Extreme value theory with applications in finance

A thesis submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

MATHEMATICAL STATISTICS

in the

DEPARTMENT OF STATISTICS

of

RHODES UNIVERSITY

by

Aphelele Matshaya

(0000-0002-0626-0095)

February 2024

Supervisor: Professor L. Raubenheimer

(0000-0003-0756-7459)

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Rhodes University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 10th September 2024

Copyright ©2024Rhodes University
All rights reserved.

Abstract

The development and implementation of extreme value theory models has been very significant as they demonstrate an application of statistics that is very much needed in the analysis of extreme events in a wide range of industries, and more recently the cryptocurrency industry. The crypto industry is booming as the phenomenon of cryptocurrencies is spreading worldwide and constantly drawing the attention of investors, the media, as well as financial institutions. Cryptocurrencies are highly volatile assets whose price fluctuations continually lead to the loss of millions in a variety of currencies in the market. In this thesis, the extreme behaviour in the tail of the distribution of returns of Bitcoin will be examined. High-frequency Bitcoin data spanning periods before as well as after the COVID-19 pandemic will be utilised. The Peaks-over-Threshold method will be used to build models based on the generalised Pareto distribution, and both positive returns and negative returns will be modelled. Several techniques to select appropriate thresholds for the models are explored and the goodness-of-fit of the models assessed to determine the extent to which extreme value theory can model Bitcoin returns sufficiently. The analysis is extended and performed on Bitcoin data from a different crypto exchange to ensure model robustness is achieved. Using Bivariate extreme value theory, a Gumbel copula is fitted by the method of maximum likelihood with censored data to model the dynamic relationship between Bitcoin returns and trading volumes at the extreme tails. The extreme dependence and correlation structures will be analysed using tail dependence coefficients and the related extreme correlation coefficients. All computations are executed in R and the results are recorded in tabular and graphical formats. Tail-related measures of risk, namely Value-at-Risk and Expected Shortfall, are estimated from the extreme value models. Backtesting procedures are performed on the results from the risk models. A comparison between the negative returns of Bitcoin and those of Gold is carried out to determine which is the less risky asset to invest in during extreme market conditions. Extreme risk is calculated using the same extreme value approach and the results show that Bitcoin is riskier than Gold.

Keywords: Bitcoin, Bivariate extremes, Extreme dependence, Extreme value theory, Generalised Pareto distribution, High-frequency data, Tail risk.

Contents

List of Figures

List of Tables

List of Abbreviations

ADF - Augmented Dickey-Fuller

AIC - Akaike information criterion

BM - Block Maxima

BTC - Bitcoin

cdf - cumulative distribution function

CLT - Central Limit Theorem

df - distribution function

ES - Expected Shortfall

EUR - Euro

EVI - Extreme Value Index

EVT - Extreme Value Theory

FTSE - Financial Times Stock Exchange

GARCH - Generalised Autoregressive Conditional Heteroscedasticity

GBM - Geometric Brownian Motion

GEVD - Generalised Extreme Value Distribution

gof - goodness-of-fit

GPD - Generalised Pareto Distribution

HFT - High-Frequency Trading

HS - Historical Simulation

i.i.d. - independent and identically distributed

IQR - Interquartile Range

IRB - Internal Ratings Based

JSE - Johannesburg Stock Exhange

K-S - Kolmogorov-Smirnov

LR - Likelihood Ratio

MC - Monte Carlo

ME - Mean Excess

MLE - Maximum Likelihood Estimation/Estimator

MRL - Mean Residual Life

MSE - Mean Squared Error

pdf - probability density function

POT - Peaks-over-Threshold

Q1 - Lower Quartile

Q3 - Upper Quartile

QQ - Quantile-Quantile

SD - Standard Deviation

TDC - Tail Dependence Coefficient

USD - United States Dollar

VaR - Value-at-Risk

ZAR - South African Rand

Acknowledgements

I would like to thank my supervisor, Professor Lizanne Raubenheimer, for introducing me to this interesting field of research and providing very useful study materials, as well as proof-reading this thesis. I would also like to express my gratitude to Mr Jeremy Baxter for his unwavering willingness to help, advice on R, as well as his continuous encouragements during the course of my postgraduate studies. I want to acknowledge the National Research Foundation (NRF) in collaboration with the Centre of Excellence: Mathematical and Statistical Sciences (CoE-MaSS) for their financial assistance towards funding this Masters degree.

Chapter 1

Introduction

1.1 Overview

The modelling of financial gains and losses in financial markets has always been a prominent area of research. This is primarily due to the turbulence experienced in the market which triggers a need for financial risk management. It is imperative that financial institutions and investors attempt to find a way to forecast and quantify potential risk in order to prevent financial losses. Traditionally, predominant financial models were contingent on the normality assumption. However, modelling financial returns using Gaussian models proved to be impractical and inaccurate, especially during periods of market stress or the occurrence of an extreme event such as a market decline or large price fluctuations. This kind of problem led to the development of Extreme Value Theory (EVT).

