U_\zeta(t)} = c$$
Where, c is of the form:
$$c = \frac{\zeta - \alpha}{\alpha - \eta} \tag{2.1,1}$$

2. Conversely, if (2.1,1) holds with $0 < c < \infty$, then $\alpha \geq 0$ and a slowly varying function L such that:
$$U_\zeta(x) \sim (\alpha - \eta)x^{\zeta-\alpha}L(x) \qquad V_\eta(x) \sim (\zeta - \alpha)x^{\eta-\alpha}L(x)$$
That is, for sufficiently large x, U_ζ and V_η resemble regularly varying functions.

Proof. Refer [5] VIII.9, Theorem 2. □

2.2 Stable Distributions

Definition 2.2.1. Let X be a non-degenerate random variable and let $X_1, X_2, ..., X_n$ be n i.i.d copies of X. Let $S_n = \sum_{i=1}^n X_i$. The distribution of X is said to be *stable* if there exist sequences $\{a_n\}$ and $\{b_n\}$ such that for all $n \geq 1$:
$$S_n \stackrel{d}{=} a_n X + b_n \tag{2.2,1}$$
The distribution is *strictly stable* if $b_n = 0$ for all n.

Now, we explore the general structure of stable distributions using the results of the previous chapter. We shall find that all stable distributions are Infinitely Divisible distributions with a special structure to their canonical measures. Explicit forms of their characteristic functions and general forms of equation (2.2,1) are also shown in the ensuing discussion.

Suppose for some sequences $\{a_n\}$ and $\{b_n\}$, we have, ${a_n}^{-1}S_n - nb_n \overset{d}{\Longrightarrow} U$ for some non-degenerate random variable U. Then, their characteristic functions satisfy:

$$\left(\phi\left(\frac{t}{a_n}\right)e^{-ib_n t}\right)^n \to \omega(t) = e^{\psi(t)}$$

Where, ϕ is the characteristic function of each X_i and ω is the characteristic function of the limit variable U.

Let F be the distribution function of each X_i. Set:

$$\phi_n(t) = \phi\left(\frac{t}{a_n}\right)e^{-ib_n t} \quad , \quad F_n(x) = F\left(a_n\left(x + b_n\right)\right)$$

$$\Longrightarrow \phi_n^n(t) \to \omega(t) = e^{\psi(t)}$$

$$\Longrightarrow n\left(\phi_n(t) - 1\right) \to \psi(t) \qquad (2.2,\ 2)$$

from Theorem 1.2.1.

(Note that, necessarily, $\phi_n(t) \to 1$. That is, $\phi\left(\frac{t}{a_n}\right)e^{-ib_n t} \to 1$) (2.2, 3)

Already, we can see that ω is the characteristic function of an infinitely divisible distribution from Theorem 1.2.2. Also, letting $\psi_n(t) = n(\phi_n(t) - 1)$, we see that,

$$\omega_n(t) := e^{\psi_n(t)} \to e^{\psi(t)} = \omega(t)$$

Note that ω_n is the characteristic function of a compound poisson variable which is an infinitely divisible distribution. Thus we may apply Theorem

1.3.2 to the canonical measures of ω_n and ω.

Expanding the left hand side of (2.2, 2), we get,

$$n\int_{-\infty}^{+\infty} \frac{e^{itx}-1}{x^2} x^2 dF_n(x) \quad \to \quad e^{\psi(t)}$$

Let $M_n\{dx\} = nx^2 F_n\{dx\}$ and let M be the canonical measure of ψ and Λ the corresponding Lévy measure. From Theorem 1.3.2, we obtain,

$$nx^2 F_n\{-x, x\} \to M\{-x, x\}, \tag{2.2, 4}$$

$$n(1 - F_n(x)) \to \Lambda\{x, \infty\}$$

$$n(F_n(-x)) \to \Lambda\{-\infty, -x\}$$

for all points of continuity $x > 0$.

Let $\mu(x)$ be the truncated moment of F over the interval $[-x, x]$. That is,

$$\mu(x) = \int_{-x}^{x} y^2 dF(y)$$

Suppose F is symmetric, that is $b_n = 0$. Then $F(a_n x) = F_n(x)$. So, (2.2, 4) now becomes:

$$\frac{n}{{a_n}^2}\mu(a_n x) \to M\{-x, x\} \tag{2.2, 5}$$

From (2,2, 3), using $b_n = 0$, we get that $\phi\left(\frac{t}{a_n}\right) \to 1$. Because of this, we get $a_n \to \infty$.

From the above we can also infer that $\frac{S_n}{a_n}$ and $\frac{S_n}{a_{n+1}}$ converge to the same distribution.

$$\implies \frac{a_{n+1}}{a_n} \to 1$$

$$\Longrightarrow \frac{\frac{n+1}{a_{n+1}^2}}{\frac{n}{a_n^2}} \to 1$$

Now, we may use Lemma 2.1.2 with $\lambda_n = \frac{n}{a_n^2}$ and $a_n = a_n$ to infer that μ is regularly varying and that $M\{-x, x\} = cx^\rho$ for some $\rho \in (-\infty, \infty)$

But for the integral $\Lambda\{x, \infty\} = \int_x^\infty y^{-2} dM(y)$ to converge and for M to be a canonical measure, we require that $\rho = 2 - \alpha$ for some $\alpha \in (0, 2]$. (The case $\alpha = 2$ corresponds to M assigning an atom at $x = 0$ which was shown to be the canonical measure of the Normal distribution in Chapter 1.)

In summary, we have obtained,

$$M\{-x, x\} = cx^{2-\alpha}$$

$$\Longrightarrow \Lambda\{x, \infty\} = c\frac{2-\alpha}{\alpha}x^{-\alpha}, \quad \alpha \in (0, 2]$$

If F is not symmetric, note that $\phi_n \to 1$ and $a_n \to \infty$ imply that $b_n \to 0$. This shows that we can use arguments similar to those we used in the symmetric case for equations (2.2, 4) to obtain:

$$M\{-y, x\} = \left(C_- y^{2-\alpha} + C_+ x^{2-\alpha}\right), \quad \alpha \in (0, 2]$$

$$\Lambda\{x, \infty\} = C_+ \frac{2-\alpha}{\alpha} x^{-\alpha}$$

$$\Lambda\{-\infty, -x\} = C_- \frac{2-\alpha}{\alpha} x^{-\alpha} \tag{2.2, 6}$$

for $x, y > 0$ and some constants $C_+, C_- > 0$.

We summarize the above results in the theorem below:

Theorem 2.2.1. Let S_n be defined as before. If there is a non-degenerate random variable U such that $a_n{}^{-1} S_n - nb_n \overset{d}{\Longrightarrow} U$ for some sequences $\{a_n\}$ and $\{b_n\}$, then U is infinitely divisible and the canonical and Lévy measures of U have the structure of (2.2, 6).

As an immediate consequence, we see that stable distributions are infinitely divisible and find the form of their canonical measures as a corollary to this theorem.

Corollary 2.2.1. A distribution is stable if and only if it is infinitely divisible with canonical measure in the form of (2.2, 6).

Proof. ($\Longrightarrow$): By definition, if X is stable, then there exist $\{a_n\}$, $\{b_n\}$ such that,

$$S_n \stackrel{d}{=} a_n X + b_n$$

$$\Longrightarrow \; a_n^{-1} S_n - n\frac{b_n}{na_n} \stackrel{d}{=} X$$

Trivially,

$$a_n^{-1} S_n - n\frac{b_n}{na_n} \stackrel{d}{\Longrightarrow} X$$

By Theorem 2.2.1, X must have canonical and Lévy measures of the form (2.2, 6).

For the ($\Longleftarrow$) proof, refer [5] XVII.3, Example (g) and (h) where the characteristic function corresponding to M of the form (2.2, 6) is explicitly found and shown to be stable. $\square$

Definition 2.2.2. We call a distribution α-stable, if its canonical measure has exponent $2 - \alpha$.

Theorem 2.2.2. Let $0 < \alpha \leq 2$.

1. A distribution of a variable X is α-stable if and only if its characteristic function is of the form:

$$\phi(t) = \begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left(\cos\frac{\pi\alpha}{2} - i\beta \sin\frac{\pi\alpha}{2}\operatorname{sgn}(t)\right) + i\delta t\right) & \alpha \neq 1 \\ \exp\left(-\gamma|t|\left(1 + i\beta\frac{2}{\pi}\operatorname{sgn}(t)\log(|t|)\right) + i\delta t\right) & \alpha = 1 \end{cases}$$

Where $-1 \leq \beta \leq 1$, $\gamma > 0$ and $-\infty < \delta < \infty$.

2. When the canonical measure of X is represented in the form of (2.2, 6), we have:

$$\gamma^\alpha = \begin{cases} (C_+ + C_-)\frac{\Gamma(3-\alpha)}{\alpha(1-\alpha)} & \alpha \neq 1 \\ (C_+ + C_-)\frac{\pi}{2} & \alpha = 1 \end{cases}$$

$$\beta = \frac{C_+ - C_-}{C_+ + C_-}$$

3. If X has a characteristic function in the form as in point 1, then (2.2, 1) is the explicit relation:

$$S_n \stackrel{d}{=} \begin{cases} n^{1/\alpha}X + (n - n^{1/\alpha})\delta & \alpha \neq 1 \\ nX + \frac{2}{\pi}\beta\gamma n\log(n) & \alpha = 1 \end{cases}$$

In particular, X is strictly stable whenever $\delta = 0$ for $\alpha \neq 1$ and $\beta = 0$ for $\alpha = 1$.

Proof. For points 1 and 2 refer [5] XVII.3, Example (h) where in our notation $p = \frac{C_+}{C_+ + C_-}$, $q = \frac{C_-}{C_+ + C_-}$ and $\delta = b$. The proofs follow from an explicit calculation of their characteristic functions from their canonical measures.

(3) For $\alpha \neq 1$, consider the random variable $Y = n^{1/\alpha}X + (n - n^{1/\alpha})\delta$. By definition of characteristic function and from part 1, we get:

$$
\begin{aligned}
\phi_Y(t) &= \mathbb{E}\left[\exp\{it(n^{1/\alpha}X + (n - n^{1/\alpha})\delta)\}\right] \\
&= \exp\{(n - n^{1/\alpha})\delta\}\mathbb{E}\left[\exp\{i(n^{1/\alpha}t)X\}\right] \\
&= \exp\{(n - n^{1/\alpha})\delta\}\phi(n^{1/\alpha}t) \\
&= \phi(t)^n
\end{aligned}
$$

That is, $Y \stackrel{d}{=} S_n$ as desired.

With exactly similar arguments, we can also show $S_n \stackrel{d}{=} nX + \frac{2}{\pi}\beta\gamma n \log(n)$ when $\alpha = 1$.

$\square$

Definition 2.2.3. We denote the α-stable distribution with characteristic function as per the last theorem as $S_\alpha(\gamma, \beta, \delta)$.

Remark 2.2.1. Now that we have the forms of M and Λ in the conditions (2.2, 4), we can restate them as follows:

$$
\frac{n}{a_n^2}\mu(a_n x) \to (C_- + C_+)x^{2-\alpha}
$$

$$
n\left(1 - F_n(x)\right) \to C_+\frac{2-\alpha}{\alpha}x^{-\alpha}
$$

$$
nF_n(-x) \to C_-\frac{2-\alpha}{\alpha}x^{-\alpha}
$$

We know that $F_n(x) = F(a_n x)$ (assuming symmetry) and we just found that $a_n = n^{1/\alpha}$. For $0 < \alpha < 2$,

$$
\implies (n^{1/\alpha}x)^\alpha(1 - F(n^{1/\alpha}x)) \to C_+\frac{2-\alpha}{\alpha}
$$

$$
(n^{1/\alpha}x)^\alpha F(-n^{1/\alpha}x) \to C_-\frac{2-\alpha}{\alpha}
$$

$$\implies x_n^\alpha(1 - F(x_n)) \to C_+\frac{2-\alpha}{\alpha}$$
$$x_n^\alpha F(-x_n) \to C_-\frac{2-\alpha}{\alpha}$$

Where $x_n = n^{1/\alpha}x$. Clearly, for all $x > 0$, $x_n \to \infty$ and for all $x < 0$, $x_n \to -\infty$.

The last two results state that the tails of any stable distribution show polynomial decay with tail index $0 < \alpha < 2$. That is, stable distributions are fat tailed, by definition.

The Normal distribution ($\alpha = 2$) cannot be inferred as fat tailed from the above analysis as the right hand side of the limit is zero. The definition of fat tails (refer Definition 1.1.2) requires that $m, M > 0$.

Remark 2.2.2. From the structure of ϕ it is easy to see that it is not continuously differentiable at $t = 0$ whenever $\alpha \leq 1$. This immediately tells us that the mean does not exist for any such $S_\alpha(\gamma, \beta, \delta)$.

However, when $\alpha > 1$ we can see that $\mathbb{E}[X] = \frac{\phi^{(1)}(0)}{i} = \delta$.

Remark 2.2.3. Consider $X = S_\alpha(1, \beta, 0)$. For any $\gamma > 0$ and $\delta \in (-\infty, \infty)$, consider the variable $Z = \gamma X + \delta$

$$\begin{aligned}\phi_Z(t) &= \mathbb{E}[\exp\{it(\gamma X + \delta)\}]\\ &= \phi_X(\gamma t)\exp\{i\delta t\}\end{aligned}$$

If $\alpha \neq 1$:

$$\phi_Z(t) = \exp\left(-\gamma^\alpha|t|^\alpha\left(\cos\frac{\pi\alpha}{2} - i\beta\sin\frac{\pi\alpha}{2}\operatorname{sgn}(t)\right) + i\delta t\right)$$

If $\alpha = 1$

$$\exp\left(-\gamma|t|\left(1 + i\beta\frac{2}{\pi}\operatorname{sgn}(t)\log(|t|)\right) + i\hat{\delta}t\right)$$

where $\hat{\delta} = \delta - \frac{2}{\pi}\beta\gamma\log(\gamma)$

That is,

$$Z = \begin{cases} S_\alpha(\gamma, \beta, \delta) & \alpha \neq 1 \\ S_\alpha(\gamma, \beta, \hat{\delta}) & \alpha = 1 \end{cases}$$

Which tells us that δ is a location parameter and γ is a scale parameter.

We can note from the structure of the canonical measure and the fact that $\beta = \frac{C_+ - C_-}{C_+ + C_-}$, that β is a skewness parameter. This means that α and β together determine the shape of the distribution.

Remark 2.2.4. From the structure of ϕ, it is clear that if $X \sim S_\alpha(\gamma, \beta, \delta)$, then $-X \sim S_\alpha(\gamma, -\beta, -\delta)$. (By simply finding $\phi(-t)$).

$\implies X$ is symmetric about 0 whenever $\beta = \delta = 0$. That is, it is neither skewed to the right nor left and is centered at 0.

2.3 Domains of Attraction

Definition 2.3.1. Let S_n be the sum of n i.i.d variables $\{X_i\}$ with common distribution F. We say F is in the *domain of attraction* of a distribution D if there exist sequences $\{a_n\}$ and $\{b_n\}$ such that:

$$\frac{S_n - b_n}{a_n} \xRightarrow{d} D \tag{2.3, 1}$$

Theorem 2.3.1. A distribution D has a domain of attraction if and only if it is stable.

Proof. If D has a domain of attraction, then by Theorem 2.2.1 and Corollary 2.2.1, D must be stable. If a distribution D is stable, then by definition, D is itself in its domain of attraction. $\square$

While this is a fairly simple theorem, it posits a heavy claim that any CLT-type limit theorem can only have stable limits.

We now present the following two theorems characterizing random variables that belong in the domain of attraction of some α-stable distribution. The proofs for the theorems can be found in [5], XVII.5, Theorem 2 and its Corollaries.

Theorem 2.3.2. Let $0 < \alpha \leq 2$. A non-degenerate random variable X belongs to the domain of attraction to an α-stable distribution D if and only if the following conditions hold:

1. The truncated moment function $\mu(x) = \mathbb{E}\left[X^2 \mathbf{I}\{|X| \leq x\}\right]$ varies regularly with exponent $2 - \alpha$ as $x \to \infty$. That is,

$$\mu(x) \sim x^{2-\alpha} L_1(x) \tag{2.3, 2}$$

   Where L_1 is a slowly varying function.

2. Either $\alpha = 2$ or the tails of X are balanced. That is,

$$\lim_{x \to \infty} \frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} = p_+ \tag{2.3, 3}$$

for $p_+ \in [0, 1]$.

Theorem 2.3.3. Let $0 < \alpha < 2$. A non-degenerate random variable X belongs to the domain of attraction to an α-stable distribution D if and only if the following conditions hold:

1. The tail probability varies regularly with exponent α as $x \to \infty$. That is,

$$\mathbb{P}(|X| > x) \sim x^{-\alpha} L_2(x) \tag{2.3, 4}$$

   Where L_2 is a slowly varying function.

2. The tails of X are balanced. That is,

$$\lim_{x \to \infty} \frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} = p_+ \tag{2.3, 5}$$

   for $p_+ \in [0, 1]$.

Remark 2.3.1. Note that D would also be in its domain of attraction as it is stable. From the two theorems above, this immediately tells us that the truncated second moment of D and the tails of D vary regularly with exponents $2 - \alpha$ and α respectively. This agrees with the result in Remark 2.2.1.

Let $0 < \alpha < 2$. The following is true for U_ζ and V_η as per Definition 2.1.3:

$$\mu(x) = U_2(x), \quad T(x) = \mathbb{P}(|X| > x) = V_0(x)$$

If the conditions of one of the above theorems holds, X is in the domain of attraction of some stable distribution. But this implies that the conditions of the other theorem must hold too.

So, if X is in some domain of attraction, then both μ and T vary regularly with their respective exponents. From Theorem 2.1.1,

$$\lim_{x\to\infty} \frac{x^2 T(x)}{\mu(x)} \to c$$

Where $c = \frac{2-\alpha}{\alpha}$.

$$\implies \frac{L_1(x)}{L_2(x)} \sim c$$

Moreover, for such X, conditions (2.2, 4) hold. So,

$$\begin{aligned} n\mathbb{P}(X > a_n x) &\to \Lambda\{x, \infty\} \\ n\mathbb{P}(X < -a_n x) &\to \Lambda\{-\infty, -x\} \end{aligned} \qquad (2.3,\ 6)$$

Equivalently, we can say,

$$n\mathbb{P}(|X| > a_n x) \to \Lambda\{y : |y| > x\}$$

In particular, $n\mathbb{P}(|X| > a_n) \to \Lambda\{y : |y| > 1\} > 0$.

Conversely, for (2.3, 1), if we can find a sequence $a_n \to \infty$ such that,

$$n\mathbb{P}(|X| > a_n) \to C \qquad (2.3,\ 7)$$

for some constant $C \in (0, \infty]$ and if (2.3, 4) and (2.3, 5) hold, then the equations in (2.3, 6) must hold. Combining (2.3, 5) and (2.3, 6), we get,

$$\frac{n\mathbb{P}(X > a_n)}{n\mathbb{P}(|X| > a_n)} \to \frac{\Lambda\{1, \infty\}}{C} = p_+$$

Using the form of $\Lambda\{x, \infty\}$, we obtain,

$$C_+ \frac{2-\alpha}{\alpha} = p_+ C$$

$$\implies \Lambda\{x, \infty\} = p_+ C x^{-\alpha}$$

Similarly, $\frac{n\mathbb{P}(X<-a_n)}{n\mathbb{P}(|X|>a_n)} = \frac{n\mathbb{P}(|X|>a_n) - n\mathbb{P}(X>a_n)}{n\mathbb{P}(|X|>a_n)} = 1 - \frac{n\mathbb{P}(X>a_n)}{n\mathbb{P}(|X|>a_n)} \to 1 - p_+$. Let $p_- = 1 - p_+$. Using the same arguments as before, we get,

$$C_- \frac{2-\alpha}{\alpha} = p_- C$$

$$\implies \Lambda\{-\infty, -x\} = p_- C x^{-\alpha}$$

$$\implies C_+ + C_- = C \frac{\alpha}{2-\alpha}$$

$$\implies C_+ - C_- = C(p_+ - p_-) \frac{\alpha}{2-\alpha}$$

$$\implies \frac{C_+ - C_-}{C_+ + C_-} = p_+ - p_-$$

So, the limit distribution corresponding to this choice of a_n, would satisfy Theorem 2.2.2 with,

$$\gamma^\alpha = \begin{cases} C\Gamma(1-\alpha) & \alpha \neq 1 \\ C\frac{\pi}{2} & \alpha = 1 \end{cases} \tag{2.3, 8}$$

Using the fact that $\frac{C}{2-\alpha} \frac{\Gamma(3-\alpha)}{(1-\alpha)} = C\Gamma(1-\alpha)$.

$$\beta = p_+ - p_- \tag{2.3, 9}$$

These results are instrumental in proving the following useful theorem that characterizes the limit distributions of any scaled and centered sum of random variables in the domain of attraction of some α-stable distribution.

Theorem 2.3.4. Let $0 < \alpha < 2$. If (2.3, 4) and (2.3, 5) hold and a_n is chosen such that (2.3, 7) holds. Let γ and β be as derived in (2.3, 8) and (2.3, 9). Then,

1. If $0 < \alpha < 1$, then,
$$\frac{S_n}{a_n} \stackrel{d}{\Longrightarrow} S_\alpha(\gamma, \beta, 0)$$

2. If $1 < \alpha < 2$, then,
$$\frac{S_n - n\mathbb{E}[X]}{a_n} \stackrel{d}{\Longrightarrow} S_\alpha(\gamma, \beta, 0)$$

3. If $\alpha = 1$, then,
$$\frac{S_n - nb_n}{a_n} \stackrel{d}{\Longrightarrow} S_1(\gamma, \beta, 0)$$
where $b_n = a_n\mathbb{E}\left[sin\left(\frac{X}{a_n}\right)\right]$.

Proof. Combine the above results with [5], XVII.5, Theorem 3. □

Now, we proceed to cite some examples from [8] where we find a suitable a_n such that we may use the above theorem to obtain limit laws.

2.4 Examples

Example 2.4.1. Let $0 < \alpha < 2$ and $C > 0$. Suppose X is a random variable such that,

$$\begin{aligned} \mathbb{P}(X > x) &\sim Cx^{-\alpha} \\ \mathbb{P}(X < -x) &= o(x^{-\alpha}) \\ \implies \mathbb{P}(|X| > x) &\sim Cx^{-\alpha} \\ \implies \frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} &\to 1 \end{aligned} \tag{2.3, 10}$$

That is, (2.3, 4) holds with $L_2(x) = C$ and (2.3, 5) holds with $p_+ = 1, p_- = 0$.

Choosing $a_n = n^{1/\alpha}$, we see that,

$$n\mathbb{P}(|X| > a_n) \to C > 0$$

Thus, (2.3, 7) holds as well. All the conditions of Theorem 2.3.4 are met, so this yields,

1. If $0 < \alpha < 1$, then,
$$\frac{S_n}{n^{1/\alpha}} \overset{d}{\Longrightarrow} S_\alpha(\gamma, 1, 0)$$

2. If $1 < \alpha < 2$, then,
$$\frac{S_n - n\mathbb{E}[X]}{n^{1/\alpha}} \overset{d}{\Longrightarrow} S_\alpha(\gamma, 1, 0)$$

3. If $\alpha = 1$, then,
$$\frac{S_n - nb_n}{n} \overset{d}{\Longrightarrow} S_1(\gamma, 1, 0)$$
where $b_n = n\mathbb{E}\left[sin\left(\frac{X}{n}\right)\right]$.

where we obtain γ from (2.3, 8) with $C = C$ and $\beta = p_+ - p_- = 1$

.

Example 2.4.2. Suppose for $0 < \alpha < 2, C > 0$, we have a natural number valued random variable X such that as $n \to \infty$,

$$\mathbb{P}(X = n) \sim cn^{-\alpha-1}$$

So for any $n \in \mathbb{N}$,

$$\mathbb{P}(X > n) = \sum_{i \in \mathbb{N}, i > n} P(X = i)$$

We know that, for $i \in \mathbb{N}, i \geq 2$ and for any $u \in [i, i+1]$,

$$\frac{1}{u^{\alpha+1}} \leq \frac{1}{i^{\alpha+1}} \leq \frac{1}{(u-1)^{\alpha+1}}$$

Integrating this inequality over the interval of u we get,

$$\frac{-1}{\alpha}\left((i+1)^{-\alpha} - i^{-\alpha}\right) \leq i^{-\alpha-1} \leq \frac{-1}{\alpha}\left(i^{-\alpha} - (i-1)^{-\alpha}\right)$$

Summing the inequalities for $i = n+1$ to N gives,

$$\frac{-1}{\alpha}\left((N+1)^{-\alpha} - (n+1)^{-\alpha}\right) \leq \sum_{i=n+1}^{N} i^{-\alpha-1} \leq \frac{-1}{\alpha}\left(N^{-\alpha} - n^{-\alpha}\right)$$

In the limit $N \to \infty$,

$$\frac{(n+1)^{-\alpha}}{\alpha} \leq \sum_{i > n, i \in \mathbb{N}} i^{-\alpha-1} \leq \frac{n^{-\alpha}}{\alpha}$$

For large n, the summation is equivalent to $\frac{1}{c}\sum_{i>n, i \in \mathbb{N}} P(X = i)$, that is as $n \to \infty$,

$$\frac{c(n+1)^{-\alpha}}{\alpha} \leq \mathbb{P}(X > n) \leq \frac{cn^{-\alpha}}{\alpha}$$

For large x,

$$\frac{(c\lceil x \rceil)^{-\alpha}}{\alpha} \leq \mathbb{P}(X > x) \leq \frac{c\lfloor x \rfloor^{-\alpha}}{\alpha}$$

So, we may write $\mathbb{P}(X > x) \sim Cx^{-\alpha}$ where $C = \frac{c}{\alpha}$. That is, we have shown that this example is identical to the previous example upto the constant C.

Now, we consider some examples for the case $\alpha = 1$ to illustrate the calculation of the centering sequence b_n.

Example 2.4.3. Let $X = 1/U$, where $U \sim U(0,1)$. Then $\mathbb{P}(X > x) = \mathbb{P}(U < x) = x^{-1}$ for $x \geq 1$. That is,

$$\mathbb{P}(X > x) \sim x^{-1}$$

that is, the constant $C = 1$ and we may use the results from Example 1. Thus, $\gamma = \frac{\pi}{2}$ and $\beta = 1$.

Further, X has the following density f,

$$f(x) = \begin{cases} x^{-2} & x > 1 \\ 0 & x \leq 1 \end{cases}$$

$$
\begin{aligned}
\implies b_n &= n\mathbb{E}\left[\sin\left(\frac{X}{n}\right)\right] \\
&= n\int_1^\infty x^{-2}\sin(x/n)dx \\
&= \int_{1/n}^\infty y^{-2}\sin(y)dy \\
&= \int_{1/n}^1 \frac{1}{y}dy + \int_{1/n}^1 \frac{\sin(y)-y}{y^2}dy + \int_1^\infty \frac{\sin(y)}{y^2}dy \\
&= \log(n) + \int_{1/n}^\infty \frac{\sin(y) - y\mathbf{I}\{y<1\}}{y^2}dy
\end{aligned}
$$

The final integral is known to be equal to $1-\tilde{\gamma}+o(1)$ where $\tilde{\gamma}$ is the Euler-Macheroni constant (refer [7]).

So, from Example 1, point 3, we get,

$$
\frac{S_n}{n} - (\log(n) + 1 - \tilde{\gamma}) \overset{d}{\implies} S_1(\frac{\pi}{2}, 1, 0)
$$

Equivalently,

$$
\frac{S_n}{n} - \log(n) \overset{d}{\implies} S_1(\frac{\pi}{2}, 1, 1-\tilde{\gamma})
$$

Finally, we may consider the case $\alpha = 2$. Here, there is no meaningful Λ as it vanishes identically. So, we choose a_n corresponding to relation (2.2, 4) such that,

$$
\frac{n\mu(a_n)}{{a_n}^2} \to C > 0 \tag{2.3, 11}
$$

Once such a choice is made, we have the following theorem analogous to Theorem 2.3.4:

Theorem 2.4.1. Let X be a random variable with truncated second moment function $\mu(x)$. Let a_n be chosen according to (2.3, 11). If $\mu(x)$ is slowly varying such that $\mu(x) \to \infty$ as $x \to \infty$, then,

$$\frac{S_n - \mathbb{E}[S_n]}{a_n} \overset{d}{\Longrightarrow} N(0, C)$$

where N is the normal distribution. This is in the form similar to the Central Limit Theorem.

Proof. Refer [5], XVII.5, Theorems 2 and 3. $\square$

Example 2.4.4. Let X be a random variable such that as $x \to \infty$,

$$\mathbb{P}(X > x) \sim Cx^{-2}$$

where $C > 0$. Also suppose $\mathbb{P}(X < -x) = o(x^{-2})$. Then, (2.3, 3) holds with $p_+ = 1$ and $p_- = 0$. Say $F(t) = \mathbb{P}(X > t)$ and $G(t) = \mathbb{P}(X < -t)$. Then as $x \to \infty$,

$$\begin{aligned}
\mu(x) &= \mathbb{E}\left(\int_0^{|X|} 2t\mathbf{I}\{|X| > x\}dt\right) \\
&= \mathbb{E}\left(\int_0^{x} 2t\mathbf{I}\{t \le |X| \le x\}dt\right) \\
&= \int_0^x 2t\mathbb{P}(t \le |X| \le x)dt \\
&= \int_0^x 2t\mathbb{P}(|X| > t)dt - x^2\mathbb{P}(|X| > x) \\
&= \int_0^x 2t(F(t) + G(t))dt + O(1) \\
&= \int_1^x 2tCt^{-2}dt + O(1) \sim 2C\log(x)
\end{aligned}$$

So, we may choose $a_n = \sqrt{n \log(n)}$. This yields,

$$\frac{n\mu(a_n)}{{a_n}^2} = \frac{nC\log(n) + nC\log(\log(n))}{n\log(n)} \to C$$

By virtue of Theorem 2.3.5, we get,

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{n\log(n)}} \overset{d}{\Longrightarrow} N(0, C)$$

Chapter 3

Empirical Analysis I

In this chapter we look into some empirical data from the Indian stock markets and observe the distributions of their log-returns. Many financial models assume them to be normally distributed. In this chapter we fit the data to normal, stable and pareto distributions and see how well they fit our data with their ability to accurately predict the Value at Risk. To this end, we adopt the methodology found in [2] to Indian data.

3.1 Introduction

The Value at Risk(VaR) is one of the main indicators for risk management of financial portfolios. VaR is expressed as the threshold that a loss over a chosen time horizon occurs with atmost a given level of confidence. Mathematically, if X is the negative profit of a portfolio over a time horizon T , then for some $0 < \alpha < 1$,

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X < x) \geq \alpha\}$$

In this chapter, we only consider the daily VaR, i.e $T = 1$. VaR can be estimated by parametric and non parametric means. Non parametric estimation uses the empirical distribution without fitting a model. Due to small amount of available data, it does not provide an accurate way to deal with extreme events. For instance if we are interested in 99% VaR, in about a year's worth of daily data only about 2 to 3 of them lie in the tail relevant for the computation of 99% VaR.
Parametric estimation on the other hand, fits some distribution (Gaussian /stable/Pareto in this case) to the daily negative profit/loss/return data (negative because we want large losses to be at the right tail), and the VaR can be computed as the α^{th} quantile of such a distribution.

3.2 Overview

Normal distributions and Brownian motion in finance go back to Bachelier's thesis and were popularised by F.Black and M.Scholes with their famous Black Scholes equations. The Black Scholes equations assumed the price process of a stock to be a geometric Brownian motion, i.e the log-returns of the stock are normally distributed. Many theories like CAPM for portfolio management take roots in the Gaussian world.
However, from the 1960's it has been shown that normal distributions do not fit the underlying market data. They fail to properly model extreme market events that happen more often than predicted from models based upon Gaussian distributions. This requires researchers to look for tractable models which reproduce heavy-tails (log-)returns. Stable distributions and processes were first proposed by B. Mandelbrot and E.F Fama. Yet choosing this model imposes $\alpha < 2$. This parameter is however difficult to estimate,

especially when α is close to 2. In addition several empirical studies backed the claim that α may be greater than 2.

Thus an alternative to both Gaussian and Stable distributions for modelling the tail is a general Pareto distribution. This distribution allows us to have tails that are heavier than the normal distribution but not as heavy as stable distributions. They are used to model the tails of the log-losses in order to estimate VaR/CVaR/Expected shortfall and other such helpful parameters for risk management.

3.3 Framework and models

Let us consider the prices of a financial asset $(S_t, t > 0)$ where time is measured in terms of days. The log-returns are defined to be the sequence $(R_t, t \in \mathbb{N})$ given by

$$R_t = \ln \frac{S_{t+1}}{S_t} = \ln S_{t+1} - \ln S_t \tag{3.1}$$

We call log-losses to be the values $L_t = -R_t$ for $R_t < 0$.

For any time $t > 0$, by definition, the α-VaR is such that,

$$\mathbb{P}(S_t - S_{t-1} < -\operatorname{VaR}_\alpha) = 1 - \alpha$$

Consider the $(1-\alpha)^{th}$ quantile of R_t, $q^R_{1-\alpha}$,

$$\mathbb{P}(R_t < q^R_{1-\alpha}) = 1 - \alpha$$

$$\implies \mathbb{P}(S_t - S_{t-1} < S_{t-1}(\exp(q^R_{1-\alpha}) - 1))$$

$$\implies \operatorname{VaR}_\alpha = S_{t-1}(1 - \exp(q^R_{1-\alpha})) \tag{3.2}$$

when given S_{t-1}.

3.3.1 Models

We consider 3 classes of parametric univariate models for the log-returns or log-losses.

(I)The Gaussian model for log-returns defined by its mean μ and standard deviation σ.

(II)The stable distribution for log-returns defined by it's characteristic function as in Theorem 2.2.2.

It can also be shown that, if $X \sim S_\alpha(\gamma, \beta, \delta)$, then as $x \to -\infty$,

$$\mathbb{P}(X \leq x) \sim \frac{C}{|x|^\alpha}$$

Where,

$$C = \gamma^\alpha \sin(\frac{\pi\alpha}{2})\frac{\Gamma(\alpha)}{\pi}(1-\beta)$$

That is, Stable distributions also have Paretian tails.

(III)The Pareto distribution for the tails of log-losses gives the cumulative distribution function F of the log-losses

$$\mathbb{P}[L_t > x] = 1 - F(x) = C/x^\alpha \text{ for } x \geq x_0 \tag{3.3}$$

This model does not make any supplementary assumption on the bulk of the distribution of the log-returns.

3.3.2 Parameter Estimation and Computation of VaR

Choosing a model that fits log returns requires that the series of stock prices (S_t) is stationary over time. This is not true for prices spanning several years. Thus we apply our estimators to a finite window of time : 1 year in our experiments, within which the prices are assumed to be stationary. We now describe the methodology for each model.

(I)**Gaussian Distribution:** Consider n successive daily log-returns $(R_1, R_2, \ldots, R_n)$ assumed to follow the Gaussian distribution. The mean and the standard deviation are estimated by:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} R_i \tag{3.4}$$

$$\hat{\sigma} = \left(\frac{1}{n-1} \sum_{i=1}^{n} (R_i - \hat{\mu})^2 \right)^1 /2 \tag{3.5}$$

and $q^R_{1-\alpha}$ is estimated by $\hat{q}^R_{1-\alpha} = \hat{\mu} + \Phi^{-1}(1-\alpha)\hat{\sigma}$, where Φ is the Standard Normal cumulative distribution function. From this, we use equation (3.2) to get the VaR estimate.

(II)**Stable Distribution:** To fit a stable distribution to the log-returns, we use the Mc-Culloch method as implemented in the R library fbasics to estimate the parameters $\alpha, \beta, \gamma, \delta$ and to compute the quantile, also using functionality in the same library.

(III)**Pareto Distribution:** For this experiment we will use a slight modification of the Hill Estimator, as in [2]. Let $(L_{(1)}, L_{(2)}, \ldots, L_{(n)})$ be the increasing order statistics of n independent log-losses $(L_1, L_2, \ldots, L_n)$ whose common distribution is assumed to satisfy $\mathbb{P}[L_1 \geq x] = C/x^\alpha$ for $x \geq x_0$ where α is the tail index .

For a large enough i,

$$\ln L_{(i)} = -\gamma \ln \left(\frac{n+1-i}{n+1} \right) + K + \epsilon_i \tag{3.6}$$

where $\gamma = 1/\alpha, K = \gamma \ln C$ and ϵ_i is noise. Plotting $\ln L_{(i)}$ as a function of $-\ln\left(\frac{n+1-i}{n+1}\right)$ gives a Pareto plot and γ is the slope of such a plot. After selecting an interval $[d_n, u_n]$, i.e., removing the highest values, the slope of the plot($=\gamma$) can be estimated as:

$$\hat{\gamma} = -\frac{\sum_{i=d_n}^{u_n} \ln(L_{(i)}).\ln\left(\frac{n+1-i}{n+1}\right)}{\sum_{i=d_n}^{u_n}\left(\ln \frac{n+1-i}{n+1}\right)^2} \tag{3.7}$$

The constant C and the threshold x_0 are estimated as in [2]. For $w \in (0,1)$ close to 1, the constant C and the threshold x_0 are estimated by

$$\hat{C} = L_{\lfloor nw \rfloor}^{\hat{\alpha}}(1-w) \text{ and } \hat{x}_0 = L_{\lfloor nw \rfloor} \tag{3.8}$$

The quantile for the log-losses at level $p > w$ used to compute the VaR is approximated by

$$\hat{q}_p^L = \left(\frac{\hat{C}}{1-p}\right)^{\hat{\gamma}} = L_{\lfloor nw \rfloor}\left(\frac{1-w}{1-p}\right)^{\hat{\gamma}} \tag{3.9}$$

Choice of parameters for Pareto estimator

We estimate the power law distribution by setting $d_n = \lfloor 0.95 \times n \rfloor, u_n = \lfloor 0.99 \times n \rfloor$. and $w = 0.90$, where n is the number of log-losses in our time-window, as in [2].

The time window is 1 year($= 252$ days) long. So we can expect roughly half of them to be losses. Since the slope of the Pareto plot is calculated between the 95^{th} and 99^{th} percentile, the model parameters are estimated from very small samples.

3.3.3 Backtesting

The backtesting procedure involves comparing the estimated VaR with the actual number of extreme losses.

Let $(S_0, S_1, \ldots, S_T)$ be the given set of daily stock prices. Let W be the window size and $S_{t:t+W} = (S_t, \ldots, S_{t+W})$ be the series of $W+1$ prices over the time interval $[t, t+W]$. From this we can find a series $R_{t:t+W}$ of log-returns and $L_{t:t+W}$ of log-losses. Using these we can fit our models and estimate the VaR (denoted by $VaR_{t:t+W}$) in each of them. Now we see if the loss incurred the next day$(t+W+1)$ is larger than the estimated VaR. For a good VaR estimator, the proportion of days in which the asset incurs a loss larger than the VaR should be close to $1-\alpha$.

For this purpose we define for $t = 0, \ldots, T-W$,

$$J_{t+W+1} = \begin{cases} 1 & \text{if } S_{t+W+1} - S_{t+W} < -VaR_{t:t+W} \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

As discussed above, the mean $\bar{J}$ for a good VaR estimator would be close to $1-\alpha$. If $\bar{J}$ is much larger than $1-\alpha$, then the model has underestimated the risks, and if $\bar{J}$ is much lower than $1-\alpha$, then the model has overestimated the risks.

In this study, we assume J_t's are independent Bernoulli random variables to compute confidence intervals. Thus $\sum_{t=0}^{T-W} J_{t+W+1}$ follows a binomial distribution. An exact 99% confidence interval of $\bar{J}$ can be computing using the R library binom.

3.4 Discussion of observed results

In this section, we discuss the results observed from using the methodology just described on data from the Indian stock market. The stocks considered in the experiments are those traded on the NSE (National Stock Exchange of India Ltd.) and BSE (BSE Limited). The data was obtained from Yahoo Finance. The stocks were chosen such that data was available throughout the period between January 2010 and September 2020 (over 10 and a half years). We chose 42 such assets from the list of most active stocks posted by Yahoo Finance (as of the start of November 2020) along with the NIFTY50 and SENSEX indices as well (that is, a total of 44 assets). The tickers of the chosen assets on Yahoo Finance can be seen in the x-axis of Figure 3.1.

In the experiments, the time series considered are the log-returns of the Adjusted Closing prices of the 44 assets. Further, we choose the window size, W, to be 252, as it provides a meaningful trade-off between having enough data to see extreme gains/losses and the stationarity assumption. In all of the analysis below, the VaR is the 0.99-VaR.

3.4.1 Analysis

The VaR estimation and backtesting methodology was applied to the 44 assets described. We make the following plots for all assets:

1. A scatter plot of $\bar{J}$.

2. The 99% confidence interval for the $\bar{J}$ estimate.

3. Price and historic volatility (over each window).

4. The relative VaR (VaR/price) over each window, where the price is the observed adjusted closing price on the day after the end of the window.

5. The evolution of the shape parameter, σ for Gaussian and γ for Stable, as estimated over each window.

6. The tail index, α, for Stable and Paretian models, as estimated over each window.

7. The logarithm of the estimated value of C of the tail distribution for Stable and Paretian models.

For a given window $[t, t+W]$, the graphs show the estimated VaR, tail index, etc as estimates for the day $t+W+1$.

Figures 3.1 and 3.2 show the scatter plot detailing the results of the backtesting method and the exact 99% confidence intervals. The x-axis for the confidence interval plot replaces the names of the tickers with serial numbers for convenience.

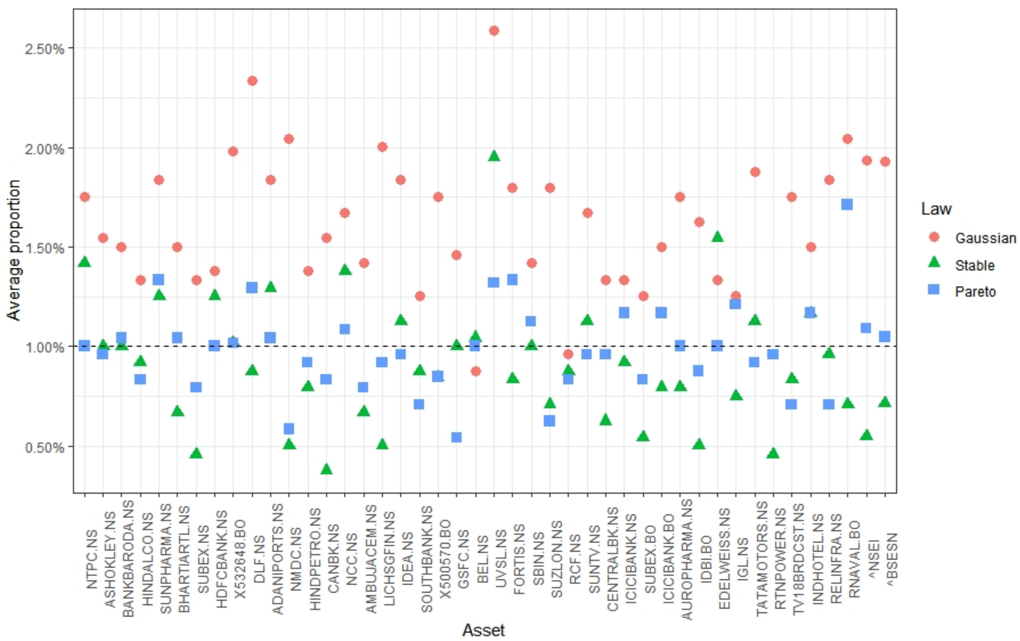

Figure 3.1: Average J

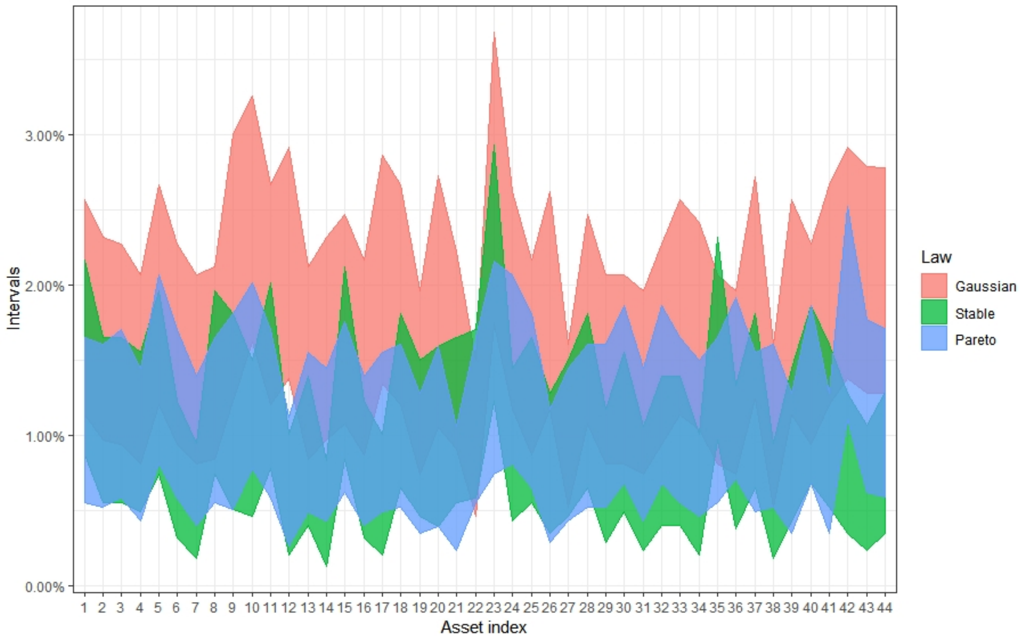

Figure 3.2: Confidence Intervals

One immediately visible fact is that the confidence intervals are quite large. This is because of the small amount of data in the backtest (in each window, that is). As these intervals are large, it is not possible to assert for each asset if the VaR provides a good measure for the loss at the 1% threshold.