EVT entails a theoretical framework that models extreme values that result from extreme or rare events. It is a probabilistic theory that uses mathematical and statistical methods to build models that can describe and predict extreme occurrences in an efficient manner. It has been extensively applied in numerous fields to study extremes in natural disasters, temperatures, price systems, and insurance claims, to name a few. Extreme events are rare events that possess substantial impact. The extreme values are then identified as the large (extreme) observations in data, and are modelled using EVT distributions.

The cryptocurrency market exhibits a continuous increase in trading activity all over the world. This is owing to numerous factors, including the advancement of technology along with the revolutionary creation of these digital assets called cryptocurrencies. Bitcoin continually emerges as the driving force behind the escalating demand to trade in cryptocurrencies. It is the longest standing cryptocurrency, thus majority of traders in the market have been exposed to it the most. Its price changes, and thus its returns, display extreme behaviour owing to its high volatility and risky nature.

The cryptocurrency market comprises of a high number of different digital currencies which provide fast and effective monetary payments. The application of extreme value thinking in this kind of market

is relatively new and is an interesting financial application for this research, as it is imperative for traders to evaluate and understand the underlying risks associated with trading during times of market stress. This enables them to protect themselves against the large losses that can accumulate from market turbulences which would be considered more extreme than the usual, and ultimately adjust their trading strategies accordingly.

Prices are constantly changing in this market hence many investors base their trade decisions on personal speculation and subjective judgement. This kind of thinking, however, is not efficient during periods when the market is faced with extreme conditions - e.g. an extreme event occurs, such as a market crash or extreme market exuberance. It is worth noting that even though these extreme events are highly unlikely, they are not impossible, and can bear heavy consequences (extreme values) such as large financial losses. On the other hand, they can also generate large fortunes (financial returns) for traders.

Although investment in the cryptomarket is widely accessible to the general public, a feasible investment typically requires a good trading strategy - one that ensures that losses are minimal and gains (i.e. profits) are maximised - compiled using market research and well-established stylised facts. Uncertainty in the market is an inherent trait since volatility is necessary for prices to increase in any financial market. However, for first time traders entering the market, financial risk (i.e. generating a loss) is not the only concern.

Over the years, it has been observed that some experienced traders manipulate new inexperienced traders into changing their trade decisions. Moreover, a growing number of unreputable 'cryptocurrency influencers' exploit ordinary citizens by giving them defective information about the market and the size of profits that can be generated, leading to millions of dollars in irrecoverable losses. Using EVT, one can build models that can be used to predict and mitigate the extreme financial losses of ordinary citizens in the crypto market. These models contribute to effective financial risk management and encourage financial sustainability.

1.2 Aims

This research seeks to demonstrate the usefulness and practicality of EVT in real world finance by proposing to solve some key problems encountered in finance using EVT models. Three main objectives will be covered in this thesis. The first is to study and compare the two main methods for classifying extremes in EVT (Block Maxima and Peaks-over-Threshold) and subsequently apply the preferable one. Various statistical models formulated using EVT methodologies will be fitted to cryptocurrency data in order to quantify and predict extreme risk. The focus will be on large losses as an extreme financial risk, although positive returns (gains) will not be disregarded. Reason being, the uncertainty that comes with risk is more amplified for events that carry negative effects. Various mod-

els for estimating risk will be explored and their performance assessed to test for their accuracy and effectiveness.

The second aim is to analyse the relationship between cryptocurrency returns and the volume of the currency traded when the market is under extreme conditions. This will be done using bivariate EVT and copula modelling. The results can be used to assist traders gain useful insight into how their trading activities can affect potential returns during the occurrence of extreme events in the market. They can then make predictions based on the extreme value models. The third objective is to perform an extreme value analysis to examine which is the better investment between Gold and Bitcoin. Ultimately, the conclusion is based on which one of these assets is less risky, i.e. carries less extreme risk.

Assessing and understanding the extreme returns in Bitcoin and their properties, the amount of risk they carry, as well as the dynamic relationship between returns and volume, using EVT is the main motivation of this research. Although the return-volume relation has been extensively researched in financial markets, an application into the cryptocurrency market is quite limited. For this thesis, a research contribution is carried out to analyse the underexplored tail relationship between trading volume and financial returns for Bitcoin at high frequency, as well as the analysis of extreme tail dependence and extreme correlation between these two variables.

In order to meet these objectives, R (R Development Core Team, 2020) was used for building and fitting the models, as well as implementing backtesting procedures.