However, it is indeed evident that the Gaussian model underestimates VaR. The VaR estimates from the Gaussian law seem to be closer to the 0.98-VaR than the 0.99-VaR (the intervals surround this value). Another fact which is clear from just the confidence intervals is that both the Pareto and Stable models clearly provide a better estimate of VaR than the Gaussian, as their confidence intervals surround 1% better. It can also be observed that the Stable law often overestimates the VaR. In our dataset, for more than half the assets, $\bar{J}$ from the Stable law is less than that from the Paretian law.

Concretely, it was observed that the 1% value was present in the confidence intervals from the Gaussian model for 21 out of the 44 assets under consideration. Whereas, for the Stable model 40 assets contain 1% and for the Paretian model 43 assets contain this value in their respective confidence intervals. Thus, in terms of the backtesting results, the Pareto model seems to outperform the Stable model which outperforms the Gaussian model.

This presents a good case to reject the Gaussian assumption and account for the heaviness of the return distribution's tails. One way that these results are interpreted is that the Gaussian distribution's tails are too thin and underestimates the possibility of extreme losses, the Stable distribution's tails are too heavy (because of the $\alpha \in (0, 2)$ constraint) and often overestimates these possibilities. Indeed, in our dataset, we could observe that the estimated α is almost always greater than 1.5 and is almost never less than 1. The Pareto distribution for the tail allows us to "choose" the heaviness of the tails and thus strike a balance between the Stable and Gaussian laws. However, we shall see that this freedom comes at a cost.

Now, we consider the price, volatility, relative VaR, shape parameter evolution, tail index and C evolution of a few representative assets. In all the following plots, we use the following colour code: all data from Gaussian models are in green, Stable in red and Pareto in blue.

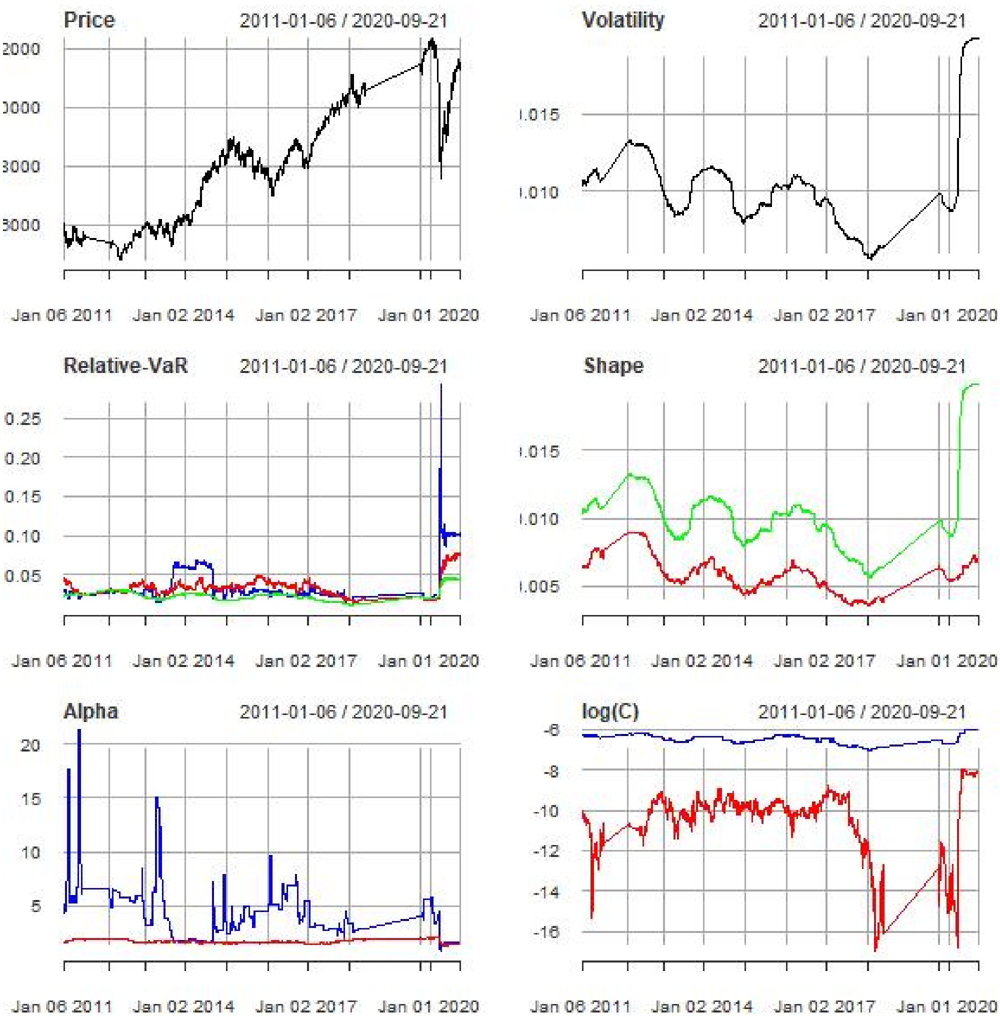

Figure 3.3: Evolution of Parameters for NIFTY50

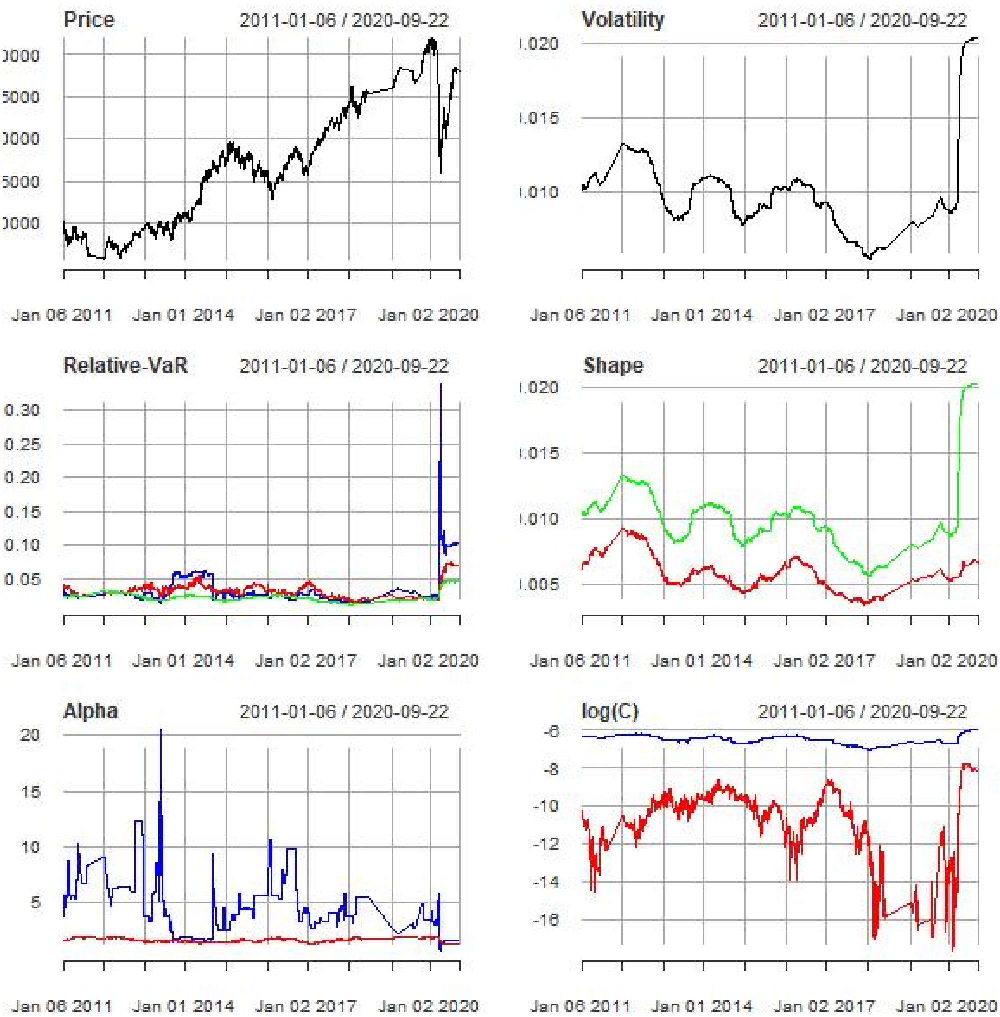

Figure 3.4: Evolution of Parameters for SENSEX

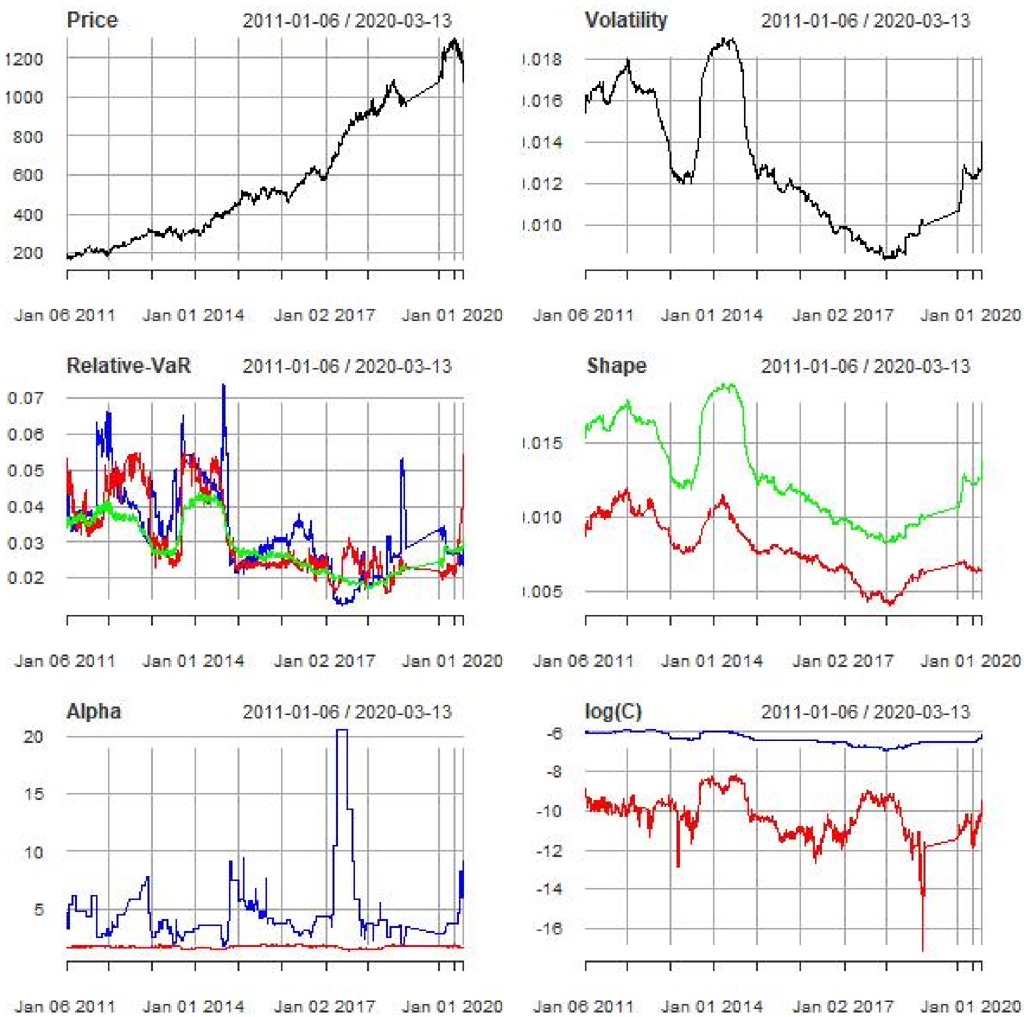

Figure 3.5: Evolution of Parameters for HDFCBANK.NS

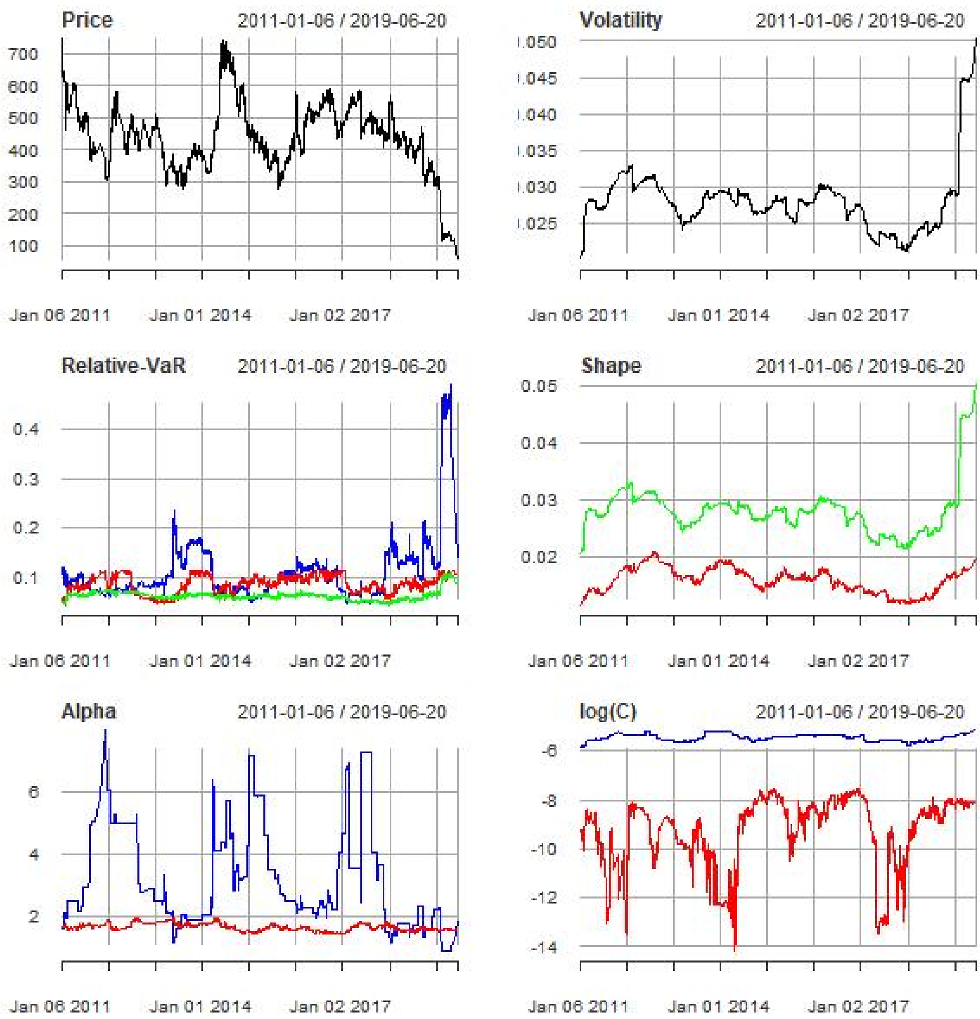

Figure 3.6: Evolution of Parameters for RELINFRA.NS

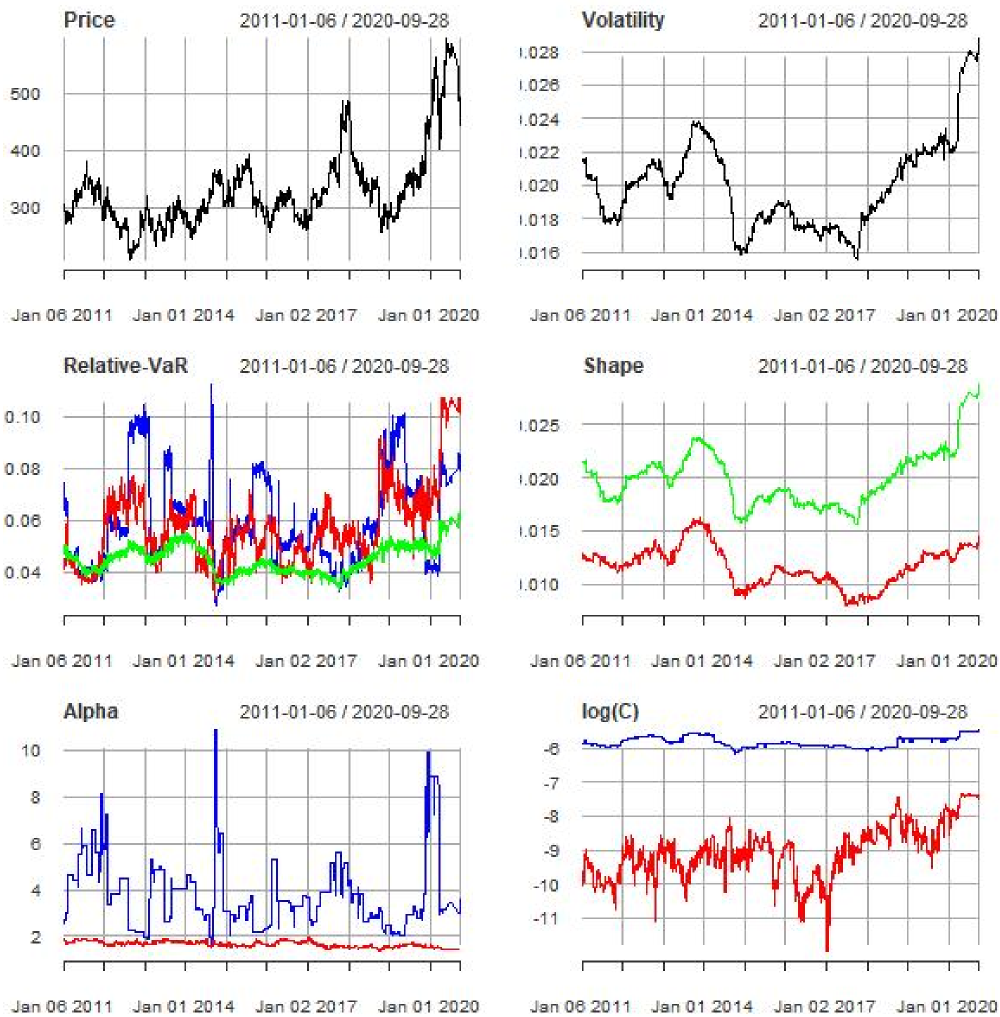

Figure 3.7: Evolution of Parameters for BHARTIARTL.NS

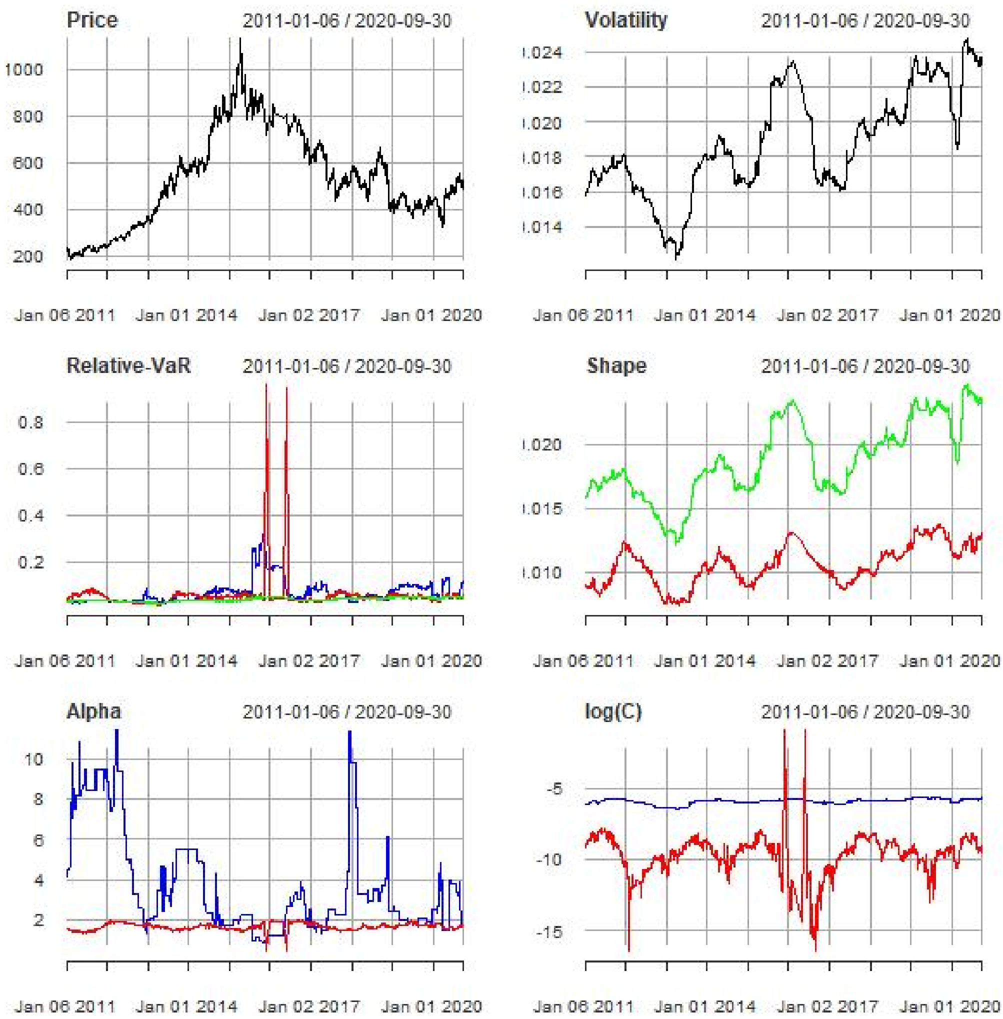

Figure 3.8: Evolution of Parameters for SUNPHARMA.NS

Several facts can be noticed upon looking at the plots for the assets:

1. Observing the volatility plots for several assets, we observe that most of them have peaks around the 2012-2013, early 2014 and 2020 periods, to varying degrees. These peaks may likely correspond to the lagged economic slowdown in 2012-2013, the 2014 general elections and COVID-19 respectively.

2. The shape parameter of the Stable distribution is neatly correlated with the volatility. The evolution of γ is typically shifted downward with respect to the volatility curve by about 1%.

3. The relative VaR estimates from the Gaussian distribution show much less fluctuation than the Stable or Pareto estimates. This is a result of our parameter estimation methods for the non-normal models which are sensitive to extreme values entering and leaving the window.

4. In particular some of the estimates from the Stable and Pareto models are unreliable. SUNPHARMA is a representative example where the Stable estimate is unreliable (notice the sharp spike around 2017) but the Pareto estimate is decent in comparison. In contrast, the Pareto estimates for RELINFRA become unreliable when we near 2020 while the Stable estimates do not.

5. Such erratic spikes were observed most often in the Stable estimates and relatively rarer in the Pareto estimates.

6. It was observed that such anomalous peaks were observed typically during times of high volatility for both the Stable and Paretian models. However, it is difficult to assert that the models actually predict dras-

tic volatility changes because of the estimators' sensitivity to random fluctuations in the data.

7. There are also cases such as HDFCBANK and BHARTIARTL where both the Stable and Pareto estimates show large amount of fluctuation, with the fluctuation increasing during periods of high volatility.

8. For the NIFTY50 and SENSEX indices, both Stable and Pareto models seem to be provide reliable estimates of VaR as can be observed from the plots (except for a brief period in 2020 where the Pareto estimate spikes. This might be an artefact of the sudden increase in the volatility due to COVID-19). In total, there were more cases of the Stable estimates being unreliable (with erratic spikes) than the Pareto estimates.