1.3 Thesis Outline

This thesis is segmented into four chapters. Chapter 2 first introduces significant concepts of univariate EVT, including relevant definitions and explanations that encompass the theoretical and mathematical background of EVT. The two main methodologies along with the distributions used in EVT modelling are explained, as well as parameter estimation techniques, and which methodology will be selected for the application to be carried out. Various models to estimate extreme tail risk are then discussed in Section 2.6 as well as how risk is measured and backtesting procedures to test for model performance. An overview of Bivariate EVT and copula theory is then provided in Section 2.10. The literature on the application of EVT in finance, and more importantly cryptocurrencies, is reviewed and a summary of the applicability of this field of research is included at the end.

Chapter 3 presents the statistical data analysis and results. The different sets of data used for the application of Chapter 2 are described and details of the procedures used in the data analysis are documented. Various techniques under the chosen methodology are explored and model robustness is assessed. The results are provided in the form of tables, graphs and interpretations that make it easier to draw conclusions needed for the last chapter. Chapter 4 documents the conclusions made from the inferences drawn and summarises the concluding remarks. The key findings of the thesis are discussed,

along with any limitations. The R code used is provided in the Appendix.

Chapter 2

Literature Review

2.1 Introduction to Univariate Extreme Value Theory

Extreme Value Theory (EVT) is a useful statistical tool for studying probabilities associated with extreme and rare events found in the tails of probability distributions. It seeks to predict these probabilities and assess the asymptotic behaviour of the extreme observations of a random variable (Beirlant et al., 2004), providing a statistical description of these extreme events. It then models the tail of the distribution – where the extremes are found – using sound extreme value methods derived from this theory.

Probabilistic EVT, and it's associated models, was first pioneered by Gumbel (1958) in the early forties of the twentieth century. According to Kotz and Nadarajah (2000), EVT initially attracted the interests of hydrologists and engineers as it was widely applied to natural occurrences such as extreme floods, freak waves and air pollution. It later attracted mainstream statisticians, as Gumbel (1958) details in his significant contribution to extreme analysis. He was the the first to instigate possible applications of the formal 'extreme value' theory, as we know it, to certain distributions which had been treated empirically in the past.

The theory of extreme values, as Gumbel (1958) states, essentially aims to analyse observed extremes and predict further extremes, outside the range of available data. This fundamental theory is a well developed body of knowledge that provides two main methods, namely the Block Maxima (BM) and Peaks-over-Threshold (POT), for the statistical modelling of these rare events and estimation of tail-related risk measures (Beirlant et al., 2004). It enables a rational extrapolation of rare or extreme occurences that exceed any historic occurence that may be considered as rather normal, giving us an idea of the magnitude of the probabilities of these extreme events occuring, even when they have never occurred before (Panjer, 2006).

This area of research is concerned with the stochastic behaviour of extremes – maxima and minima of independent and identically distributed (i.i.d.) random variables. Assume that we have the realisations

from a sample of data $X_1, X_2, \ldots, X_n$ (i.i.d. random variables) which can be transformed into the ordered data $X_{1,n} \leq \ldots \leq X_{n,n}$ with $X_{1,n} = \min(X_1, X_2, \ldots, X_n)$ and $X_{n,n} = \max(X_1, X_2, \ldots, X_n)$ (Beirlant et al., 2004). The sample data is then used to study properties associated with the distribution function

$$F(x) = P(X \leq x),$$

as well as its inverse function (the quantile function) defined as:

$$Q(p) := \inf\{x : F(x) \geq p\}.$$

The distributional properties of these extremes, order statistics and exceedances over high thresholds or below low thresholds are determined by the upper and lower tails of the underlying distribution (Kotz and Nadarajah, 2000). The converse is true, that is we can use statistical procedures based on extreme order statistics or exceedances over high thresholds to evaluate the tail of the underlying distribution.

A substantial amount of literature exists on the application of EVT for estimating risk in finance and insurance. Throughout the world, financial markets have been characterised by significant instabilities. This has led to criticism regarding the existing risk management systems and has stimulated the search for more improved methodologies to cope with rare events that bear heavy consequences (Gilli and Këllezi, 2006). The first researchers pertaining to the 'theory of largest values', according to Gumbel (1958), stem from the normal distribution.

EVT continues to play an increasingly significant and effective role in modelling events that occur in a wide range of fields, such as finance, insurance, engineering, quality control, geology and hydrology; demonstrating the successfulness of its application in modern science problems (Gumbel, 1958). According to Chikobvu and Jakata (2020), the implementation of EVT became more popular in finance after the occurrences of several international financial crises; such as the Global financial crisis of 2007-2008 and the Chinese stock market crash (2015-2016).

These crises revealed the flaws of financial risk models, especially with the traditional methods of estimating risk under the assumption of normality in the distribution of financial data falling short when the data is heavy-tailed. Traditional models may be effective in the centre of the distribution, but they are inefficient in the tails of the distribution where the data is sparse (Dicks et al., 2014). Using EVT methods in analysing financial data becomes more efficient as EVT focuses on the tails without forcing a single distribution to fit the entire sample, provides improved extreme quantile fits for fat-tailed data, and also allows for asymmetry (Chinhamu et al., 2015b).