9. Overall, Pareto models seem to strike the balance between the uncertain, overestimates of the Stable model and the stable, underestimates of the Gaussian model.

10. Further, from the plots of the evolution of α and C, we observe that there is a large amount of fluctuation indicating the presence of considerable uncertainty in our estimation procedure.

11. In particular, for α, we see strong uncertainty in the Pareto estimation whereas we see relatively less in the Stable estimates. This could be explained by the fact that Stable distributions restrict α to $(0, 2)$ but Pareto does not, so there is inherently lesser chance for much fluctuation. In contrast, we see more fluctuation for C in the Stable case relative to Pareto.

12. Also, a key point to notice is that α is typically larger than 2 for the Pareto models and almost always larger than 1.5 for the Stable models. This suggests that the Stable distributions might indeed be too heavy tailed for stock data and strengthens the case for Paretian models.

In conclusion, based on the results of the backtesting and the analysis of the estimates' uncertainty through the plots, the Pareto tail distribution comes out to be the better performer. Even though the uncertainty in the estimation of α prevents us from trusting the VaR estimates with high confidence, it certainly avoids the pitfall of the Gaussian distribution in underestimating risk and that of the Stable distribution of overestimation and erratic spikes.

Chapter 4

Multivariate Distributions

In this chapter, we start by defining the multivariate Gaussian distribution and work our way up to more complex distributions that fit financial data better.

4.1 Multivariate Gaussian

Definition 4.1.1. A random vector $X = (X_1, \ldots, X_d)'$ has a multivariate normal distribution if

$$X \stackrel{d}{=} \mu + AZ \tag{4.1}$$

where $Z \sim \mathcal{N}_k(0, I_k)$, that is, it is a vector of k iid univariate standard normals $Z_1, \ldots, Z_k$, $A \in \mathbb{R}^{d \times k}$ and $\mu \in \mathbb{R}^d$.

From the definition, it is clear that the mean vector $\mathbb{E}[X] = \mu$ and the covariance matrix, $\text{Cov}(X) = \Sigma = AA'$.

The characteristic function of X is given by,

$$\phi_X(t) = \mathbb{E}[\exp(it'X)] = \exp\left(it'\mu - \frac{1}{2}t'\Sigma t\right) \tag{4.2}$$

From this, it is clear that the multivariate Gaussian is completely specified by its mean vector and covariance matrix, hence the notation, $X \sim \mathcal{N}_d\left(\mu, \Sigma\right)$.

Further, X is non-degenerate if and only if $\text{rank}(A) = d \leq k$. In this case, Σ is non-singular and we have the density,

$$f(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(x-\mu\right)' \Sigma^{-1}\left(x-\mu\right)\right) \quad x \in \mathbb{R}^d \tag{4.3}$$

where $|\Sigma|$ is the determinant of Σ. (Whenever not mentioned, we assume $\text{rank}(A) = d$).

Clearly, this density has ellipsoidal contours corresponding to constant values of $(x-\mu)'\Sigma^{-1}(x-\mu)$. That is, the multivariate Gaussian is a special case of the more general class of *Elliptical* distributions.

Other important properties such as invariance to linear transformations, normal marginal and conditional distributions and convolutions are summarized in [4], equations (3.12) to (3.14). It is important to note that all of these properties generalize to the class of Elliptical distributions as well.

4.2 Normal Variance-Mixture Distributions

In this section, we study the normal variance-mixture models which introduce randomness into the covariance of the multivariate normal distribution. This allows for heavier tails of the marginals and tail-dependencies.

Definition 4.2.1. $X = (X_1, \ldots, X_d)'$ has a normal variance-mixture distribution if

$$X \stackrel{d}{=} \mu + \sqrt{W}AZ \tag{4.4}$$

where $\mu \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times k}$, $Z \sim \mathcal{N}_k\left(0, I_k\right)$ and $W \geq 0$ is a scalar valued random variable independent of Z. W is typically called the mixing variable and its distribution, the mixing distribution.

Clearly, we can observe that if $\Sigma = AA'$, then $X|W = w \sim \mathcal{N}_d(\mu, w\Sigma)$. This leads us to the interpretation of Definition 4.2.1 as follows: Consider a family of multivariate normal distributions, each with the same mean vector but with a scaled version of Σ as the covariance matrix. The choice of the scaling factor (and thus, the multivariate normal distribution) is determined by the distribution of W.

This interpretation allows us to sample easily from the normal variance-mixture distributions. We first sample a w from the distribution of W. This leads to a choice of the distribution from the family, that is $\mathcal{N}_d(\mu, w\Sigma)$. We now sample from this multivariate normal using well-known methods. This gives us one realization of X.

4.2.1 Properties

1. Assuming $\mathbb{E}\left[\sqrt{W}\right] < \infty$, the mean vector remains unchanged.

$$\begin{aligned} \mathbb{E}[X] &= \mu + \mathbb{E}\left[\sqrt{W}AZ\right] \\ &= \mu + \mathbb{E}\left[\sqrt{W}\right] A\mathbb{E}[Z] \\ &= \mu \end{aligned} \tag{4.5}$$

   as $\mathbb{E}[Z] = 0$ and W, Z are independent.

2. Assuming $\mathbb{E}[W] < \infty$, the covariance matrix is scaled.

$$\begin{aligned} \mathrm{Cov}(X) &= \mathbb{E}\left[\sqrt{W}AZ\sqrt{W}Z'A'\right] \\ &= \mathbb{E}[W] A\mathbb{E}[ZZ'] A' = \mathbb{E}[W]\Sigma \end{aligned} \tag{4.6}$$

   That is, X has the same correlation matrix as the "unmixed" $\mu + AZ$.

We refer to μ as the *location* vector and Σ as the *dispersion* matrix.

3. Let $X = (X_1, X_2)'$ have a normal variance-mixture distribution with $A = I_2$ and $\mathbb{E}[W] < \infty$ so that $\mathrm{Cov}(X) = \begin{bmatrix} \mathbb{E}[W] & 0 \\ 0 & \mathbb{E}[W] \end{bmatrix}$. That is, X_1 and X_2 are uncorrelated. Then, they are independent if and only if W is almost surely a constant. (WLOG assume $\mu = (0, 0)'$)

Proof. Clearly, if W is almost surely a constant, the result follows from the basic property of uncorrelated normals. For the other direction consider the following.

Suppose X_1 and X_2 are independent. By definition, $X \overset{d}{=} \sqrt{W}AZ$ where $Z = (Z_1, Z_2)$, Z_1, Z_2 are iid standard normals.

$$\begin{aligned}\implies \mathbb{E}\left[|X_1||X_2|\right] &= \mathbb{E}\left[W\right]\mathbb{E}\left[|Z_1|\right]\mathbb{E}\left[|Z_2|\right] \\ &\geq (\mathbb{E}\left[\sqrt{W}\right])^2\mathbb{E}\left[|Z_1|\right]\mathbb{E}\left[|Z_2|\right] = \mathbb{E}\left[|X_1|\right]\mathbb{E}\left[|X_2|\right]\end{aligned}$$

from the Cauchy-Schwartz inequality ($\mathbb{E}[\sqrt{W} \cdot 1] \leq \sqrt{\mathbb{E}[W] \cdot 1}$).

However, as X_1 and X_2 are independent, the equality must hold in the inequality. Again, from Cauchy-Schwartz, this can only happen if $\sqrt{W} = \alpha \cdot 1$ almost surely, for some $\alpha \geq 0$. This establishes that W is constant almost surely. $\square$

This property illustrates that lack of correlation $\implies$ independence if and only if (X_1, X_2) is bivariate normal for the variance-mixture distributions. This suggests that we need a better dependence measure among random variables. This will be addressed in the chapter about Copulas.

4. The characteristic function of a normal variance-mixture is given by,

$$\begin{aligned}
\phi_X(t) &= \mathbb{E}\left[\exp(it'X)\right] \\
&= \mathbb{E}\left[\exp it'\mu + it'\sqrt{W}AZ\right] \\
&= \exp\left(it'\mu\right) \quad \mathbb{E}\left[\mathbb{E}\left[\exp(it'\sqrt{W}AZ) \quad | \quad W\right]\right] \\
&= \exp\left(it'\mu\right) \quad \mathbb{E}\left[\exp(-\frac{1}{2}Wt'\Sigma t)\right] \\
&= \exp\left(it'\mu\right) \hat{F}_W\left(\frac{1}{2}t'\Sigma t\right)
\end{aligned} \tag{4.7}$$

where $\hat{F}_W(s) = \int_0^\infty \exp(-sx)dF_W(x)$ is the Laplace-Stieltjes transform of the cdf of W, F_W. Based on this result, we use the notation $X \sim \mathcal{M}_d\left(\mu, \Sigma, \hat{F}\right)$ for normal variance-mixture distributions.

5. The normal variance-mixtures are elliptical as well. We observe the density f_X given by,

$$f_X(x) = \int_0^\infty f_{X|W}(x|w)dF_W(w)$$

where $f_{X|W}$ is the W-conditional density of X. That is,

$$f_X(x) = \int_0^\infty \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}w^{d/2}} \exp\left(-\frac{(x-\mu)'\,\Sigma^{-1}\,(x-\mu)}{2w}\right) dF_W(w) \tag{4.8}$$

which clearly is a constant whenever $(x-\mu)'\,\Sigma^{-1}\,(x-\mu)$ is a constant. That is, it has elliptsoidal contours.

6. Let $X \sim \mathcal{M}_d\left(\mu, \Sigma, \hat{F}\right)$ and $Y = BX + b$ where $B \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^m$. Then $Y \sim \mathcal{M}_d\left(B\mu + b, B\Sigma B', \hat{F}\right)$. That is, the class of normal

variance-mixture distributions is closed under linear transformations.

Proof. The characteristic function of Y is given by,

$$\begin{aligned}\phi_Y(t) &= \mathbb{E}\left[\exp\left(it'(BX+b)\right)\right] \\ &= \exp\left(it'b\right)\phi_X(B't) \\ &= \exp\left(it'b\right)\exp\left(it'B\mu\right)\hat{F}_W\left(\frac{1}{2}t'B\Sigma B't\right) \\ &= \exp\left(it'\left(B\mu+b\right)\right)\hat{F}_W\left(\frac{1}{2}t'\left(B\Sigma B'\right)t\right)\end{aligned}$$

which completes the proof. □

In particular, the univariate marginals are given by,

$$X_i = e_i'X \sim \mathcal{M}_1\left(\mu_i, \Sigma_{ii}, \hat{F}\right) \tag{4.9}$$

4.2.2 Examples

Example 4.2.1. (*n-Point Mixtures*) Let W be a discrete random variable taking values in $\{k_1, \ldots, k_n\}$ with probabilities $\{p_1, \ldots, p_n\}$. This choice of W can be interpreted as dividing the outcomes into n different regimes. Regimes with large k are seen as stressed as they correspond to high variance and those with relatively smaller values are seen as ordinary regimes.

Example 4.2.2. (*Multivariate Student-t distribution*) If $W \sim \mathrm{Ig}\left(\frac{1}{2}\nu, \frac{1}{2}\nu\right)$ then, we say that X follows the Multivariate Student-t distribution.

In general, $Y \sim \mathrm{Ig}\left(\alpha, \beta\right)$ if it has density,

$$f(y) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}y^{-(\alpha+1)}\exp\left(-\frac{\beta}{y}\right)$$

This is called the Inverse Gamma Distribution.

Substituting this density into (4.8), we get the density of the multivariate Student-t as,

$$f(x) = \frac{\Gamma(\frac{1}{2}(\nu + d))}{\Gamma(\frac{1}{2}\nu)(\pi\nu)^{d/2}|\Sigma|^{1/2}} \left(1 + \frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{\nu}\right)^{-\frac{1}{2}(\nu+d)} \qquad x \in \mathbb{R}^d \tag{4.10}$$

Because of this form, we denote this distribution by $t_d(\nu, \mu, \Sigma)$. In later sections, we will see that the marginals of this distribution have heavier tails than the Gaussian and also that it has more tendency to generate joint extreme values.

Example 4.2.3. (*Symmetric Generalized Hyperbolic distribution*) This popular distribution is obtained by choosing $W \sim \mathcal{N}^{-}(\lambda, \chi, \psi)$, which is the Generalized Inverse Gaussian distribution (GIG).

GIG has the density,

$$f_W(x) = \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}\right)^{\lambda}}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\frac{\chi}{x} + \psi x\right)\right) \tag{4.11}$$

where, the normalizing constant involves K_λ, a modified Bessel function of the third kind with index λ. The parameters must satisfy,

$$\lambda < 0 \implies \chi > 0, \psi \geq 0$$

$$\lambda = 0 \implies \chi > 0, \psi > 0$$

$$\lambda > 0 \implies \chi \geq 0, \psi > 0$$

Extensive details of the GIG can be found in [9].

We now cite the following important results from appendix A.2.5 of [4].

1. $K_\lambda(x) = K_{-\lambda}(x)$.

2. $K_\lambda(x) \sim \Gamma(\lambda) 2^{\lambda-1} x^{-\lambda}$ as $x \to 0^+$ for $\lambda > 0$.

3. $K_\lambda(x) \sim \Gamma(-\lambda) 2^{-\lambda-1} x^{\lambda}$ as $x \to 0^+$ for $\lambda < 0$.

Whenever $\chi > 0, \psi > 0$, we have

$$\mathbb{E}\left[X^\alpha\right] = \left(\frac{\chi}{\psi}\right)^{\alpha/2} \frac{K_{\lambda+\alpha}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} \quad \alpha \in \mathbb{R} \tag{4.12}$$

$$\mathbb{E}\left[\ln(X)\right] = \left.\frac{d\mathbb{E}\left[X^\alpha\right]}{d\alpha}\right|_{\alpha=0} \tag{4.13}$$

So for $\lambda > 0$,

$$f_W(x) = \frac{\left(\frac{\psi}{2}\right)^\lambda}{\Gamma(\lambda)} \exp\left(-\frac{\psi}{2}x\right)$$

as $\chi \to 0^+$.

This is exactly the density of Gamma $\left(\frac{\psi}{2}, \lambda\right)$. Similarly, when $\lambda < 0$ and $\psi \to 0^+$, we get the density of Ig $\left(-\lambda, \frac{\chi}{2}\right)$. So, the Gamma and Inverse Gamma distributions are both special cases of the GIG. This means that the multivariate student-t is a special case of the Symmetric Generalized Hyperbolic distributions as well.

The normal variance-mixture distributions from such a choice of W has a density of the form (by substituting (4.11) into (4.8)),

$$f(x) = \frac{\left(\sqrt{\chi\psi}\right)^{-\lambda}\left(\psi^{d/2}\right)}{(2\pi)^{d/2}|\Sigma|^{d/2}K_\lambda(\sqrt{\chi\psi})} \frac{K_{\lambda-(d/2)}\left(\sqrt{(\chi + (x-\mu)'\Sigma^{-1}(x-\mu))\,\psi}\right)}{\left(\sqrt{(\chi + (x-\mu)'\Sigma^{-1}(x-\mu))\,\psi}\right)^{(d/2)-\lambda}} \tag{4.14}$$

The mixture density corresponding to the special case $\lambda = -\frac{1}{2}$ is called the symmetric Normal Inverse Gaussian distribution (NIG). The one-dimensional marginals of the case $\lambda = 1$ are called the hyperbolic distributions.

4.3 Normal Mean-Variance Mixture Distributions

The normal variance-mixture distributions preserved the elliptical symmetry of the multivariate Gaussian distribution. However, this symmetry may not exist in financial return data as it is sometimes observed that the right-tail (gains) is lighter than the left-tail (losses). We can attempt to capture this asymmetry by introducing randomness into the mean of the Gaussian distribution as well.

Definition 4.3.1. $X = (X_1, \ldots, X_d)'$ has a normal mean-variance mixture distribution if,

$$X \stackrel{d}{=} m(W) + \sqrt{W} AZ \tag{4.15}$$

where, $Z \sim \mathcal{N}_k(0, I_k)$, $W \geq 0$ is a scalar valued random variable independent of Z, $A \in \mathbb{R}^{d \times k}$ and $m : [0, \infty) \to \mathbb{R}^d$ is a measurable function. Again, we typically consider cases where $\text{rank}(A) = d \leq k$.

Let $\Sigma = AA'$. In contrast to the variance mixtures, $X|W = w \sim \mathcal{N}_d\left(m(w), w\Sigma\right)$. Similar to the variance mixtures we interpret these variables as follows: Consider a family of multivariate normal distributions, each associated with a value $w \geq 0$ which determines its mean and the scaling factor for Σ. The choice of this w is determined by the distribution of W.

This allows us to sample easily from the mean-variance distributions. We first sample a w from the distribution of W. This leads to a choice of the distribution $\mathcal{N}_d\left(m(w), w\Sigma\right)$. We now sample from this multivariate normal distribution using well-known methods. This gives us one realization of X.

A common specification for m is,

$$m(W) = \mu + W\gamma \tag{4.16}$$

where $\mu, \gamma \in \mathbb{R}^d$ are parameter vectors.

4.3.1 Properties

Throughout the properties, we assume m takes the form in (4.16).

1. Assuming $\mathbb{E}[W] < \infty$,

$$\mathbb{E}[X] = \mathbb{E}[m(W)] = \mu + \mathbb{E}[W]\gamma \tag{4.17}$$

2. Assuming $\mathbb{E}[W^2] < \infty$, the covariance matrix is given by,

$$\begin{aligned} \mathrm{Cov}(X) &= \mathbb{E}[\mathrm{Cov}(X|W)] + \mathrm{Cov}(\mathbb{E}[X|W]) \\ &= \mathbb{E}[W]\Sigma + \mathrm{Cov}(\mu + W\gamma) \\ &= \mathbb{E}[W]\Sigma + Var(W)\gamma\gamma' \end{aligned} \tag{4.18}$$

   That is, neither the mean nor the covariance (and correlation) matrix of X is identical to the "unmixed" multivariate normal unless $\gamma = 0$.

3. For $x \in \mathbb{R}^d$, the density f_X is given by,

$$f_X(x) = \int_0^\infty \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}w^{d/2}} \exp\left(-\frac{(x - m(w))'\Sigma^{-1}(x - m(w))}{2w}\right) dF_W(w) \tag{4.19}$$

   using the same idea as in (4.8).

   In particular, for the choice of m as in (4.16), the density is,

$$f_X(x) = \int_0^\infty \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}w^{d/2}} \exp\left(-\frac{(x-\mu-w\gamma)'\,\Sigma^{-1}\,(x-\mu-w\gamma)}{2w}\right) dF_W(w) \tag{4.20}$$

On simplification, this yields,

$$f_X(x) = \int_0^\infty \frac{e^{\left((x-\mu)'\Sigma^{-1}\gamma\right)}}{(2\pi)^{d/2}|\Sigma|^{1/2}w^{d/2}} \exp\left(-\frac{(x-\mu)'\,\Sigma^{-1}\,(x-\mu)}{2w} - \frac{w\gamma'\Sigma^{-1}\gamma}{2}\right) dF_W(w) \tag{4.21}$$

4. The characteristic function is given by,

$$\begin{aligned} \phi_X(t) &= \mathbb{E}\left[\exp\left(it'\left(\mu + W\gamma + \sqrt{W}AZ\right)\right)\right] \\ &= \exp\left(it'\mu\right)\hat{F}_W(\frac{1}{2}t'\Sigma t - it'\gamma) \end{aligned} \tag{4.22}$$

using the same argument from (4.7).

5. These distributions are closed under linear transformations as well (for the specific choice of m under consideration).

Let $X = \mu + W\gamma + \sqrt{W}AZ$ and $Y = BX + b = B\mu + WB\gamma + \sqrt{W}BAZ + b$. Clearly, the following transformations have occurred,

$$\begin{aligned} \mu &\to B\mu + b \\ \gamma &\to B\gamma \\ A &\to BA \end{aligned}$$

The last transformation could be rewritten as $\Sigma \to B\Sigma B'$ as well. Arguments similar to Property (6) of variance-mixture distributions could be applied to obtain the characteristic function of Y.

6. It can be shown that X is infinitely divisible if and only if W is infinitely divisible (refer [5], XIII).

4.3.2 Examples

Example 4.3.1. (*Generalized Hyperbolic Distribution*) In Example 4.2.3, we saw a special class of the Generalized Hyperbolic distributions. We can generate the full class using the mean-variance mixing idea. Here, we use the form of (4.16) for m and the GIG as the mixing distribution. That is, $W \sim \mathcal{N}^{-}(\lambda, \chi, \psi)$.

Plugging the density (4.11) into (4.21) gives us the density of this class of distributions,

$$f(x) = c\frac{K_{\lambda-(d/2)}\left(\sqrt{(\chi+(x-\mu)'\Sigma^{-1}(x-\mu))\,(\psi+\gamma'\Sigma^{-1}\gamma)}\right)e^{\left((x-\mu)'\Sigma^{-1}\gamma\right)}}{\left(\sqrt{(\chi+(x-\mu)'\Sigma^{-1}(x-\mu))\,(\psi+\gamma'\Sigma^{-1}\gamma)}\right)^{(d/2)-\lambda}} \tag{4.23}$$

where the constant c is given by,

$$c = \frac{\left(\sqrt{\chi\psi}\right)^{-\lambda}\psi^{\lambda}(\psi+\gamma'\Sigma^{-1}\gamma)^{d/2-\lambda}}{(2\pi)^{d/2}|\Sigma|^{d/2}K_{\lambda}(\sqrt{\chi\psi})} \tag{4.24}$$

We denote this distribution by $GH_d\,(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$.

We can observe that the density differs from (4.14) only by shifts in ψ by $\gamma'\Sigma^{-1}\gamma$. That is, if $\gamma = 0$ then we revert to the Symmetric Generalized Hyperbolic distribution. However, in general, this density defines a non-elliptical distribution with asymmetric marginals.

Using (4.12) in (4.17) and (4.18), we obtain the mean vector and covariance matrix for this distribution as,

$$\mathbb{E}[X] = \mu + \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}\gamma \tag{4.25}$$

$$\mathrm{Cov}(X) = c_m\Sigma + c_v\gamma\gamma' \tag{4.26}$$

where,

$$c_m = \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}$$

and

$$c_v = \left(\frac{\chi}{\psi}\right)\left(\frac{K_{\lambda+2}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})} - \frac{K^2_{\lambda+1}(\sqrt{\chi\psi})}{K^2_{\lambda}(\sqrt{\chi\psi})}\right)$$

It can be shown that GIG is an infinitely divisible distribution (refer [6]). Using Property (6), this shows that the generalized hyperbolic distributions are infinitely divisible as well.

It is important to notice from (4.23) and (4.24) that the density of $GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ is identical to that of $GH_d(\lambda, \chi/k, k\psi, \mu, k\Sigma, k\gamma)$ for any $k > 0$. This creates an identifiability problem while estimating parameters for such distributions. Typically we overcome this problem by forcing $|\Sigma| = 1$. While this would alter the values of χ, ψ obtained, the value of $\chi\psi$ can be used as a summary parameter.

From Property (5) of mean-variance mixtures, we can see that if $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$, then $Y = BX + b \sim GH_d(\lambda, \chi, \psi, B\mu + b, B\Sigma B', B\gamma)$.