The history of the application of mathematical modelling in finance can be traced back to the famous stochastic process: the Geometric Brownian Motion (GBM), which provided the basis of the Black-Scholes option pricing model. However, it has been observed in a number of financial markets over the years that the distribution of market movements are usually heavy-tailed and not normally distributed.

This heavy-tailedness of a distribution can be described by a quantitative measure called a "kurtosis" greater than 3 – where the standard normal distribution is assigned a kurtosis equal to 3 and distribution functions with a kurtosis greater than 3 are considered heavy-tailed. As a consequence of its particularly stringent assumptions of normality, the Black-Scholes model is not very practical for real world markets.

For instance, large price variations in the market caused by a change in volatility may possibly be induced by some extreme events. Since these extreme events have the potential of high risks, they can cause even larger shocks to the underlying financial system. This has led to an increasing need for models that are able to explain how classical probabilistic results change if the underlying assumptions allow for larger shocks to the system. EVT models and techniques do not only outline the statistical nature of problems in finance, but they allow a precise mathematical description of certain notions of extreme events (Embrechts et al., 1997).

Drawing statistical inference about rare events has a distinct link to the observations that are considered "extreme" in a sample. Extreme values are linked to small probabilities, which justifies the usefulness of EVT in fitting a distribution to the observations that lie in the tail, providing adequate models that can extrapolate beyond what has been observed in the tail (Wentzel and Mare, 2007). Estimating tail probabilities is quite important as predictive power is one of the leading objectives of statistical modelling. Below is an example from Beirlant et al. (2004) 's explanation of EVT applications:

Suppose that if the percentage change in price exceeds 20% we may run the risk of market turbulence. This is a state where a financial market experiences wide unpredictable upward and downward swings in prices. Stock markets are inherently unstable when it comes to prices, as price fluctuations are inevitable. One might seek to model the extreme price changes, X, of a highly volatile financial variable in the stock market.

Of major interest would be the estimation of the tail probability $p = P(X > 20\%)$. Conventionally the empirical distribution function is defined as

$$F_n(x) = \frac{i}{n} \text{ if } x \in [x_{i,n}, x_{i+1,n}]$$

where $x_{i,n}$ is the i^{th} value of the ordered sample (Beirlant et al., 2004).

This will lead to

$$\hat{p} = 1 - \hat{F}_n(20\%).$$

However, according to Beirlant et al. (2004), the classical theory largely based on the normal distribution and the central limit theory would not apply. This traditional approach fails to answer pertinent questions concerning the extent of risk or financial losses caused by extreme changes in price. For instance, what if we wish to estimate $p = P(X > x)$, where $x > x_{n,n}$ and our estimate $\hat{p}$ evaluates to the

value 0? The value of x would be considered as extreme as this change in price would exceed even the maximum observed price change in the sample, meaning it would result in significantly adverse effects on the financial stability of the market.

We certainly cannot assume that such an event ($x > x_{n,n}$) would be impossible as that would restrict the forecasting or predictive capabilities of our modelling, and further obscure the extent of risk associated with this extreme event. Large and unexpected price fluctuations in the stock market (both up and down) can lead to market turbulence - which neither benefits the investors nor the companies, since investors will tend to make impulsive and drastic changes in their portfolios during this period - resulting in substantial fluctuations in their demand to invest.

Posing an even greater concern are the macroeconomic impacts of a market turbulence, such as faster interest rate rises and increased inflation. However, it is important to note that not all financial instruments are able to have such a profound effect on the stock market and the economy at large. Therefore, market turbulence is uncommon and rarely disrupts the extensive movements of the market.

From the perspective of the empirical quantile function

$$\hat{Q}(p) = \inf\left\{x : \hat{F}_n(x) \geq p\right\},$$

this kind of problem (as mentioned above) occurs when we consider very high quantiles $\hat{Q}(1-p)$ with small probability $p < \frac{1}{n}$ (Beirlant et al., 2004). This kind of problem demonstrates the necessity to develop special techniques that focus on the extreme values of a sample, or more inclusively the extreme high quantiles or the small tail probabilities (Beirlant et al., 2004). As described in the above example on price variations, the application of extreme value reasoning or, in simple terms, extreme values can be of key interest in the real world and should not be disregarded.

Wentzel and Mare (2007) claim that the disregard of extreme values in the statistical literature, as well as in practice, dates back to the traditional three standard deviations rule adapted from Fourier's findings in 1824 (which suggested that all observations lying more than three standard deviations away from the mean can be neglected). As previously mentioned, extreme observations typically lie in the tails and may carry the potential of high risk. Thus, neglecting them could result in large financial gains or losses, depending on which perspective is undertaken in the industry - i.e. extremely large financial gains or losses made by profit-driven organisations offering a product or service in the market, or consumers and investors utilising the product or service.