In particular, the univariate marginals are given by,

$$X_i = e_i'X \sim GH_1(\lambda, \chi, \psi, \mu_i, \Sigma_{ii}, \gamma_i) \tag{4.27}$$

Example 4.3.2. (*d-dimensional Hyperbolic Distributions*) These distributions correspond to the class $GH_d\left(\frac{d+1}{2}, \chi, \psi, \mu, \Sigma, \gamma\right)$. Note that by (4.27), the marginals also have $\lambda = \frac{d+1}{2}$ and thus are not one-dimensional hyperbolic distributions.

(Only in the case $\lambda = 1$, we get one-dimensional hyperbolic marginals.)

Example 4.3.3. (*Normal Inverse Gaussian (NIG)*) These distributions correspond to the class $GH_d\left(-\frac{1}{2}, \chi, \psi, \mu, \Sigma, \gamma\right)$

Example 4.3.4. (*Generalized Laplace or Variance-Gamma (VG)*) We obtain the generalized Laplace distribution in the limit $\chi \to 0^+$ for $\lambda > 0$ in (4.23) and (4.24). Using the limit properties from Example 4.2.3, we get,

$$f(x) = c\frac{K_{\lambda-(d/2)}\left(\sqrt{Q(x)\left(\psi + \gamma'\Sigma^{-1}\gamma\right)}\right)e^{\left((x-\mu)'\Sigma^{-1}\gamma\right)}}{\left(\sqrt{Q(x)\left(\psi + \gamma'\Sigma^{-1}\gamma\right)}\right)^{(d/2)-\lambda}} \tag{4.28}$$

Where,

$$c = \frac{\psi^{\lambda}(\psi + \gamma'\Sigma^{-1}\gamma)^{d/2-\lambda}}{(2\pi)^{d/2}|\Sigma|^{d/2}\Gamma(\lambda)2^{\lambda-1}} \tag{4.29}$$

and

$$Q(x) = (x-\mu)'\Sigma^{-1}(x-\mu) \tag{4.30}$$

Example 4.3.5. (*Skewed-t distribution*) This corresponds to the limit $\psi \to 0^+$ for $\lambda = -\frac{1}{2}\nu$, $\chi = \nu > 0$. Similar to the last example, we can get the density of this distribution as,

$$f(x) = c\frac{K_{(\nu+d)/2}\left(\sqrt{(\nu + Q(x))\gamma'\Sigma^{-1}\gamma}\right)e^{\left((x-\mu)'\Sigma^{-1}\gamma\right)}}{\left(\sqrt{(\nu + Q(x))\gamma'\Sigma^{-1}\gamma}\right)^{-(\nu+d)/2}\left(1 + \frac{Q(x)}{\nu}\right)^{(\nu+d)/2}} \tag{4.31}$$

Where,

$$c = \frac{2^{1-(\nu+d)/2}}{\Gamma(\frac{1}{2}\nu)(\pi\nu)^{d/2}|\Sigma|^{1/2}} \tag{4.32}$$

and $Q(x)$ is as in (4.30). It can be seen using the limit properties that this density converges to the multivariate Student-t density as $\gamma \to 0$.

4.4 General Spherical and Elliptical Distributions

In this section we present a brief overview of generalized classes of spherical and elliptical distributions, whose properties we use in the next section. The previous sections discussed the properties of specific members of these classes.

4.4.1 Spherical Distributions

Definition 4.4.1. The random vector $X = (X_1, \ldots, X_d)'$ has a spherical distribution if for every orthogonal matrix $U \in \mathbb{R}^{d \times d}$ it satisfies,

$$UX \stackrel{d}{=} X \tag{4.33}$$

That is, X is invariant under rotations and is spherically symmetric.

Theorem 4.4.1. The following statements are equivalent:

1. X is spherically distributed.
2. There exists a scalar function ψ such that,

$$\phi_X(t) = \mathbb{E}\left[\exp\left(it'X\right)\right] = \psi(t't) = \psi(\|t\|_2^2) \tag{4.34}$$

3. For all $a \in \mathbb{R}^d$,

$$a'X \stackrel{d}{=} \|a\|_2 X_1 \tag{4.35}$$

Proof. (1) $\implies$ (2):

Because of (4.33), for any orthogonal matrix U, we have:

$$\phi_X(t) = \phi_{UX}(t) = \mathbb{E}\left[\exp\left(it'UX\right)\right] = \phi_X(U't) \tag{4.36}$$

For this to be true for any orthogonal matrix U, it must be true for every vector $\hat{t} \in \mathbb{R}^d$ which is a rotation of t with the same length. That is, ϕ_X has to only depend on $\|t\|_2$. That is, there exists a scalar function ψ such that,

$$\phi_X(t) = \psi(t't) = \psi(\|t\|_2^2) \tag{4.37}$$

(2) $\implies$ (3):

First of all, note that for any $t \in \mathbb{R}$ and $b \in \mathbb{R}^d$,

$$\phi_{b'X}(t) = \mathbb{E}\left[\exp\left(itb'X\right)\right] = \phi_X(tb) = \psi(t^2\|b\|_2^2) \tag{4.38}$$

Now, we have $\|a\|_2 X_1 = \|a\|_2 e_1' X$. So, using $b = \|a\|_2 e_1$ in (1.38) we get,

$$\phi_{\|a\|_2 X_1}(t) = \psi(t^2\|a\|_2^2) = \phi_{a'X}(t) \tag{4.39}$$

where the last equality follows from using $b = a$ in (4.38). Also, note that the choice of X_1 was arbitrary. Using the same arguments we may show that for any $i \in \{1, \ldots, d\}$, $a'X \stackrel{d}{=} \|a\|_2 X_i$.

(3) $\implies$ (1):

Let $U \in \mathbb{R}^{d\times d}$ be an orthogonal matrix. Also, we have the relation,

$$\phi_{UX}(t) = \phi_X(U't) = \mathbb{E}\left[\exp\left(i(U't)'X\right)\right] \tag{4.40}$$

By (3), we have that $(U't)'X \stackrel{d}{=} \|U't\|_2 X_1 = \|t\|_2 X_1$ as U is orthogonal.

However, by (3) we also have $t'X \stackrel{d}{=} \|t\|_2 X_1$. So, $(U't)'X \stackrel{d}{=} t'X$. This reduces (4.40) to,

$$\phi_{UX}(t) = \mathbb{E}\left[\exp\left(it'X\right)\right] = \phi_X(t) \tag{4.41}$$

This completes the proof. $\square$

The function ψ in point (2) of the above theorem is called the *characteristic generator* as it completely determines the characteristic function of the spherical distribution. So, we adopt the notation $X \sim S_d\left(\psi\right)$. For example, if $X \sim \mathcal{N}(0, I_d)$, then $\psi(t) = \exp\left(-\frac{1}{2}t\right)$.

With the help of the inverse Fourier transform and Theorem 4.4.1, we can also infer that the density of $X \sim S_d\left(\psi\right)$ has spherical contours. That is, there exists a scalar function g, called the *density generator*, such that,

$$f(x) = g(x'x) \quad x \in \mathbb{R}^d \tag{4.42}$$

The following theorem provides another useful characterization of the spherical distributions.

Theorem 4.4.2. X has a spherical distribution if and only if there exists a *stochastic representation* given by,

$$X \stackrel{d}{=} RS \tag{4.43}$$

where $R \geq 0$ is a scalar random variable (called the radial random variable) independent of S, which is uniformly distributed on the unit d-sphere, $\mathcal{S}^{d-1} = \{s \in \mathbb{R}^d : s's = 1\}$.

Proof. ($\Longleftarrow$) We show that RS is spherically distributed. Clearly, S is spherically distributed. Let its charaacteristic generator be Ω_d. So,

$$\begin{aligned}
\phi_{RS}(t) &= \mathbb{E}\left[\exp\left(it'RS\right)\right] = \mathbb{E}\left[\mathbb{E}\left[\exp\left(iRt'S\right)|R\right]\right] \\
&= \mathbb{E}\left[\Omega_d\left(R^2t't\right)\right] \\
&= \int_0^\infty \Omega_d\left(r^2t't\right)dF_R(r)
\end{aligned} \tag{4.44}$$

where F_R is the cdf of R. Clearly, ϕ_{RS} depends only on $t't$. So, RS is spherically distributed.

($\Longrightarrow$) Let $X \sim S_d\left(\psi\right)$. For any $t \in \mathbb{R}^d$ and $s \in \mathcal{S}^{d-1}$, we have:

$$\phi_X(t) = \phi_X(\|t\|_2 s) = \psi(t't) \tag{4.45}$$

Let S be uniformly distributed over $\mathcal{S}^{d-1}$. Clearly,

$$\begin{aligned}
\psi(t't) = \int_{\mathcal{S}^d}\psi(t't)dF_S(s) &= \int_{\mathcal{S}^d}\phi_X(\|t\|_2 s)dF_S(s) \\
&= \int_{\mathcal{S}^d}\mathbb{E}\left[\exp\left(i\|t\|_2 s'X\right)\right]dF_S(s) \\
&= \mathbb{E}\left[\Omega_d\left(\|t\|_2^2\|X\|_2^2\right)\right] \\
&= \int_0^\infty \Omega_d\left(\|t\|_2^2 r^2\right)dF_{\|X\|_2}
\end{aligned} \tag{4.46}$$

where the second last equality follows from Fubini's theorem and the last equality by definition ($F_{\|X\|_2}$ is the cdf of $\|X\|_2$). This is exactly in the form of (4.44) where R is replaced with $\|X\|_2$, proving our claim. $\square$

Simply put, this theorem tells us that any spherical distribution can be thought of as a uniform distribution over the unit sphere with randomized length.

Concretely, if $X \sim S_d\left(\psi\right)$, $\mathbb{P}\left(X = 0\right) = 0$ and $X \stackrel{d}{=} RS$, then

$$\left(\|X\|_2, \frac{X}{\|X\|_2}\right) \stackrel{d}{=} (R, S) \tag{4.47}$$

Example 4.4.1. If $X \sim \mathcal{M}_d\left(0, I_d, \hat{F}\right)$, by (4.7) we see that $\phi_X(t) = \hat{F}(\frac{1}{2}t't) := \psi(t't)$. That is, the normal-variance mixture distributions are spherically distributed whenever they have 0 mean vector and uncorrelated marginals.

4.4.2 Elliptical Distributions

This class of distributions generalizes the spherical distributions through their affine transformations.

Definition 4.4.2. The random vector X has an elliptical distribution if it takes the representation,

$$X \stackrel{d}{=} \mu + AY \tag{4.48}$$

where $Y \sim S_k(\psi)$, $\mu \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d\times k}$.

The characteristic function for such variables is given by,

$$\phi_X(t) = \exp(it'\mu)\mathbb{E}\left[\exp\left(it'AY\right)\right] = \exp(it'\mu)\psi\left(t'\Sigma t\right) \tag{4.49}$$

where $\Sigma = AA'$. So, we denote $X \sim E_d(\mu, \Sigma, \psi)$.

Clearly, if $X \sim E_d(\mu, \Sigma, \psi)$, as a direct consequence of Theorem 4.4.2, it admits the representation,

$$X \stackrel{d}{=} \mu + RAS \tag{4.50}$$

where R and S are as in Theorem 4.4.2.

In the case where $A \in \mathbb{R}^{d\times d}$ and invertible, we get the following relation from (4.50),

$$A^{-1}(X-\mu) \stackrel{d}{=} RS \sim S_d(\psi) \tag{4.51}$$

From (4.47), we get,

$$\left(\sqrt{(X-\mu)'\Sigma^{-1}(X-\mu)}, \frac{A^{-1}(X-\mu)}{\sqrt{(X-\mu)'\Sigma^{-1}(X-\mu)}}\right) \stackrel{d}{=} (R,S) \tag{4.52}$$

In (4.48), if Y has density generator g, then from (4.51) we can infer that the density of X takes the form (note that $|A| = |\Sigma|^{1/2}$),

$$f(x) = \frac{1}{|\Sigma|^{1/2}} g\left((x-\mu)'\Sigma^{-1}(x-\mu)\right) \tag{4.53}$$

Properties that generalize the linear transformation, conditional and marginal attributes of the multivariate Gaussian to the full class of elliptical distributions are given in [4], Section 3.3.3.

4.5 Tail Behaviour

In this section we investigate the tail behavior of some of the distributions discussed so far. We see that the tails of the marginals of the normal mixtures are directly connected to the nature of the mixing distribution. We start by citing the following important theorem proved in [3].

Theorem 4.5.1. Let the real-valued random variable X be given by $X = YZ$, where Y and Z are independent and non-negative random variables such that,

1. Y has a regularly varying tail with tail index α. That is,

$$\mathbb{P}(Y > x) \sim x^{-\alpha} L(x), \quad x \to \infty \tag{4.54}$$

   where L is a slowly varying function.

2. $\mathbb{E}\left[Z^{\alpha+\varepsilon}\right] < \infty$ for some $\varepsilon > 0$.

Then, X has a regularly varying tail with tail index α such that,

$$\mathbb{P}(X > x) \sim \mathbb{E}\left[Z^{\alpha}\right] \mathbb{P}(Y > x), \quad x \to \infty \tag{4.55}$$

We can use this to check that a given spherical distribution has marginals with power-law tails. To this end, we state the following theorem.

Theorem 4.5.2. Let $X \overset{d}{=} RS \sim S_d(\psi)$ be a spherical variable. If R has a regularly varying tail with index α, then so does $|X_i|, i = 1, \ldots, d$. If $\mathbb{E}\left[R^k\right] < \infty$ for all $k > 0$ then $|X_i|$ does not have regularly varying tails.

Proof. Let R have a regularly varying tail with index α. By definition, $X_i = RS_i$. Clearly, $|S_i|$ is a non-negative random variable over a finite support $[0, 1]$ with finite moments. That is, R satisfies point (1) and $|S_i|$ satisfies point (2) of Theorem 4.5.1. So, $|X_i| = R|S_i|$ has a regularly varying tail with index α.

If R is such that $\mathbb{E}\left[R^k\right] < \infty$ for all $k > 0$, then all moments of $X_i = RS_i$ are finite, so X_i cannot have regularly varying tails for any $i = 1, \ldots, d$. $\square$

This theorem allows us to analyze the spherical distributions generated by normal variance-mixtures. That is distributions of the form $\mathcal{M}_d\left(0, I_d, \hat{F}_W\right)$.

Theorem 4.5.3. Let $X \sim \mathcal{M}_d\left(0, I_d, \hat{F}_W\right)$, that is, $X \stackrel{d}{=} \sqrt{W}Z$ where W is the mixing variable and $Z \sim \mathcal{N}(0, I_d)$. If W has regularly varying tails with index α, then $|X_i|$ has regularly varying tails with index 2α for $i = 1, \ldots, d$.

Further, if $\mathbb{E}\left[W^k\right] < \infty$ for all $k > 0$, then $|X_i|$ is not regularly varying for any $i = 1, \ldots, d$.

Proof. Clearly, Z is spherical with the representation $Z \stackrel{d}{=} \tilde{R}S$ where $\|Z\|_2^2 \stackrel{d}{=} \tilde{R}^2 \sim \chi_d^2$ (using 4.47) and S is uniformly distributed over $\mathcal{S}^{d-1}$. It is known that if $Y \sim \chi_d^2$, for all $k > 0$, $\mathbb{E}\left[Y^k\right] < \infty$. So we get,

$$\mathbb{E}\left[\tilde{R}^k\right] = \mathbb{E}\left[\left(\tilde{R}^2\right)^{k/2}\right] < \infty \quad \forall k > 0 \tag{4.56}$$

If W has regularly varying tails with index α, that is, for large x, $\mathbb{P}(W > x) \sim x^{-\alpha}L_1(x)$, then,

$$\mathbb{P}\left(\sqrt{W} > x\right) = \mathbb{P}\left(W > x^2\right) \sim x^{-2\alpha}L_1(x^2) := x^{-2\alpha}L_2(x) \tag{4.57}$$

which shows that $\sqrt{W}$ has regularly varying tails with index 2α.

Combining (4.56) and (4.57), from Theorem 4.5.1, we get that $\sqrt{W}\tilde{R}$ is regularly varying with index 2α. As $X = \left(\sqrt{W}\tilde{R}\right)S$, by Theorem 4.5.2, $|X_i|$ for $i = 1, \ldots, d$ has regularly varying tails with index 2α.

If all moments of W are finite, then all moments of $\sqrt{W}\tilde{R}$ are finite as well, so by Theorem 4.5.2, none of $|X_i|$ can have regularly varying tails.

$\square$

We can now consider the distributions obtained in section 1.2 and examine if they have power-law tails.

Example 4.5.1. The multivariate student-t distribution was obtained in Example 4.2.2 as a normal variance-mixture with $W \sim \mathrm{Ig}\left(\frac{1}{2}\nu, \frac{1}{2}\nu\right)$ whose density is given by,

$$f(x) = cx^{-(\frac{1}{2}\nu+1)} \exp\left(-\frac{\nu}{2x}\right) \tag{4.58}$$

where $c = \frac{(\sqrt{\frac{\nu}{2}})^{\nu}}{\Gamma(\frac{1}{2}\nu)}$. Also, the function $L(x) = c\exp\left(-\frac{\nu}{2x}\right)$ is slowly varying as $\lim_{x\to\infty} \exp\left(-\frac{\nu}{2x}\right) = 1$.

So, we get the tail probability for large x,

$$\begin{aligned}
\mathbb{P}(W > x) &= \int_x^\infty f(x)dx \\
&= \int_x^\infty x^{-(\frac{1}{2}\nu+1)} L(x)dx \\
&= \frac{2}{\nu} x^{-\frac{1}{2}\nu} L(x) \\
&:= x^{-\frac{1}{2}\nu} L_1(x)
\end{aligned} \tag{4.59}$$

where the second last equality follows from Karamata's theorem (refer [4], A.1.3) and the last equality follows by defining $L_1(x) = \frac{2}{\nu}L(x)$.

So, the multivariate student-t distribution has marginals with power-law tails with index $2 \times \frac{1}{2}\nu = \nu$ by Theorem 4.5.3.

Example 4.5.2. For the symmetric generalized hyperbolic distributions (Example 4.2.3) that are spherical (that is, the class, $GH_d(\lambda, \chi, \psi, 0, I_d, 0)$, assuming $\chi > 0, \psi > 0$), we see from equation (4.12) that W does not have regularly varying tails as all of its moments are finite. So, by Theorem 4.5.3, this class of distributions does not have marginals with power-law tails.

However, the tails of this class of distributions are seen to be heavier than that of the Gaussian distribution but lighter than power-law. We have the following asymptotic relation (refer [1]), for the density of the marginals obtained using (4.9) and (4.14) for the symmetric models,

$$f(x) \sim x^{\lambda-1} \exp\left(-\sqrt{\psi}x\right), \quad x \to \infty \tag{4.60}$$

This tail-behaviour is seen as "semi-heavy" in the sense that it shows power-law decay with exponential "tempering".

The elliptical equivalents of these examples can also be analyzed in a similar fashion by suitably transforming the variable to a spherical distribution and using the results stated in this section.

Chapter 5

Copulas

Copulas are an important tool to study the dependence structures in multivariate random variables. In this chapter we introduce copulas and study their various properties.

Definition 5.0.1. A copula is a d-dimensional distribution function on $[0, 1]^d$ with standard uniform marginal distributions.
In other words, $C(u) = C(u_1, \ldots, u_d)$ is a copula if and only if

1. $C(u_1, \ldots, u_d)$ is increasing in each component u_i for all $i \in \{1, \ldots, d\}$.

2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}$ and $u_i \in [0, 1]$.

3. For all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0, 1]^d$ with $a_i \leq b_i$ for all $i \in \{1, \ldots, d\}$, we have

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\ldots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0 \tag{5.1}$$

   where $u_{j_1} = a_j$ and $u_{j_2} = b_j$ for all $j \in \{1, \ldots, d\}$.

The first property is a direct implication of any multivariate probability distribution function. The second property can be implied from the fact that the marginal distributions of a copula are uniform. The third property, also called the rectangle-property ensures that if a random variable $(X_1, \ldots, X_d)$ has a distribution function C, $\mathbb{P}\left(a_1 \leq X_1 \leq b_1, \ldots, a_d \leq X_d \leq b_d\right) \geq 0$.

Theorem 5.0.1 (*Sklar's Theorem*)**.** Let F be a d-dimensional joint distribution function with marginals $F_1, \ldots, F_d$. Then, there exists a copula $C : [0,1]^d \to [0,1]$ such that for all $x_1, \ldots, x_d \in [-\infty, \infty]$,

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{5.2}$$

If the marginals are continuous, C is uniquely determined. Conversely, if C is a d-dimensional copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F defined in (5.2) is a joint distribution function with marginals $F_1, \ldots, F_d$. (For detailed proof, refer [11].)

Example 5.0.1. (*Independence Copula*) If X and Y are independent random variables, with distribution function F_X and F_Y respectively, then the joint distribution function of X and Y is $F(x, y) = F_X(x)F_Y(y)$. Hence the independence copula denoted by $C^{\perp}(u_1, u_2) = u_1 u_2$.
More generally the d-dimensional independence copula is given by,

$$C^{\perp}(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i$$

Definition 5.0.2. (*Survival Copulas*) A version of Sklar's theorem (5.2) also applies to multivariate survival functions. Let X be a random vector with multivariate survival function $\bar{F}$, marginal distribution functions $F_1, \ldots, F_d$ and marginal survival functions $\bar{F}_1, \ldots, \bar{F}_d$, i.e. $\bar{F}_i = 1 - F_i$. We have the identity

$$\bar{F}(x_1, \ldots, x_d) = \hat{C}(\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)) \tag{5.3}$$

for a copula $\hat{C}$, which is known as the survival copula. If the marginals are continuous, we get

$$\begin{aligned}\bar{F}(x_1,\ldots,x_d) &= \mathbb{P}(X_1 > x_1,\ldots,X_d > x_d) \\ &= \mathbb{P}(1-F_1(X_1) \leq \bar{F}_1(x_1),\ldots,1-F_d(X_d) \leq \bar{F}_d(x_d))\end{aligned} \tag{5.4}$$

Hence $\hat{C}$ is the distribution function of $1-U$ where $U := (F_1(X_1),\ldots,F_d(X_d))'$. In general, the term survival copula of a copula C will be used to denote the distribution function of $1-U$ when U has the distribution function C.

Definition 5.0.3. (*Radial symmetry*) A d-dimensional random vector X is radially symmetric about $a \in \mathbb{R}^d$ if $X - a \stackrel{d}{=} a - X$. Clearly, the class of spherical distributions are radially symmetric about 0 (Let $U = -I$ in 4.33) and in general, the class of elliptical distributions $E_d(\mu,\Sigma,\psi)$ are radially symmetric about μ.

Definition 5.0.4. (*Concordance ordering of copulas*) We say that the copula C_1 is smaller than copula C_2 i.e., $C_1 \leq C_2$ if

$$\forall (u_1,\ldots,u_d) \in [0,1]^d,\ C_1(u_1,\ldots,u_d) \leq C_2(u_1,\ldots,u_d). \tag{5.5}$$

Theorem 5.0.2. For any copula C, there are bounds

$$C^- \leq C \leq C^+ \tag{5.6}$$

where

$$C^-(u_1,\ldots,u_d) = \max\left(\sum_{i=1}^{d} u_i + 1 - d, 0\right) \tag{5.7}$$

$$C^+(u_1,\ldots,u_d) = \min(u_1,\ldots,u_d) \tag{5.8}$$

C^+ and C^- are called Fréchet upper and lower bounds respectively.

Proof. $C \le C^+$ can be inferred from the fact that, for all i,

$$\bigcap_{1 \le j \le d} \{U_j \le u_j\} \subset \{U_i \le u_i\} \tag{5.9}$$

Hence $\mathbb{P}\left[\bigcap_{1 \le j \le d} \{U_j \le u_j\}\right] \le \min_{i=1}^{n} \mathbb{P}\{U_i \le u_i\}$ which directly implies $C \le C^+$.

For the inequality $C^- \le C$, observe that

$$\begin{aligned} C(u) = \mathbb{P}\left(\bigcap_{1 \leqslant i \leqslant d} \{U_i \leqslant u_i\}\right) &= 1 - \mathbb{P}\left(\bigcup_{1 \leqslant i \leqslant d} \{U_i > u_i\}\right) \\ &\geqslant 1 - \sum_{i=1}^{d} \mathbb{P}\left(U_i > u_i\right) = 1 - d + \sum_{i=1}^{d} u_i. \end{aligned} \tag{5.10}$$

□

From this, we can observe the relation,

$$C^- \le C^{\perp} \le C^+ \tag{5.11}$$

In the next section we look into different types of dependence between random variables and discuss measures of dependence.

5.1 Dependence

Before we study dependence measures, we discuss co-monotonicity and counter-monotonicity of random variables and the copulas associated with them. These are examples where the random variables have perfect dependence with each other and all random variables can be simulated with a single random variable.

Definition 5.1.1. (*Co-monotonicity*) The random variables $X_1, \dots X_d$ are said to be comonotic if the copula associated with them is the Fréchet upper bound C^+.

Proposition 5.1.1. $X_1, \dots, X_d$ are comonotonic if and only if

$$(X_1, \dots, X_d) \stackrel{d}{=} (v_1(Z), \dots, v_d(Z)) \tag{5.12}$$

for some random variable Z and increasing functions $v_1, \dots, v_d$.

Proof. Let $X_1, \dots, X_d$ be comonotonic according to Definition 5.1.1. Let U be any uniform random variable and let $F, F_1, \dots, F_d$ be the joint distribution function and the marginal distribution functions respectively. We have

$$\begin{aligned} F(x_1, \dots, x_d) &= \min\{F_1(x_1), \dots, F_d(x_d)\} \\ &= P(U \leqslant \min\{F_1(x_1), \dots, F_d(x_d)\}) \\ &= P(U \leqslant F_1(x_1), \dots, U \leqslant F_d(x_d)) \\ &= P(F_1^{\leftarrow}(U) \leqslant x_1, \dots, F_d^{\leftarrow}(U) \leqslant x_d) \end{aligned}$$

where $F_1^{\leftarrow}, \dots, F_d^{\leftarrow}$ are the generalised inverses of the marginal distribution functions, which are increasing. Hence,

$$(X_1, \dots, X_d) \stackrel{d}{=} (F_1^{\leftarrow}(U), \dots, F_d^{\leftarrow}(U)) \tag{5.13}$$

Conversely if (5.12) holds

$$F(x_1, \dots, x_d) = \mathbb{P}(v_1(Z) \leq x_1, \dots, v_d(Z) \leq x_d) = \mathbb{P}(Z \in A_1, \dots, Z \in A_d) \tag{5.14}$$

Since v_i are increasing functions, A_i are of the form $(-\infty, k_i)$ or $(\infty, k_i]$, so one interval A_i is a subset of all other intervals. Therfore,

$$F(x_1, \ldots, x_d) = \min\{\mathbb{P}(Z \in A_1), \ldots, \mathbb{P}(Z \in A_d)\} = \min\{F_1(x_1), \ldots, F_d(x_d)\} \tag{5.15}$$

which proves co-monotonicity. □

Definition 5.1.2 (*Counter-Monotonicity*)**.** Random variables X_1 and X_2 are said to be counter-monotonic if the copula associated with them is the Fréchet lower bound.

Remark 5.1.1. The Fréchet lower bound is not a copula for $d > 2$.

Proof. Consider the cube $[1/2, 1]^d \subset [0, 1]^d$. If the Fréchet lower bound was a distribution function in $[0, 1]^d$, then (5.1) implies that probability mass $P(d)$ of this cube will be:

$$\begin{aligned}
P(d) =& \max(1 + \cdots + 1 - d + 1, 0) - d \max\left(\frac{1}{2} + 1 + \cdots + 1 - d + 1, 0\right) \\
&+ \begin{pmatrix} d \\ 2 \end{pmatrix} \max\left(\frac{1}{2} + \frac{1}{2} + \cdots + 1 - d + 1, 0\right) - \cdots \\
&+ \max\left(\frac{1}{2} + \cdots + \frac{1}{2} - d + 1, 0\right) \\
=& 1 - \frac{1}{2}d
\end{aligned} \tag{5.16}$$

Since probability can't be negative, the Fréchet lower bound can't be a copula for $d > 2$. This implies that the concept of counter-monotonicity is only applicable for a pair of random variables. □

Before moving to the next section, we state the following important theorem whose proof can be found in [4].

Theorem 5.1.1. Let $(X_1, \ldots, X_d)'$ be a random vector with copula C and let

$T_1, \ldots, T_d$ be strictly increasing functions. The random vector $(T_1(X_1), \ldots, T_d(X_d))$ also has the same copula C.

This theorem is important as it suggests that the copula of a particular random vector contains the full information of the dependence among them as it is preserved under monotonic transformations. That is, measures of dependence solely based on copulas would be invariant under monotonic transformations, a desirable property.

5.1.1 Dependence Measures

In this section we discuss 3 different measures of dependence namely linear correlation, rank correlation and coefficient of tail dependence.

Linear Correlation

Definition 5.1.3 (*Linear Correlation*)**.** The linear correlation between random variables X_1 and X_2 (denoted by $\rho(X_1, X_2)$) is defined if $\mathbb{E}(X_1^2) < \infty$ and $\mathbb{E}(X_2^2) < \infty$ by the expression,

$$\rho(X_1, X_2) = \frac{\operatorname{Cov}(X_1, X_2)}{\sqrt{\operatorname{Var}(X_1)\operatorname{Var}(X_2)}} \tag{5.17}$$

This is a measure of linear dependence and takes values in $[-1, +1]$. If X_1 and X_2 are independent, then $\rho(X_1, X_2) = 0$, but the converse is not true, i.e the uncorrelatedness of the random variables does not imply independence.

Also, correlation is invariant under strictly increasing linear transformations. In other words, for $\beta_1, \beta_2 > 0$ and for any $\alpha_1, \alpha_2 \in \mathbb{R}$,

$$\rho(\alpha_1 + \beta_1 X_1, \alpha_2 + \beta_2 X_2) = \rho(X_1, X_2) \tag{5.18}$$

However, correlation is not invariant under non-linear strictly increasing transformations.

Another important remark is that linear correlation is only defined when the variances of X_1 and X_2 are finite. This may cause problems when we are working with heavy tailed random variables.

Now, we discuss two more pitfalls of using linear correlation as a measure of dependence by demonstrating the following fallacies:

Fallacy 1: The marginal distributions and the pairwise correlations of a random vector determine its joint distribution.

Counterexample : Consider $(X_1, X_2) \sim \mathcal{N}(0, I_2)$. Both X_1 and X_2 have standard normal marginals and the independence copula $C^{\perp}(x, y) = xy$. Clearly $\rho(X_1, X_2) = 0$ as implied by independence.

Now, consider an independent discrete random variable V such that $\mathbb{P}(V = 1) = \mathbb{P}(V = -1) = 0.5$. Let $(Y_1, Y_2) = (X_1, VX_1)$. Clearly one can observe that $\rho(Y_1, Y_2) = 0$. However notice that the copula associated with this pair of random variables is $C(x, y) = 0.5 \max(x + y - 1, 0) + 0.5 \min(x, y)$ which is a mixture of two-dimensional Fréchet bound copulas. Hence (X_1, X_2) and (Y_1, Y_2) do not share the same joint distribution despite having same marginals and correlation.

Fallacy 2: A linear correlation between a pair of random variables can be determined by the copula associated with it.

Counterexample : As we have seen copulas are invariant with strictly increasing transformations. Here we give an example where a strictly increasing transform changes the correlation of the random variables hence giving a different correlation for the same copula.

Consider a pair of random variables X_1, X_2 such that $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 = X_1$. Hence, the copula associated with this pair is the Fréchet upper bound. The correlation between these variables is 1. Now consider a strictly increasing transformation $T(x_1, x_2) = (x_1, x_2^2)$ applied to this pair. That is, $(Y_1, Y_2) = (X_1, X_2^2)$. The covariance between Y_1 and Y_2 is given by

$$\mathrm{Cov}(Y_1, Y_2) = \mathbb{E}\left[(Y_1 - \mathbb{E}\left[Y_1\right])\left(Y_2 - \mathbb{E}\left[Y_2\right]\right)\right]$$

With $Y_1 = X_1$ and $Y_2 = X_2^2$, we get $\mathbb{E}[Y_1] = 0$ and $\mathbb{E}[Y_2] = 1$. Plugging these values in the above equation we get $\mathrm{Cov}(Y_1, Y_2) = \rho(Y_1, Y_2) = 0$. Hence correlation is not invariant with strictly increasing transformation whereas copulas are. The following theorem presents this concretely.

Theorem 5.1.2. If (X_1, X_2) has a joint distribution function F and marginal distribution function F_1 and F_2, then the covariance of X_1 and X_2, when finite is given by

$$\mathrm{Cov}(X_1, X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(F(x_1, x_2) - F_1(x_1)F_2(x_2))dx_1 dx_2 \tag{5.19}$$

Proof. Let (X_1, X_2) have distribution function F and $(\hat{X}_1, \hat{X}_2)$ be an independent copy. We have

$$2\,\mathrm{Cov}(X_1, X_2) = \mathbb{E}\left[(X_1 - \hat{X}_1)(X_2 - \hat{X}_2)\right]$$

We will now use an identity that says for any $a, b \in \mathbb{R}$, we can write

$$a - b = \int_{-\infty}^{\infty}\left(\mathbb{I}_{\{b \leqslant x\}} - \mathbb{I}_{\{a \leqslant x\}}\right)\mathrm{d}x$$

Applying this to random pairs $(X_1 - \hat{X}_1)$ and $(X_2 - \hat{X}_2)$, we get

$$\begin{aligned}
&2\,\mathrm{Cov}\,(X_1, X_2) \\
&= \mathbb{E}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\mathbb{I}_{\{\tilde{X}_1 \leqslant x_1\}} - \mathbb{I}_{\{X_1 \leqslant x_1\}}\right)\left(\mathbb{I}_{\{\tilde{X}_2 \leqslant x_2\}} - \mathbb{I}_{\{X_2 \leqslant x_2\}}\right) \mathrm{d}x_1\ \mathrm{d}x_2\right] \\
&= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\mathbb{P}\left(X_1 \leqslant x_1, X_2 \leqslant x_2\right) - \mathbb{P}\left(X_1 \leqslant x_1\right)\mathbb{P}\left(X_2 \leqslant x_2\right)\right) \mathrm{d}x_1\ \mathrm{d}x_2
\end{aligned} \tag{5.20}$$

which proves the theorem. □

Hence, covariance (and correlation) are inextricably linked to the marginal distributions and are not fully determined by the copula.

Now that we have seen the drawbacks of linear correlation as a measure of dependence, we now look at other measures which can be expressed in terms of copulas.

Rank Correlation

Rank correlations are measures of dependence that can be empirically calculated by looking at the rank of the data points alone. In this section we will be looking at two such measures : Kendall's tau and Spearman's rho.

Definition 5.1.4 (*Kendall's tau*)**.** For random variables X_1 and X_2, Kendall's tau is given by,

$$\rho_\tau(X_1, X_2) = \mathbb{E}\left[\mathrm{sign}((X_1 - \hat{X}_1)(X_2 - \hat{X}_2)\right] \tag{5.21}$$

where $(\hat{X}_1, \hat{X}_2)$ are independent copies of (X_1, X_2).

Alternatively (2.21) can also be written as:

$$\rho_\tau(X_1, X_2) = \mathbb{P}\left((X_1 - \hat{X}_1)(X_2 - \hat{X}_2) > 0\right) - \mathbb{P}\left((X_1 - \hat{X}_1)(X_2 - \hat{X}_2) < 0\right) \tag{5.22}$$

That is, it represents the fraction of concordant pairs (pairs of data with the same sign) minus the fraction of discordant pairs (pairs of data with opposite sign). Thus, it is natural that it is invariant under monotonic transformations, as we shall see further.

Proposition 5.1.2. Suppose X_1 and X_2 have continuous marginal distributions and a unique copula C. Then Kendall's tau is given by,

$$\rho_\tau(X_1, X_2) = 4\int_0^1\int_0^1 C(u_1,u_2)dC(u_1,u_2) - 1 \tag{5.23}$$

Proof. The fact that the marginals are continuous imply

$$\mathbb{P}\left((X_1-\hat{X}_1)(X_2-\hat{X}_2)<0\right) = 1-\mathbb{P}\left((X_1-\hat{X}_1)(X_2-\hat{X}_2)>0\right)$$

Plugging this relation in (5.22), we get

$$\rho_\tau(X_1, X_2) = 2\mathbb{P}\left((X_1-\hat{X}_1)(X_2-\hat{X_2})>0\right) - 1 \tag{5.24}$$

From the fact that the pairs (X_1, X_2) and $(\hat{X}_1, \hat{X}_2)$ are interchangeable, we get,

$$\begin{aligned}\rho_\tau(X_1, X_2) &= 4\mathbb{P}\left(X_1<\hat{X}_1, X_2<\hat{X}_2\right) - 1\\ &= 4\mathbb{E}\left[\mathbb{P}(X_1<\hat{X}_1, X_2<\hat{X}_2|X_1,X_2)\right] - 1\\ &= 4\int_{-\infty}^{\infty}\int_{\infty}^{\infty}\mathbb{P}(X_1<x_1, X_2<x_2)dF(x_1,x_2) - 1\end{aligned} \tag{5.25}$$

Since X_1 and X_2 have continuous margins, we may infer,

$$\rho_\tau(X_1, X_2) = 4\int_0^1\int_0^1 C(F_1(x_1),F_2(x_2))dC(F_1(x_1),F_2(x_2)) - 1 \tag{5.26}$$

Upon substituting $u_1 = F_1(x_1)$ and $u_2 = F_2(x_2)$ we get,

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1$$

□

As we can see, ρ_τ depends only on the copula, thus confirming our idea that it is invariant under monotonic transformations.

Definition 5.1.5 (*Spearman's rho*)**.** For random variables X_1 and X_2 with marginal distributions F_1 and F_2 respectively, Spearman's rho is given by the relation

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)) \tag{5.27}$$

Proposition 5.1.3. Suppose X_1 and X_2 have continuous marginal distributions and a unique copula C. Then Spearman's rho is given by

$$\begin{aligned} \rho_S(X_1, X_2) &= 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2 \\ &= 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3 \end{aligned} \tag{5.28}$$

Proof. Observe that $F_1(X_1)$ and $F_2(X_2)$ are uniform random variables with variance $\frac{1}{12}$. Hence we can write the Spearman's rho as

$$\rho_S(X_1, X_2) = 12 \operatorname{Cov}(F_1(X_1), F_2(X_2)) \tag{5.29}$$

Since $F_1(X_1)$ and $F_2(X_2)$ have a joint distribution C, with (5.19), we get the required result. □

From (5.29), we also get the representation,

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3 \tag{5.30}$$

Additionally, we can also show the following useful result,

$$\rho_S(X_1, Y_1) = 3\left(\mathbb{P}\left((X_1 - X_2)(Y_1 - Y_3) > 0\right) - \mathbb{P}\left((X_1 - X_2)(Y_1 - Y_3) < 0\right)\right) \tag{5.31}$$

where $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ are iid copies.

(For more detailed analysis on dependence measures based on copulas, refer [11]).

Coefficient of Tail Dependence

The motivation for looking at coefficients of tail dependence is that they provide measures of extremal dependence in bivariate random variables, or in other words, measure the strength of dependence in the tails of a bivariate distribution.
In case of upper tail dependence, we look at the probability that X_2 exceeds its q-quantile, given that X_1 exceeds its q-quantile. More formally we define it as follows:

Definition 5.1.6 (*Coefficients of Tail Dependence*)**.** Let X_1 and X_2 be random variables with distribution functions F_1 and F_2. The coefficient of upper tail dependence of X_1 and X_2 is

$$\lambda_u(X_1, X_2) = \lim_{q \to 1^-} \mathbb{P}(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q)) \tag{5.32}$$

provided the limit $\lambda_u \in [0, 1]$ exists. If $\lambda_u \in (0, 1]$, then X_1 and X_2 are said to have upper tail dependence or extremal dependence in the upper tail . If $\lambda_u = 0$, they are said to be asymptotically independent in the upper tail.

Analogously, the coefficient of lower tail dependence is

$$\lambda_l(X_1, X_2) = \lim_{q \to 0^+} \mathbb{P}(X_2 \leq F_2^{\leftarrow}(q) | X_1 \leq F_1^{\leftarrow}(q)) \tag{5.33}$$

provided the limit $\lambda_l \in [0, 1]$ exists.

It is simple to see that $\lambda_u(X_1, X_2) = \lambda_u(X_2, X_1)$ and similarly for λ_l as well. If the distribution functions F_1 and F_2 are continuous, then we get simple expressions in terms of the unique copula C of the bivariate distribution.

Proposition 5.1.4. Let (X_1, X_2) be a bivariate random variable with continuous marginals F_1 and F_2 and copula C, then tail dependence coefficeients are given by

$$\lambda_l(X_1, X_2) = \lim_{q \to 0^+} \frac{C(q, q)}{q} \tag{5.34}$$

$$\lambda_u(X_1, X_2) = \lim_{q \to 0^+} \frac{\hat{C}(q, q)}{q} \tag{5.35}$$

where $\hat{C}$ is the survival copula of (X_1, X_2).

Proof. By using the expression for conditional probability in (5.32), we get

$$\begin{aligned} \lambda_l &= \lim_{q \to 0^+} \frac{\mathbb{P}(X_2 \leq F_2^{\leftarrow}(q), X_1 \leq F_1^{\leftarrow}(q))}{\mathbb{P}(X_1 \leq F_1^{\leftarrow}(q))} \\ &= \lim_{q \to 0^+} \frac{C(q, q)}{q} \end{aligned} \tag{5.36}$$

Similarly, for upper tail dependence, we get,

$$\begin{aligned} \lambda_u &= \lim_{q \to 1^-} \frac{\mathbb{P}(X_2 \geq F_2^{\leftarrow}(q), X_1 \geq F_1^{\leftarrow}(q))}{\mathbb{P}(X_1 \geq F_1^{\leftarrow}(q))} \\ &= \lim_{q \to 1^-} \frac{\hat{C}(1-q, 1-q)}{1-q} \\ &= \lim_{q \to 0^+} \frac{\hat{C}(q, q)}{q} \end{aligned} \tag{5.37}$$

where $\hat{C}$ is the survival copula of (X_1, X_2). $\square$

Clearly, for radially symmetric distributions, $C = \hat{C}$ and thus $\lambda_l = \lambda_u$. Also, using L'Hôpital's rule in (5.34), we get the alternative expression,

$$\lambda_l(X_1, X_2) = \lim_{q \to 0^+} \frac{dC(q, q)}{dq} \tag{5.38}$$

5.2 Normal Mixture Copulas

A particularly useful class of copulas arise from the normal mixture distributions studied in Chapter 4. We begin with the variance-mixture distributions and look at their tail dependence and rank correlation structures.

5.2.1 Tail Dependence

First, we establish the definition of exchangeability of random vectors.

Definition 5.2.1. (*Exchangeability*) The random vector $(X_1, \ldots X_d)'$ is called exchangeable if

$$(X_1, \ldots, X_d)' \stackrel{d}{=} (X_{\pi(1)}, \ldots, X_{\pi(d)})' \tag{5.39}$$

for any permutation $(\pi(1), \ldots, \pi(d))$ of $(1, \ldots, d)$. Equivalently, we call its cdf exchangeable as well. Thus, we call a copula exchangeable if it is the distribution function of an exchangeable random vector with uniform marginals.

For example, $X \sim \mathcal{N}_2(0, I_d)$ is clearly exchangeable. In general, we can say that if $X \sim \mathcal{N}_2(\mu, \Sigma)$, then $\tilde{X} \sim \mathcal{N}_2(0, P)$ is exchangeable, where $\tilde{X}$ is X but with its components standardized to have mean 0 and variance 1 (so that $P = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, where $\rho = \rho(X_1, X_2)$).

As the standardizing operation is a strictly increasing one, X and $\tilde{X}$ have the same copula. We have already shown that $\tilde{X}$ is exchangeable, so, it follows that the general two dimensional Gaussian copula is exchangeable as well.

Theorem 5.2.1. Spherically distributed random vectors are exchangeable.

Proof. Let $X \sim S_d(\psi)$. By definition, $X \stackrel{d}{=} UX$ for any orthogonal matrix U. Corresponding to any permutation $(X_{\pi(1)}, \ldots, X_{\pi(d)})$, we have the permutation matrix,

$$P_\pi = \begin{bmatrix} e'_{\pi(1)} \\ \vdots \\ e'_{\pi(d)} \end{bmatrix}$$

where e_i is the i^{th} column of I_d. Clearly, $(X_{\pi(1)}, \ldots, X_{\pi(d)})' = P_\pi(X_1, \ldots, X_d)'$. P_π is trivially orthogonal, so $(X_1, \ldots, X_d) \stackrel{d}{=} (X_{\pi(1)}, \ldots, X_{\pi(d)})$. $\square$

In a similar fashion, we can show that the d-dimensional Gaussian copula is exchangeable whenever it admits an equicorrelation matrix of the form $P = \rho J_d + (1-\rho)I_d$ where $J_d \in \mathbb{R}^{d\times d}$ is a matrix with all entries equal to 1 and $|\rho| \leq 1$. Note that this agrees with our earlier discussion for 2 dimensions where the correlation matrix is always equicorrelational.

Theorem 5.2.2. $X = (X_1, X_2)' \sim \mathcal{M}_2\left(\mu, \Sigma, \hat{F}_W\right)$ has an exchangeable copula.

Proof. By definition, $X \stackrel{d}{=} \mu + \sqrt{W}AZ$, where the notation takes usual meaning. Here $Y = AZ$ follows a bivariate normal distribution, which has an exchangeable copula. Clearly, $X - \mu = \sqrt{W}Y$ is exchangeable as well. This corresponds to a component-wise strictly increasing transformation of X, meaning that the copulas of $X - \mu$ and X are the same. Thus X has an exchangeable copula.

That is, all bivariate normal variance mixture models have exchangeable copulas. □

Theorem 5.2.3. The conditional distribution of a bivariate copula C is given by,

$$C_{U_2|U_1}(u_2|u_1) = \frac{\partial C(u_1, u_2)}{\partial u_1} \tag{5.40}$$

wherever the partial derivative exists almost everywhere.