Often, in financial markets we consider modelling the returns of some instrument. On account of the asymmetry or skewness typically encountered in financial data, it is sometimes necessary to model the left and right tails separately in order to distinguish between gains and losses. Extreme events in finance have the prime advantage that they are usually quantifiable in monetary units, which makes the mathematical modelling more manageable. Market participants are often interested in calculating and pricing their risks. Using extremal events, one is able to determine the probable maximal loss of a risk

or investment portfolio (Embrechts et al., 1997). In order to solve problems in finance using EVT, it is necessary to carry out statistical inference about the extreme events in a population.

What constitutes an extreme value depends on the financial problem at hand and what kind of losses or gains we are trying to predict. Some examples of quantities that can assist in forecasting extreme events, and thus safeguard against the inauspicious effects they cause (Embrechts et al., 1997), include

- the distribution of the largest values in a portfolio
- the expected maximum loss of an investment portfolio
- the frequency of rare events
- the distribution of the excesses
- the distribution of financial returns.

Some large data sets exhibit skewness and heavy tails, and are therefore described poorly by normal/ Gaussian models. Financial data, more particularly one that describes extreme events, is contradicted by the traditional normality assumption due to the characteristics it inherits – such as volatility clustering, skewness and fat tails (Beirlant et al., 2004). These characteristics can be explained by the presence of extreme values and they result in a peaked density. Fortunately, a significant amount of empirical research shows that EVT is able to accommodate the aforementioned properties and provide models that are practical for the data encountered in real world finance.

Of particular concern in financial risk management are the maximum losses from a financial instrument that can be expected. EVT characterises extreme observations using the two previously mentioned methods in the beginning of this section (BM and POT). Using statistical modelling, EVT builds models that quantify and forecast the unusually large financial losses or gains of some financial instrument (such as stock market indices) that carry the potential of a high risk. These models can assist risk analysts and investors to outline the risks associated with these instruments, and manage future extreme losses by putting appropriate measures in place to reduce risk exposure, in both the short and the long run (Chikobvu and Jakata, 2020).

This notion not only benefits the finance sector, but contributes to the growth of the economy as a whole. Notably, the performance of EVT models fitted to financial data, as well as parameter estimation techniques, can be assessed to ensure that the best fit to the data is chosen. This enhances the efficiency of EVT modelling.

2.2 Limit Theory

According to De Haan and Ferreira (2006), the asymptotic theory of extremes in a sample was developed analogous to the central limit theorem (CLT), and that these two are quite similar to one another.

If we have $X_1,...,X_n$ i.i.d. random variables, the CLT deals with the limit behaviour of the partial sums $X_1+X_2+X_3+\ldots+X_n$ as $n\to\infty$, whilst the theory of sample extremes pertains to the limit behaviour of the extreme values in the sample: $X_{n,n}$ or $X_{1,n}$ as $n\to\infty$.

Beirlant et al. (2004) further elaborate that contrary to the convergence in the CLT, the normal distribution cannot be the resulting limiting distribution of the sample extremes owing to the inherent skewness which is observed in a distribution of maxima for instance. In the next section we will look at the probability theory of how the limits of extreme values in extreme value analysis are expressed.

2.3 Generalised Extreme Value Distribution (GEVD)

In extreme value methodology we assume that the unknown underlying distribution F belongs to the domain of attraction of a GEVD. Let $X_1,...,X_n$ be a random sample from F and write $X_{n,n}=\max(X_1,...,X_n)$ and assume there exists sequences $a_n>0$ and b_n such that as $n\to\infty$,

$$P\left(\frac{X_{n,n}-b_n}{a_n}\leq x\right)\to G_\gamma(x)\,, \tag{2.1}$$

where γ is known as the extreme value index (EVI) and γ^{-1} is the rate of tail decay, Beirlant et al. (2004).

This limiting distribution is inspired by the Fisher-Tippet theorem (also known as the extreme value theorem) which provides a general description of the asymptotic distribution of extreme order statistics and was formerly defined by Gnedenko (1948) as the following theorem:

Theorem. *Fisher–Tippet theorem*

Given a sequence of i.i.d. random variables $X_1,...,X_n$ from an unknown distribution F. Let $X_{n,n}=max(X_1,...,X_n)$ and assume there exist sequences of real numbers $a_n>0$ and $b_n\,\varepsilon\,\mathbb{R}$ such that the following limit exists

$$\lim_{n\to\infty}P\left(\frac{X_{n,n}-b_n}{a_n}\leq x\right)=G(x)$$

where $G(x)$ is a non-degenerate distribution function. Then, $G(x)$ belongs to the 'Generalised Extreme Value' (GEV) class of distributions $G_\gamma(x)$.

All extreme value distributions, under this parametric approach, can be expressed as the one-parameter standard GEVD

$$G_\gamma(x)=\begin{cases}\exp\left[-(1+\gamma x)^{-1/\gamma}\right], & \text{for}\quad \gamma\neq 0,\\ \exp\left[-\exp(-x)\right], & \text{for}\quad \gamma=0,\end{cases} \tag{2.2}$$

when $1+\gamma x > 0$, $x \in \mathbb{R}$, can occur as limits in (2.1). And the related location-scale family GEVD can be obtained by replacing x with $\frac{x-\mu}{\sigma}$ for $\mu \in \mathbb{R}$, $\sigma > 0$ and $1+\gamma\left(\frac{x-\mu}{\sigma}\right) > 0$,

$$G_{\gamma,\mu,\sigma}(x) = \begin{cases} \exp\left[-\left(1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\gamma}\right], & \text{for} \quad \gamma \neq 0, \\ \exp\left[-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right], & \text{for} \quad \gamma = 0, \end{cases} \tag{2.3}$$

where μ and σ represent the location and scale parameters, respectively.

Analogously, X can be replaced by $-X$ to obtain the limiting distribution of the minima (Kotz and Nadarajah, 2000).

As stated in Beirlant et al. (2004), the extreme value index, γ, defines which class type the generalised extreme value distribution belongs to

- When $\gamma = 0$, it belongs to the light-tailed Gumbel class.
- When $\gamma > 0$, it belongs to the heavy-tailed Fréchet class.
- When $\gamma < 0$, it belongs to the short-tailed Weibull class.

Figures 2.1 and 2.2 illustrate the probability density plots of these three classes for different values of the EVI parameter.

The Fisher–Tippet theorem implies that if the maxima of a sample, properly normalised, converge in distribution to a non-degenerate limit, the limiting distribution belongs to one of the above three classes of the GEVD.

The Fisher–Tippet theorem unfortunately does not guarantee that such sequences exist, as they in fact do not exist for numerous distributions. Nevertheless, they generally do exist for several common distributions (De Haan and Ferreira, 2006). The role played by this extreme value theorem for maxima can be paralleled to that of the CLT for averages. The latter explains that the limit of the normalised mean statistic (average) from any distribution that has a finite variance converges to the normal distribution, while the former does not state that the normalised maxima converge, but rather requires them to converge in order for their limiting distribution to be a GEVD.

Gumbel-class distributions ($\gamma = 0$)

Also known as Extremal type 1 class distributions (Beirlant et al., 2004), the maxima converge to the following distribution function:

$$P(X \leq x) = G_{\gamma,\mu,\sigma}(x) = \exp\left[-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right].$$

Fréchet-class distributions ($\gamma > 0$)

Under this distribution, the maxima have the limiting distribution:

$$G_{\gamma,\mu,\sigma}(x) = P(X \leq x) = \begin{cases} 0, & \text{for} \quad x < \mu - \frac{\sigma}{\gamma}, \\ \exp\left[-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right], & \text{for} \quad 1+\gamma\frac{x-\mu}{\sigma}, \text{ i.e. } x \geq \mu - \frac{\sigma}{\gamma}. \end{cases}$$

Weibull-class distributions ($\gamma < 0$)

This class gives the limiting distribution:

$$G_{\gamma,\mu,\sigma}(x) = P(X \leq x) = \begin{cases} \exp\left[-\left(1+\gamma\frac{\mu-x}{\sigma}\right)^{\frac{1}{\gamma}}\right], & \text{for } 1+\gamma\frac{\mu-x}{\sigma}, \text{ i.e. } \quad x \leq \mu - \frac{\sigma}{\gamma}, \\ 0, & \text{for} \quad x > \mu - \frac{\sigma}{\gamma}. \end{cases}$$

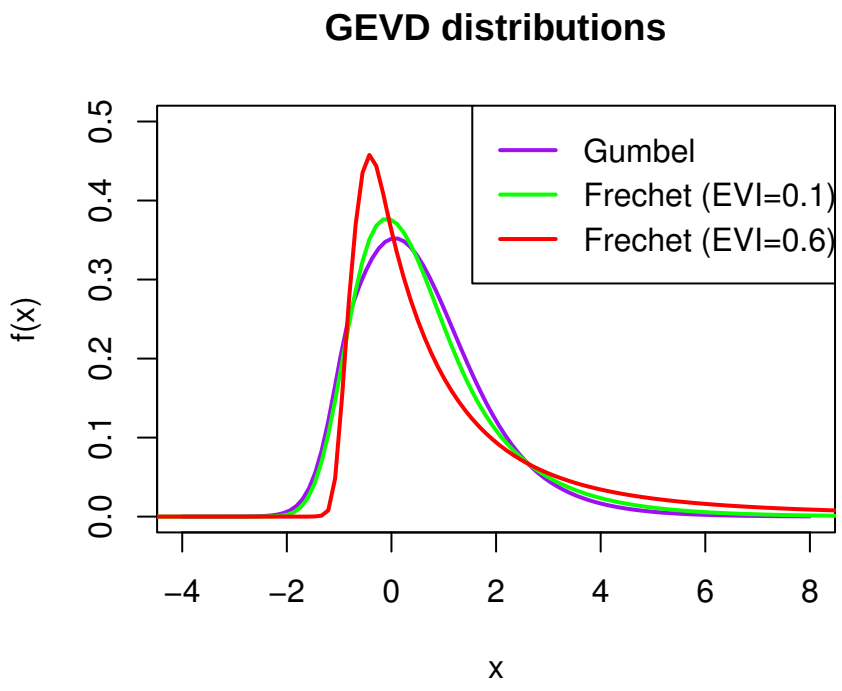


Figure 2.1: Gumbel and Fréchet class distributions.