Proof. By definition,

$$\begin{aligned} C_{U_2|U_1}(u_2|u_1) &= \mathbb{P}\left(U_2 \le u_2 | U_1 = u_1\right) \\ &= \lim_{\delta \to 0} \frac{\mathbb{P}\left(U_2 \le u_2, u_1 \le U_1 \le u_1 + \delta\right)}{\mathbb{P}\left(u_1 \le U_1 \le u_1 + \delta\right)} \\ &= \lim_{\delta \to 0} \frac{C(u_1 + \delta, u_2) - C(u_1, u_2)}{\delta} \\ &= \frac{\partial C(u_1, u_2)}{\partial u_1} \end{aligned}$$

(For more details and precision refer [11]). □

From this, we can see that if $(U_1, U_2) \sim C$ and C is exchangeable, then,

$$\mathbb{P}\left(U_2 \le u_2 | U_1 = u_1\right) = \mathbb{P}\left(U_1 \le u_2 | U_2 = u_1\right) \tag{5.41}$$

Now, let $(X_1, X_2) \sim \mathcal{M}_2\left(\mu, \Sigma, \hat{F}_W\right)$ with exchangeable copula C. The normal variance mixtures are elliptical and thus are radially symmetric. So, $\lambda_l = \lambda_u$. Combining (5.38) and (5.40), we get,

$$\begin{aligned} \lambda_l(X_1, X_2) &= \lim_{q \to 0^+} \mathbb{P}\left(U_2 \le q | U_1 = q\right) + \mathbb{P}\left(U_1 \le q | U_2 = q\right) \\ &= \lim_{q \to 0^+} 2\mathbb{P}\left(U_2 \le q | U_1 = q\right) \end{aligned} \tag{5.42}$$

where $(U_1, U_2) \sim C$ and the second equality follows from (2.41). We now calculate this value for some examples.

Example 5.2.1. (Gaussian) Let $(X_1, X_2) \sim \mathcal{N}_2(0, P)$ where $P = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ where ρ (assume < 1) is the correlation between X_1 and X_2 with copula denoted as C_ρ^{Ga}. That is, $(\Phi(X_1), \Phi(X_2)) \sim C_\rho^{Ga}$ where Φ is the cdf of $\mathcal{N}(0, 1)$. So, we have the lower (and upper) tail dependence,

$$\begin{aligned}
\lambda_l(X_1, X_2) &= \lim_{q \to 0^+} 2\mathbb{P}\left(\Phi(X_2) \le q | \Phi(X_1) = q\right) \\
&= \lim_{q \to 0^+} 2\mathbb{P}\left(X_2 \le \Phi^{-1}(q) | X_1 = \Phi^{-1}(q)\right) \\
&= \lim_{x \to -\infty} 2\mathbb{P}\left(X_2 \le x | X_1 = x\right)
\end{aligned}$$

We know that $X_1 | X_2 = x \sim \mathcal{N}(\rho x, 1 - \rho^2)$, which gives us,

$$\lambda_l = \lim_{x \to -\infty} 2\Phi\left(\frac{x(1-\rho)}{\sqrt{1-\rho^2}}\right) = \lim_{x \to -\infty} 2\Phi\left(\frac{x\sqrt{1-\rho}}{\sqrt{1+\rho}}\right) = 0$$

That is, regardless of the correlation between the variables, extremal events occur independently for each marginal.

Example 5.2.2. (Student-t) Let $(X_1, X_2) \sim t_2(\nu, 0, P)$ where P is the correlation matrix. Let us also denote the bivariate t copula by $C_{\nu,\rho}^t$. Thus, $(t_\nu(X_1), t_\nu(X_2)) \sim C_{\nu,\rho}^t$ where t_ν is the cdf of the univariate Student-t distribution with ν degrees of freedom. So, we have the tail dependence,

$$\lambda_l = \lim_{q \to 0^+} 2\mathbb{P}\left(X_2 \le t_\nu^{-1}(q) | X_1 = t_\nu^{-1}(q)\right)$$

Further, it can be shown that,

$$\left(\left(\frac{\nu+1}{\nu+x^2}\right)^{1/2} \frac{X_2 - \rho x}{\sqrt{1-\rho^2}}\right)_{|X_1 = x} \sim t_{\nu+1} \tag{5.43}$$

Using this, we get:

$$\lambda_l = \lim_{x\to-\infty} 2t_{\nu+1}\left(\frac{x\sqrt{1-\rho}}{\sqrt{1+\rho}}\sqrt{\frac{\nu+1}{\nu+x^2}}\right) = 2t_{\nu+1}\left(-\frac{\sqrt{1-\rho}\sqrt{\nu+1}}{\sqrt{1+\rho}}\right)$$

That is, provided that $\rho > -1$, the bivariate student-t distribution shows dependent extremal behaviour. This is as expected, because we saw in Example 4.5.1, that this distribution has heavy tailed marginals due to its mixing distribution's power-law tails.

It is clear from the formula obtained that for a fixed ρ, tail dependence increases with decreasing ν and for a fixed ν, it increases with increasing ρ. Also, as $\nu \to \infty$, $\lambda_l \to 0$ which is consistent with the fact that the t-distribution approaches the Gaussian distribution in the same limit.

Proceeding in a similar fashion one can obtain the tail dependence coefficients for the other distributions in the Examples of Chapter 4. In doing so, one would find that the symmetric hyperbolic distributions would have 0 tail dependence. This is, again, due to the fact that their mixing distributions do not have power-law tails as was shown in Example 4.5.2. Along these lines, we cite the following theorem from [10] without proof,

Theorem 5.2.4. Let $X \stackrel{d}{=} \mu + RAS \sim E_d(\mu, \Sigma, \psi)$ with symbols taking their usual meaning and $\Sigma_{ii} > 0$ for all $i = 1, \ldots, d$. The following statements are equivalent.

1. R has a regularly varying tail with index $\alpha > 0$.

2. For all $i \neq j$ (X_i, X_j) has tail dependence.

When these conditions are met, the tail dependence is given by,

$$\lambda_l(X_i, X_j) = \lambda_u(X_i, X_j) = \frac{\int_{(\pi/2-\arcsin\rho_{ij})/2}^{\pi/2} cos^\alpha(t)dt}{\int_0^{\pi/2} cos^\alpha(t)dt} \tag{5.44}$$

where $\rho_{ij} = \rho(X_i, X_j)$ which can be inferred from Σ (which is in turn proportional to the covariance matrix). That is, it is the $(i, j)^{th}$ element of $P(\Sigma) = S^{-1}\Sigma S^{-1}$ where $S = \text{diag}(\sqrt{\Sigma_{11}}, \ldots, \sqrt{\Sigma_{dd}})$. We refer to $P(\Sigma)$ as the correlation operator.

5.2.2 Rank Correlation

We begin with the theorem below.

Theorem 5.2.5. Let $X \sim E_2(0, \Sigma, \psi)$ and $\rho = P(\Sigma)_{12}$ (simply, the correlation between X_1 and X_2). If $\mathbb{P}(X = 0) = 0$, then

$$\mathbb{P}(X_1 > 0, X_2 > 0) = \frac{1}{4} + \frac{\arcsin \rho}{2\pi} \tag{5.45}$$

Proof. Clearly, when we standardize the components of X as $Y \sim E_2(0, P, \psi)$ where $P = P(\Sigma)$, $\mathbb{P}(X_1 > 0, X_2 > 0) = \mathbb{P}(Y_1 > 0, Y_2 > 0)$.

Also, note that P admits the Cholesky decomposition $A = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$ such that $P = AA'$. Thus, by definition, Y has the representation,

$$Y = (Y_1, Y_2)' \stackrel{d}{=} \left(Z_1, \rho Z_1 + \sqrt{1-\rho^2} Z_2\right)'$$

where $Z = (Z_1, Z_2)' \sim S_2(\psi)$. This reduces Y to,

$$Y \stackrel{d}{=} R(\cos\Theta, \rho\cos\Theta + \sqrt{1-\rho^2}\sin\Theta)'$$

where we have used the stochastic representation $Z = R(\cos\Theta, \sin\Theta)'$ where R is a radial random variable and Θ is distributed uniformly over $[-\pi, \pi)$.

As $|\rho| \leq 1$, we can say $\rho = \sin\phi$. This allows us to write $Y_2 = R\sin(\phi + \Theta)$.

Since, $\mathbb{P}(R=0) = \mathbb{P}(Y=0) = 0$, we get,

$$\mathbb{P}(Y_1 > 0, Y_2 > 0) = \mathbb{P}(\cos\Theta > 0, \sin(\phi + \Theta) > 0)$$

This would require $\Theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \cap (-\phi, \pi - \phi)$. It is clear that this region always has length $\frac{\pi}{2} + \phi$ and thus probability $\frac{\frac{\pi}{2}+\phi}{2\pi}$ which proves our result. $\square$

With this, we now compute the two rank correlations discussed, for some distributions.

Example 5.2.3. (Gaussian copula) Let X be some bivariate distribution with Gaussian copula, C_ρ^{Ga}. Since the rank correlations are a purely copula property, we can simply assume the bivariate distribution as $\mathcal{N}_2(0, P)$ where P is a correlation matrix with off-diagonal elements as ρ and others as 1.

From (5.25), Kendall's tau is given by,

$$\rho_\tau(X_1, X_2) = 4\mathbb{P}\left(\tilde{X}_1 - X_1 > 0, \tilde{X}_2 - X_2 > 0\right) - 1$$

where $\tilde{X}$ is an independent copy of X. Setting $Y = \tilde{X} - X$, it is clear that $Y \sim \mathcal{N}_2(0, 2P)$ by the convolution property of the multivariate normal. As the covariance matrix has just been scaled, $\rho(Y_1, Y_2) = \rho$. Thus, using Theorem 5.2.5, we get,

$$\begin{aligned} \rho_\tau(X_1, X_2) &= 4\mathbb{P}(Y_1 > 0, Y_2 > 0) - 1 \\ &= \frac{2\arcsin\rho}{\pi} \end{aligned} \tag{5.46}$$

From (5.28), Spearman's rho is given by,

$$\begin{aligned}\rho_S(X_1, X_2) &= 12 \int_0^1 \int_0^1 \mathbb{P}\left(X_1 \leq \Phi^{-1}(u_1), X_2 \leq \Phi^{-1}(u_2)\right) du_1 du_2 - 3 \\ &= 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{P}\left(X_1 \leq x_1, X_2 \leq x_2\right) \phi(x_1)\phi(x_2) dx_1 dx_2 - 3\end{aligned}$$

where ϕ is the standard normal density. Clearly, this is equivalent to,

$$\begin{aligned}\rho_S(X_1, X_2) &= 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{P}\left(X_1 \leq x_1, X_2 \leq x_2 | Z_1 = x_1, Z_2 = x_2\right) \phi(x_1)\phi(x_2) dx_1 dx_2 - 3 \\ &= 12\mathbb{P}\left(X_1 < Z_1, X_2 < Z_2\right) - 3 \\ &= 12\mathbb{P}\left(Y_1 > 0, Y_2 > 0\right) - 3\end{aligned}$$

where $Z = (Z_1, Z_2)'$, Z_1 and Z_2 are independent of each other and of X, and we set $Y = Z - X$. By the convolution property, $Y \sim \mathcal{N}_2(0, P + I_2)$ where $\rho(Y_1, Y_2) = \frac{\rho}{2}$. Thus, from Theorem 5.2.5, we get,

$$\rho_S(X_1, X_2) = \frac{6 \arcsin\left(\frac{\rho}{2}\right)}{\pi} \tag{5.47}$$

We can show that the result for Kendall's tau for the Gaussian copula holds in general for Elliptical distributions (and thus, the normal variance-mixture distributions as well) as summarized in the theorem below.

Theorem 5.2.6. Let $X \sim E_2\left(0, P, \psi\right)$ where P is a correlation matrix with off-diagonal entries ρ. Then, $\rho_\tau(X_1, X_2) = \frac{2 \arcsin \rho}{\pi}$ holds.

Proof. Again, from (5.25), Kendall's tau is given by,

$$\rho_\tau(X_1, X_2) = 4\mathbb{P}\left(\tilde{X}_1 - X_1 > 0, \tilde{X}_2 - X_2 > 0\right) - 1$$

where $\tilde{X}$ is an independent copy of X.

From the convolution property of elliptical distributions (refer [4], section 3.3.3), $Y = \tilde{X} - X \sim E_2\left(0, P, \tilde{\psi}\right)$ for some characteristic generator $\tilde{\psi}$. However, we have shown in Theorem 5.2.5 that $\mathbb{P}\left(Y_1 > 0, Y_2 > 0\right)$ is independent of the characteristic generator. Thus, by using Theorem 5.2.5, we get,

$$\rho_\tau(X_1, X_2) = \frac{2 \arcsin \rho}{\pi} \tag{5.48}$$

□

Of course, this applies to any elliptical distribution with arbitrary location and dispersion matrices as well, as those correspond to increasing linear transformations of the components of the random vector considered in this theorem.

In contrast, it can be shown that (5.47) does not hold for all elliptical distributions for Spearman's rho. (In fact, no closed form is known for the elliptical distributions other than the Gaussian.) We construct a counterexample as follows: Let $X = RAS \sim E_2\left(0, P, \psi\right)$ with symbols taking their usual meaning. Consider $W \stackrel{d}{=} AS$. It can be shown that,

$$\rho_S(W_1, W_2) = 3\frac{\arcsin \rho}{\pi} - 4\left(\frac{\arcsin \rho}{\pi}\right)^3 \tag{5.49}$$

This can be proved using the (5.31) formulation of Spearman's rho along with the fact that $S \stackrel{d}{=} (\cos\Theta, \sin\Theta)$ where Θ is uniform over $[-\pi, \pi)$, similar to the proof of Theorem 5.2.5. The full proof can be found in [10].

Chapter 6

Empirical Analysis II

In this chapter, we look at some data from the Indian markets to test the applicability of the distributions discussed so far. First, we perform univariate analysis using the 1 dimensional marginals of the distributions in the same way as Chapter 3. In Chapter 3, we observed the Pareto-tail distribution to be better than Stable and Gaussian distributions. Here, we first perform the same analysis to the normal mixture distributions.

Further, we also fit the multidimensional mixture distributions to multivariate data and look at their goodness of fit.

For estimating the parameters of the GH_d distributions, we use the Expectation Maximization (EM) approach explained in [4], Section 3.2.4. We briefly summarize the method below:

Suppose we have iid realizations $X_1, X_2, \ldots, X_n \in \mathbb{R}^d$ to which we wish to fit the GH_d or one of its special cases (holding the corresponding parameters constant). We do this by maximizing the log-likelihood function,

$$\ln L\left(\theta; X_1, \ldots, X_n\right) = \sum_{i=1}^{n} \ln f_X(X_i; \theta) \tag{6.1}$$

where $\theta = (\lambda, \chi, \psi, \mu, \Sigma, \gamma)'$ is the parameter vector and $f_X(x;\theta)$ denotes the generalized hyperbolic density in (4.23). This problem is not particularly easy as we would have to maximize over several parameters which include the dispersion matrix Σ as well. However, if we were able to "observe" the W_i's corresponding to the mixing variable, we may write,

$$\ln \tilde{L}\left(\theta; X_1, \ldots, X_n, W_1, \ldots, W_n\right) = \sum_{i=1}^{n} \ln f_{X|W}(X_i; W_i, \mu, \Sigma, \gamma) + \sum_{i=1}^{n} \ln h_W(W_i; \lambda, \chi, \psi) \tag{6.2}$$

where $f_{X|W}$ would be a normal density, h_W is the GIG density and $\ln \tilde{L}$ is called the augmented-log-likelihood. However, $\{W_i\}$ are "latent variables" in the sense that they cannot be observed. To overcome this, we use the EM algorithm wherein we iteratively replace W_i in (6.2) by estimates of W_i obtained from previous estimates of θ. In short, it consists of two steps, namely:

1. *E-Step*: We calculate the conditional expectation of the augmented likelihood (6.2) given the data $X_1, ..., X_n$ using the parameter values $\theta^{[k]}$ from the previous iteration. This results in the objective function,

$$Q(\theta; \theta^{[k]}) = \mathbb{E}\left[\ln \tilde{L}\left(\theta; X_1, \ldots, X_n, W_1, \ldots, W_n\right) | X_1, \ldots, X_n; \theta^{[k]}\right] \tag{6.3}$$

2. *M-Step*: We maximize Q in (6.3) to obtain the next set of parameters $\theta^{[k+1]}$.

In this manner we obtain a sequence of improved parameter estimates. We overcome the problem of the latency of $\{W_i\}$ by replacing every instance of any function $g(W_i)$ with $\mathbb{E}\left[g(W_i)|X_i, \theta^{[k]}\right]$. These expectations can be found with the help of (4.12)-(4.13).

Full details of this algorithm are explained in [4].

6.1 Univariate Analysis

We employ the same methodology from [2], again. We estimate the J's, the mean $\bar{J}$ and its corresponding confidence interval as defined in Section 3.3.3. We summarize the results in the scatter plot in Figure 6.1 and the confidence intervals in Figure 6.2.

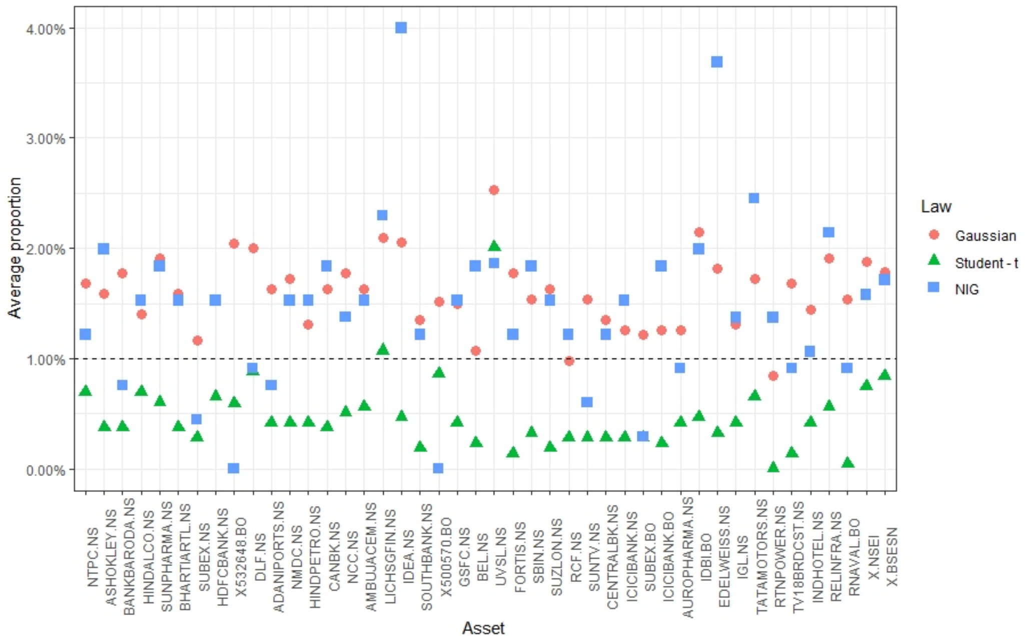

Figure 6.1: Average J, Gaussian, Student-t and NIG

6.1.1 Observations

1. Both the plots immediately suggest that the Student-t distribution gives the best fit in terms of VaR estimation.

2. Further, we see that the NIG distribution shows large variability in the

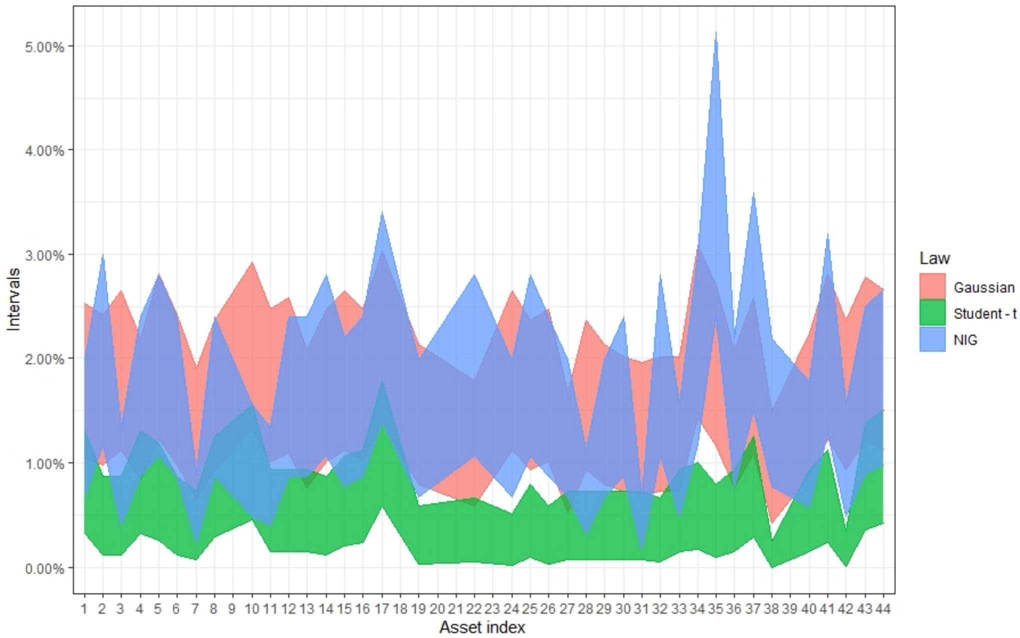

Figure 6.2: Confidence Intervals, Gaussian, Student-t and NIG

confidence interval and $\bar{J}$ values (sometimes even reaching 4%). This may be a product of the parameter estimation methods. Similar to the Stable distributions, we see large variability in the parameter estimates for the NIG distribution which seems to translate into the confidence interval estimates as well.

3. For 28 out of 44 assets, the 1% value is within the estimated confidence interval for the NIG distribution and the same holds true for 43 assets in the Student-t case. In terms of the backtesting results, it seems as if the Student-t performs as well as the Paretian model and outperforms NIG and Gaussian distributions.

4. This could be because of the fact that the Student-t has power-law tails but the NIG has the tempered semi-heavy tails as discussed in Example 4.5.2.

6.2 Multivariate Analysis

In this section, we fit the multivariate models that we have discussed so far to the log-losses of three assets listed on NSE whose Yahoo Finance tickers are HDFCBANK.NS, SBIN.NS and ICICIBANK.NS. We have chosen these assets as they are very likely to show dependence between each other and call for multivariate analysis. Also, we have chosen the timeline from January 2007 to December 2010 as there is a larger chance for marginal and joint extremal events in this range. Additionally, it was noticed that as the size of the dataset increased the stability of the parameter estimates (like those seen in the univariate analysis) improved. Also, it is to be noted that we assume stationarity over the whole time interval and thus fit these static distributions. While this assumption may not be completely accurate, it helps us to understand the heavy-tailed nature of financial returns. To this end, we fit both symmetric and asymmetric variants of the discussed mixture distributions and compare them.

6.2.1 Symmetric models

Table 6.1 summarizes the Log-likelihood of the fitted model for each symmetric distribution. As we can see, the Gaussian distribution underperforms to a great extent in comparison to the heavier-tailed mixture models. We also notice that among the special cases, the symmetric NIG distribution comes out as the best model with the Student-t as a close second. There seems to be only a minor improvement by using a general symmetric GH_3 distribution.