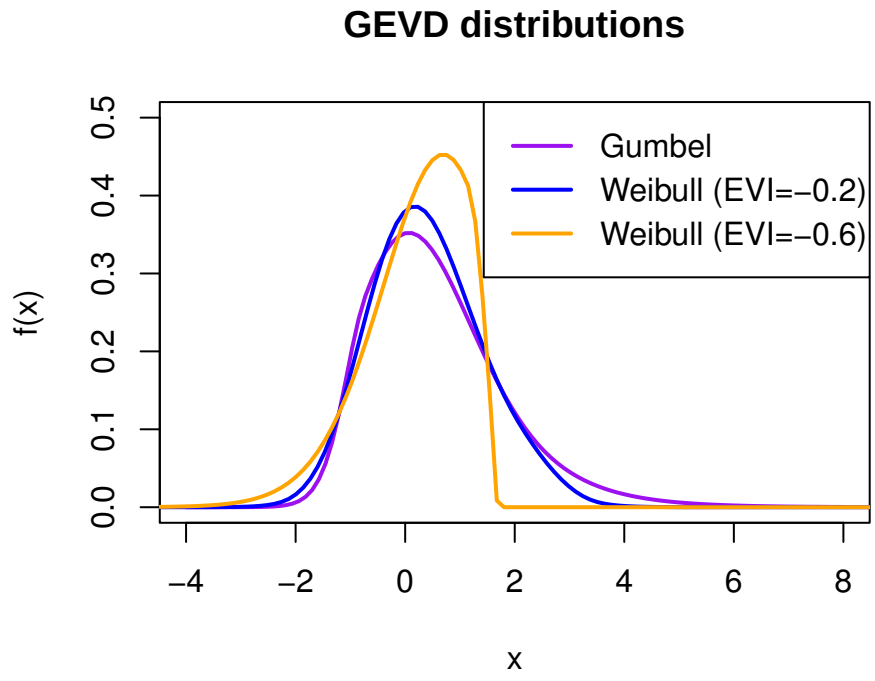


Figure 2.2: Gumbel and Weibull class distributions.

Some examples of distributions that belong in each class include:

- Gumbel class: normal, gamma, and exponential distributions.
- Fréchet class: Student's t, Pareto, and Cauchy distributions.
- Weibull class: uniform and beta distributions.

Nonetheless, Kotz and Nadarajah (2000) highlight that these distributions do not represent distributions of all kinds of extreme value (for instance in small samples). Additionally, they can be used empirically (i.e. they do not have to be used within an extreme value model).

The corresponding probability density functions can be obtained upon differentiating the distribution functions specified in Equations (2.2) and (2.3) as

$$g_\gamma(x) = \begin{cases} (1+\gamma x)^{-1-1/\gamma} \exp\left[-(1+\gamma x)^{-1/\gamma}\right], & \text{for} \quad \gamma \neq 0, \\ \exp\left[-\exp(-x) - x\right], & \text{for} \quad \gamma = 0, \end{cases} \tag{2.4}$$

and

$$g_{\gamma,\mu,\sigma}(x) = \begin{cases} \frac{1}{\sigma}\left(1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-1-1/\gamma} \exp\left[-\left(1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\gamma}\right], & \text{for} \quad \gamma \neq 0, \\ \frac{1}{\sigma}\exp\left[-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right] \exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right), & \text{for} \quad \gamma = 0. \end{cases} \tag{2.5}$$

The choice of GEVD model to fit a set of given data depends on the tail behaviour of the population's distribution – that is if this graphical information is available (Embrechts et al., 1997). More constructively, parameters are estimated using, for example, the maximum likelihood estimation method. Following that, the goodness-of-fit (gof) of the GEVD to the data can be evaluated by means of analysing diagnostic plots such as QQ (Quantile-quantile) plots, probability plots and excess distribution plots (Beirlant et al., 2004). This will be discussed later on in the chapter.