Figure 6.3 shows a pairwise plot matrix of the fitted symmetric NIG distribution. The rows and columns are ordered as "HDFCBANK.NS, SBIN.NS,

Table 6.1: Symmetric models, Log-likelihoods

	Gaussian	Student-t	NIG	Hyperbolic	VG	GH_3
LLH	6853.53	7027.57	7029.39	7023.44	7024.27	7029.46

Table 6.2: Asymmetric models, Log-likelihoods and $\|\gamma\|_2$

	Skewed Student-t	NIG	Hyperbolic	VG	GH_3
LLH	7028.85	7030.82	7024.45	7025.11	7030.88
$\|\gamma\|_2$	3.32×10^{-3}	3.44×10^{-3}	2.88×10^{-3}	2.36×10^{-3}	3.44×10^{-3}

ICICIBANK.NS". The plot on the i^{th} diagonal entry corresponds to the QQ plot of the fitted Gaussian, the fitted symmetric NIG against the empirical quantiles of the i^{th} asset. The (i, j) plot on the lower triangle corresponds to a 2-dimensional normalized histogram of the joint distribution of assets i and j. Similarly, the upper triangle plots correspond to the scatter plots of the joint distribution of assets i and j. We plot the same for the Student-t distribution as well (Figure 6.4).

As we can observe, the Gaussian quantiles show considerable deviation from the sample quantiles whereas the symmetric NIG (and Student-t) fit shows very little deviation. We also observe on the histogram and scatter plots that there have been joint extremal events (on either tail). The same is true in the Student-t case. This will be discussed further in the next section.

6.2.2 Asymmetric models

Now, we fit the asymmetric mixture models to the same data and check if there are any considerable benefits by allowing the possibility of asymmetry.

Table 6.2 summarizes the Log-likelihoods and estimated $\|\gamma\|_2$ of the fitted

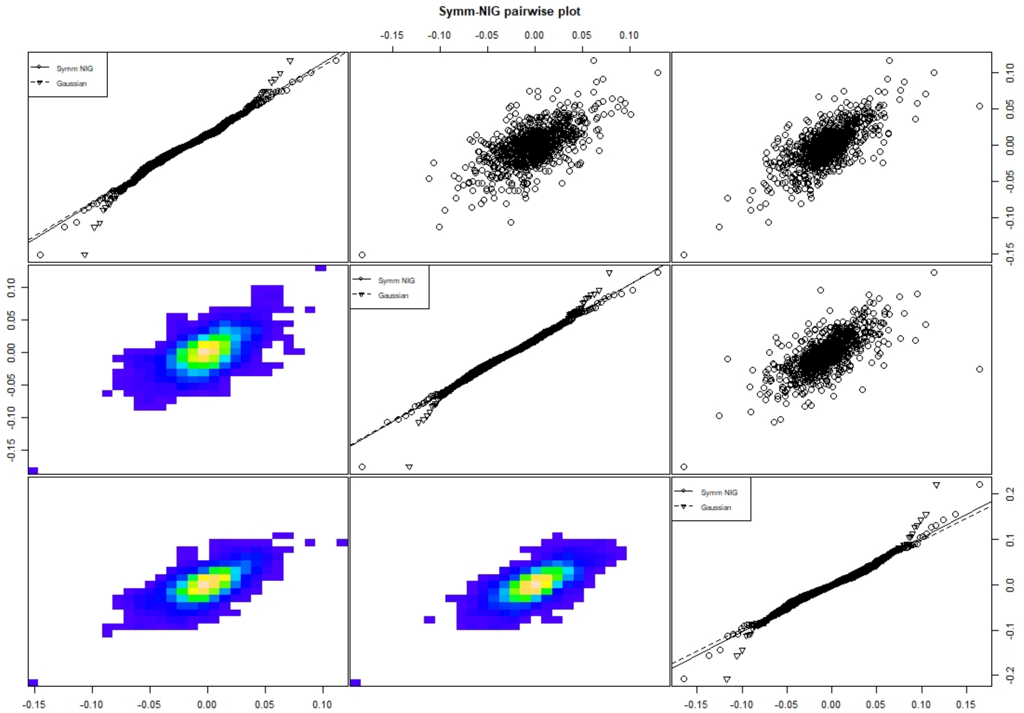

Figure 6.3: Symmetric NIG plot matrix

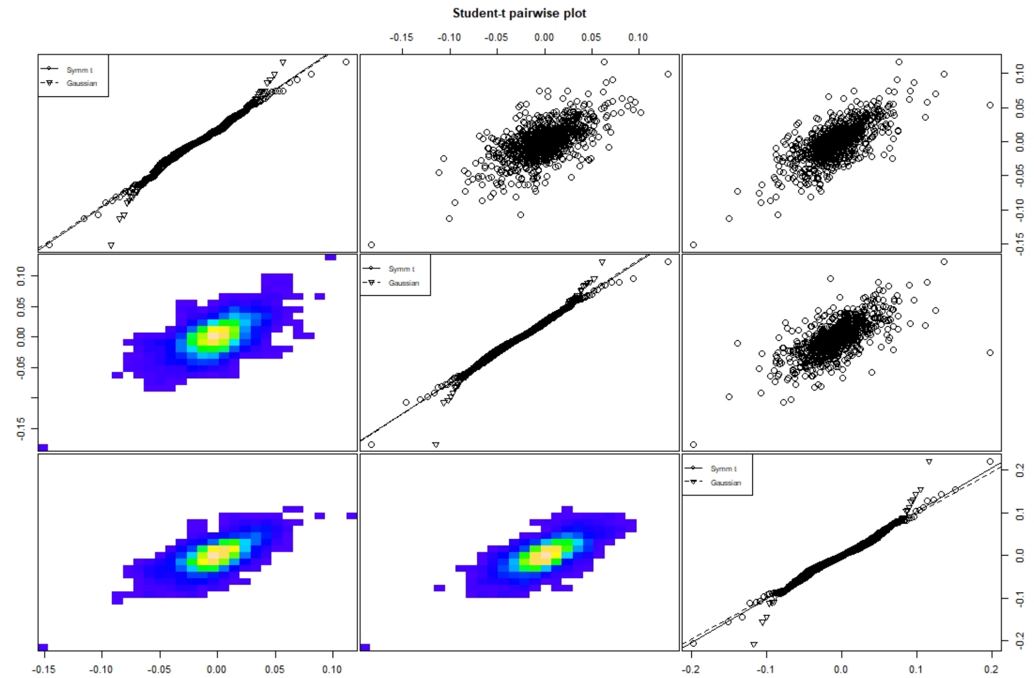

Figure 6.4: Student-t plot matrix

models. As we can see, the asymmetric models provide very little improvement over the symmetric, elliptical models. We can see this by observing that the parameter γ causes the asymmetry and its estimated norm is small. This could be due to the fact that the considered assets show high inter-dependence and are thus unlikely to show asymmetric behaviour between themselves.

In contrast, when we consider four assets "HDFCBANK.NS, SBIN.NS, BHARTIARTL.NS, SUNPHARMA.NS" which are not expected to be highly inter-dependent, we see that there is a larger difference in log-likelihoods of the asymmetric models compared to the symmetric equivalents. For this case, it was observed that the asymmetric VG distribution provided the best fit. The 4×4 pairwise plot is shown in Figure 6.7.

Once again, in the three asset case, we see that the asymmetric NIG is the best out of the special cases with the Skewed Student-t coming in second, and the general GH_3 provides a very minor improvement. Similar to Figure 6.1, Figure 6.5 shows the pairwise plot matrix of the fitted asymmetric NIG distribution and Figure 6.6 for the Skewed Student-t.

In total, we observe that the symmetric NIG is the best of the multivariate mixture models for the considered three assets. We choose this over the asymmetric model which adds complexity without much benefit in return. In addition, it is evident that the NIG performed worse than Student-t in the univariate analysis because of the small window size $W = 252$ which caused high uncertainty in the parameter estimation for the NIG distribution. This highlights the importance of the trade-off between the stationarity assumption and reliable parameter estimation while choosing the size of data.

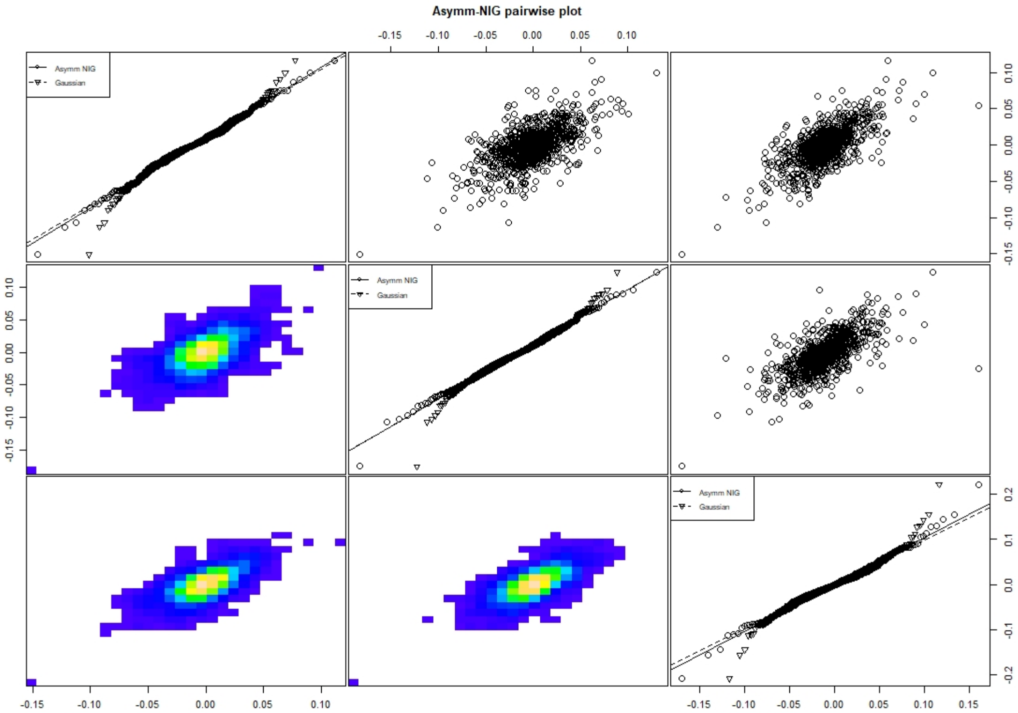

Figure 6.5: Asymmetric NIG plot matrix

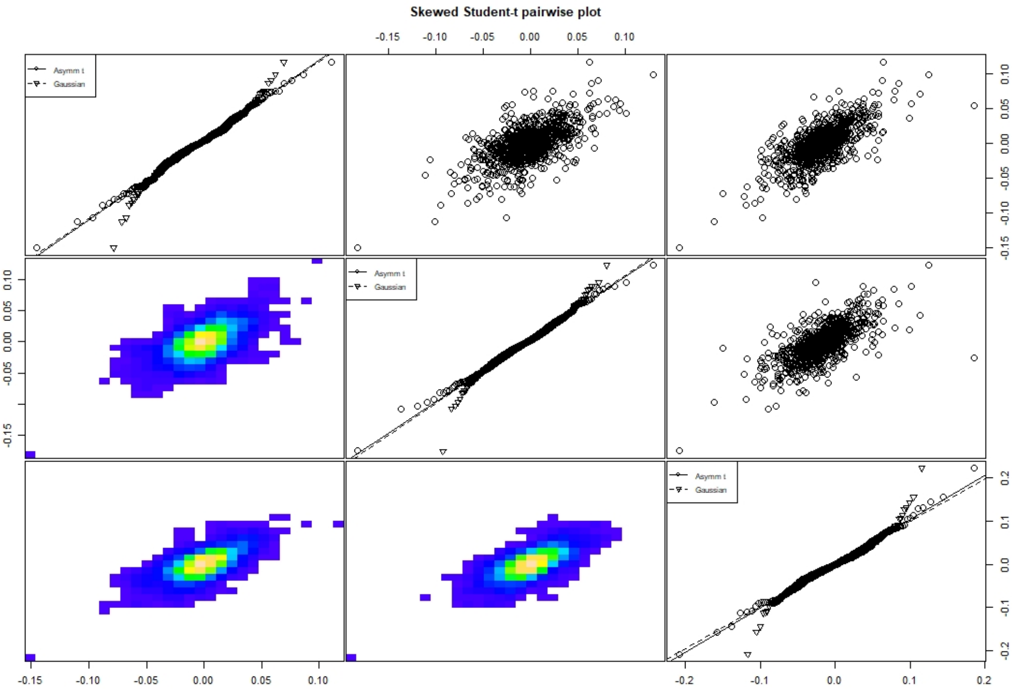

Figure 6.6: Skewed Student-t plot matrix

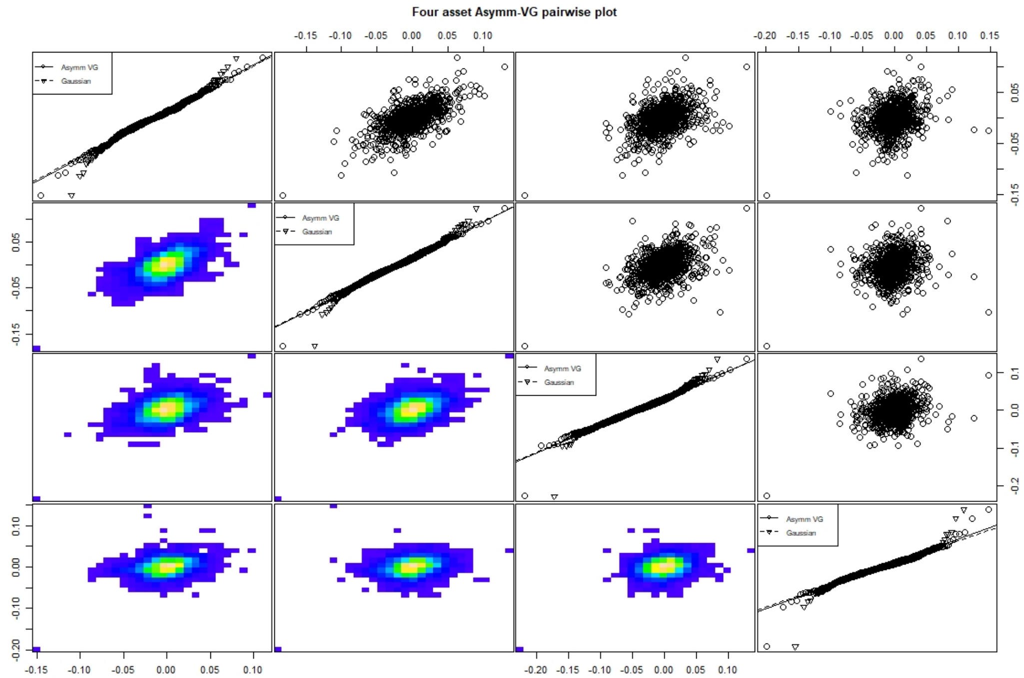

Figure 6.7: Asymmetric VG plot matrix for HDFCBANK.NS, SBIN.NS, BHARTIARTL.NS, SUNPHARMA.NS

6.3 Copula Analysis

In this section we fit a Gaussian copula and a t-copula to the log returns data of HDFCBANK.NS, ICICIBANK.NS and SBIN.NS. We wish to calculate the 95% joint exceedence probability of the log losses for the set of stocks with the fitted copula and compare this against the empirical exceedance frequency. The empirical exceedance frequency is simply the proportion of times the daily log losses are in the top 95 percentile for all the 3 stocks.

Definition 6.3.1. (*Joint Exceedance Probabilities*) Joint exceedence probability at level α of a multivariate random variable $X = (X_1, \ldots, X_d)$ with marginal cdfs $(F_1, \ldots, F_d)$, is defined as the probability with which all the random variables are larger than their respective α-quantiles. Let us denote this probability by $J_X(\alpha)$. So,

$$J_X(\alpha) = \mathbb{P}\left(X_1 > F_1^{-1}(\alpha), \ldots, X_d > F_d^{-1}(\alpha)\right) \tag{6.4}$$

Let $X = (X_1, \ldots, X_d)$ be the (continuous) log-losses of d assets. Thus, $J_X(\alpha)$ gives us the probability that the losses of all assets jointly exceed their α quantiles, which would be a noteworthy extremal event. We can also get an expression for $J_X(\alpha)$ in terms of the log-returns. Clearly,

$$\begin{aligned} J_X(\alpha) &= \mathbb{P}\left(-X_1 < -F_1^{-1}(\alpha), \ldots, -X_d < -F_d^{-1}(\alpha)\right) \\ &= \mathbb{P}\left(-X_1 < \tilde{F}_1^{-1}(1-\alpha), \ldots, -X_d < \tilde{F}_d^{-1}(1-\alpha)\right) \\ &= \tilde{C}(1-\alpha, \ldots, 1-\alpha) \end{aligned} \tag{6.5}$$

where $\tilde{F}_i$ is the cdf of $-X_i$ and $\tilde{C}$ is the copula for $-X$.

Thus, our problem for the three assets has reduced to fitting a copula to their log-returns, finding the value of $\tilde{C}(0.05, 0.05, 0.05)$ and comparing this

Table 6.3: Empirical vs Fitted $J_X(0.95)$

Method	$J_X(0.95)$
Empirical	0.012560
C_P^{Ga}	0.008334
$C_{\nu,P}^t$	0.011826

against the empirical exceedance frequency to check for goodness of fit in the tails.

If the empirical exceedence frequency is greater than the $\tilde{C}(0.05, 0.05, 0.05)$ value, then the copula underestimates the tail dependence and if it is lower than the $\tilde{C}(0.05, 0.05, 0.05)$ value then the copula overestimates the tail dependence.

6.3.1 Observations

We collect daily price data for the stocks HDFCBANK.NS, ICICIBANK.NS and SBIN.NS for the period from 01-01-2007 to 01-01-2017. We choose such a large time-range due to the relative rarity of joint exceedance events. We then fit a Gaussian copula C_P^{Ga} and a t copula $C_{\nu,P}^t$, where P is the correlation matrix for both cases and ν is the degrees of freedom, using the `copula` package in R. It uses a maximum likelihood estimator to fit the parameters. We then calculate the $C_P^{Ga}(0.05, 0.05, 0.05)$ and $C_{\nu,P}^t(0.05, 0.05, 0.05)$ as the joint exceedance probability we would expect in case of the fitted Gaussian and t-Copula respectively. The empirical joint exceedance and the joint exceedance expected from the fitted copula are showin in Table 6.3

This shows that normal copula greatly underestimates joint exceedence events and t Copula is better at estimating the joint exccedence probability.

Chapter 7

Conclusion

In this project, we first took up a study on the probability theory concerning the class of Infinitely Divisible distributions. We saw that these distributions could be characterized by the Canonical and Lévy measures.

Following this, we studied the class of Stable distributions which are a special class of Infinitely Divisible distributions. From the structure of the canonical and Lévy measures, we could conclude that the stable distributions are fat tailed, making them a candidate to model financial (log-)return data. Explicit formulae for the characteristic function of these distributions were also calculated.

Further, we saw that the stable distributions were the only possible limit distributions of appropriately scaled and centered sums of random variables. This entailed the study of Domains of Attraction in which we characterized the random variables that can be present in a Domain of Attraction, that is, have a CLT-type limit law. To close the chapter, we saw a method to actually obtain limit laws for such variables and applied it to several examples.

In Chapter 3, we undertook an empirical analysis of stocks from the Indian markets wherein we use VaR estimates to assess the goodness of fit in the tails of the distributions discussed so far. Here, we found that the Gaussian vastly underestimated the tail events, whereas the stable distributions overestimated them, and the Pareto-tail distribution struck the balance between the two. This study suggests that, while the log-returns are fat-tailed and violate the assumptions made in classical financial models, they can, however, be reasonably modelled with distributions with finite variance and hence, certain aspects of portfolio theory and CAPM may still apply.

Chapter 4 onward, we discussed multivariate distributions. The primary focus was on the Normal Mixture distributions which are popular choices to model multivariate log-return data. We characterized these distributions as "randomized" multivariate normals and looked at special cases such as the Generalized Hyperbolic distributions. Further, we discussed the more general classes of spherical and elliptical distributions and undertook an analysis of the tail behaviours of the marginals of the same. This allowed us to show that the Student-t had power-law tails whereas the rest of the Generalized Hyperbolic distributions had semi-heavy tempered power-law tails, thus making all of these distributions good candidates for modelling log-returns.

Following this, Chapter 5 introduced Copulas as an important tool to study the dependence structures among random variables. To this end, we demonstrated some fallacies concerning linear correlation and showed the benefits of using Copula-based dependence measures such as the rank correlations. Further, we also introduced tail dependence, a measure of asymptotic dependence between random variables. Finally, we explored the properties of the Normal Mixture copulas by studying their tail dependence and rank correlation structures.

Finally, in Chapter 6, we first repeated the univariate analysis from Chapter 3 to the marginals of the mixture models and found that the Student-t provides the best VaR estimates and the NIG had parameter variability due to the small window size, with both of them outperforming the Gaussian. Then, we took up a multivariate analysis where we fit the Symmetric and Asymmetric GH_d distributions and found that the symmetric NIG gives the best fit for the considered three asset case and the asymmetric VG gives the best fit for the four asset case. In the last section, we also did a copula analysis where we found that the Student-t-copula provides better estimates of the joint exceedance probabilities than the Gaussian copula, further corroborating the heavy-tailed nature of log-returns.

In this study, we had looked at both univariate and multivariate models. Using the best-fit multivariate model (such as the NIG), one may attempt to optimize a portfolio of stocks in the vein of Markowitz portfolio theory. There is also the possibility to borrow ideas from Extreme Value Theory to model the largest possible loss from a portfolio of stocks. One could also consider a stochastic process of stock prices whose log-returns follow one of the mixture models and use it to price options, perhaps improving on the Black-Scholes framework. In conclusion, our analysis highlighted the importance of the heavy tailed nature of stock returns and also the dependence structures that arise from multivariate heavy tailed models and presents a lot of scope for future research to build on these ideas.

Bibliography

[1] O. Barndorff-Nielsen , P. Blaesild. Hyperbolic distributions and ramifications: Contributions to theory and application. 1981.

[2] N. Champagnat , M. Deaconu , A. Lejay , N. Navet , S. Boukherouaa. An empirical analysis of heavy-tails behavior of financial data: The case for power laws. *HAL Archives-Ouvertes*, 2013.

[3] L. Breiman. On some limit theorems similar to the arc-sin law. 1965.

[4] A. J. McNeil , R. Frey , P. Embrechts. *Quantitative Risk Management - Concepts, Techniques and Tools.* Princeton Series in Finance. Princeton University Press, 2005.

[5] W. Feller. *An Introduction to Probability Theory and its Applications, Volume II.* Second Edition. Wiley, New York, 1971.

[6] O. Barndorff-Nielsen , C. Halgreen. Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. 1977.

[7] S. Janson. Some integrals related to the gamma integral. Notes, 2011.

[8] S. Janson. Stable distributions. Notes, 2011.

[9] B. Jørgensen. *Statistical Properties of the Generalized Inverse Gaussian Distribution.* Lecture Notes in Statistics. Springer, New York, 1982.

[10] H. Hult , F. Lindskog. Multivariate extremes, aggregation and dependence in elliptical distributions. 2002.

[11] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, 2006.