2.3.1 Block Maxima Method

In EVT we can characterise extreme observations using the Block Maxima (BM) method. The Block maxima (BM) approach groups a sample of data (commonly displayed as a time series) into blocks of successive periods of fixed and equal length (e.g. months or years) and identifies the extremes or large values in each block as the maxima (Beirlant et al., 2004). It then fits the generalised extreme value distribution (GEVD) to these maxima, since GEVD asymptotically describes the limiting behaviour of these maxima. Initially developed by Gumbel (1958), De Haan and Ferreira (2006) summarise this method as follows: Given a sample of i.i.d. random variables $X_1, X_2, \dots, X_n$ from an unknown distribution F assumed to belong to the maximum domain of attraction of a GEVD,

1. Divide the sample into m non-overlapping subsamples of r observations each (i.e. m blocks of size r) for given integers $0 < m$, $r < n$.

2. We then consider the maxima in each sub-sample and denote $M_{(r,j)}$ as the maximum of the j^{th} of the subsample.

3. Fit the GEVD model to the sequence of block maxima $M_{(r,1)}, \dots, M_{(r,m)}$ to determine estimates of it's parameters.

Analogously, this same method and thinking can be applied to study the asymptotic distribution of the minima. The following relation would then be required

$\min(X_1, X_2, \dots, X_n) = \max(-X_1, -X_2, \dots, -X_n)$.

This method is the traditional method used to analyse data with seasonality (Gilli and Këllezi, 2006). However, as Abad et al. (2014) state, it does not seem very suitable for application in financial time series mainly because of the clusters commonly observed in financial returns.

The main problem encountered when utilising this method is the question of how many subgroups or blocks to choose. The more blocks we have, the more maxima we have and we risk estimation biasedness. The less blocks we have, on the other hand, the less maxima we have and the larger the variance of estimates will be. This method only uses partial information from the time series, considering that it may happen that one block contains more extremes than another block. This reduces its efficiency as the data is underutilised in modelling the extremes.

2.4 Generalised Pareto Distribution (GPD)

The generalised Pareto distribution is a limiting distribution characterised by two parameters and its distribution function is given by

$$G_{\gamma,\beta}(x) = \begin{cases} 1-\left(1+\frac{\gamma x}{\beta}\right)^{-1/\gamma}, & \text{for} \quad \gamma \neq 0, \text{ i.e. } 1+\frac{\gamma x}{\beta} > 0, \\ 1-\exp\left(-\frac{x}{\beta}\right), & \text{for} \qquad \gamma = 0, \end{cases} \tag{2.6}$$

for $x > 0$ when $\gamma \geq 0$, $0 \leq x \leq \frac{\beta}{\gamma}$ when $\gamma < 0$ and $\beta > 0$, where β is the scale parameter. The shape parameter or tail index,γ, acts as the main factor in determining the tail properties of this distribution (it is the same as the EVI under the GEVD). It indicates the heaviness of the tail – the larger the value of γ, the heavier the tail (Beirlant et al., 2004). For $\gamma = 0$ the tail decreases exponentially, while for $\gamma < 0$ the tail exhibits a finite right end point .

Figures 2.3 and 2.4 illustrate the density plots of the GPD with a negative, positive and zero tail index. It can be seen that the tail heaviness increases gradually as the transition from the negative tail indexes ($\gamma = -0.5$ to $\gamma = -0.2$), to the zero tail index, and finally to the positive tail indexes ($\gamma = 0.2$ and $\gamma = 0.5$).

The related 3-parameter GPD with scale and location parameters is given by

$$G_{\gamma,u,\beta}(x) = \begin{cases} 1-\left(1+\gamma\left(\frac{x-u}{\beta}\right)\right)^{-1/\gamma}, & \text{for} \quad \gamma \neq 0, \\ 1-\exp\left[-\left(\frac{x-u}{\beta}\right)\right], & \text{for} \quad \gamma = 0, \end{cases} \tag{2.7}$$

where u is the threshold, with conditions $x-u \geq 0$ when $\gamma \geq 0$ and $0 \leq x-u \leq -\frac{\beta}{\gamma}$ when $\gamma < 0$ and $\beta > 0$.

The probability density functions (pdfs) are obtained upon differentiating the distribution function specified in Equations (2.6) and (2.7):

$$g_{\gamma}(x) = \begin{cases} \frac{1}{\beta}\left(1+\frac{\gamma x}{\beta}\right)^{-1-1/\gamma}, & \text{for} \quad \gamma \neq 0, \\ \frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right), & \text{for} \quad \gamma = 0, \end{cases} \tag{2.8}$$

and

$$g_{\gamma,u,\beta}(x) = \begin{cases} \frac{1}{\beta}\left(1+\gamma\left(\frac{x-u}{\beta}\right)\right)^{-1-1/\gamma}, & \text{for} \quad \gamma \neq 0, \\ \frac{1}{\beta}\exp\left[-\left(\frac{x-u}{\beta}\right)\right], & \text{for} \quad \gamma = 0. \end{cases} \tag{2.9}$$

Similarly for the GPD, we can assess the gof of the model using various graphical tools which will be explored later.

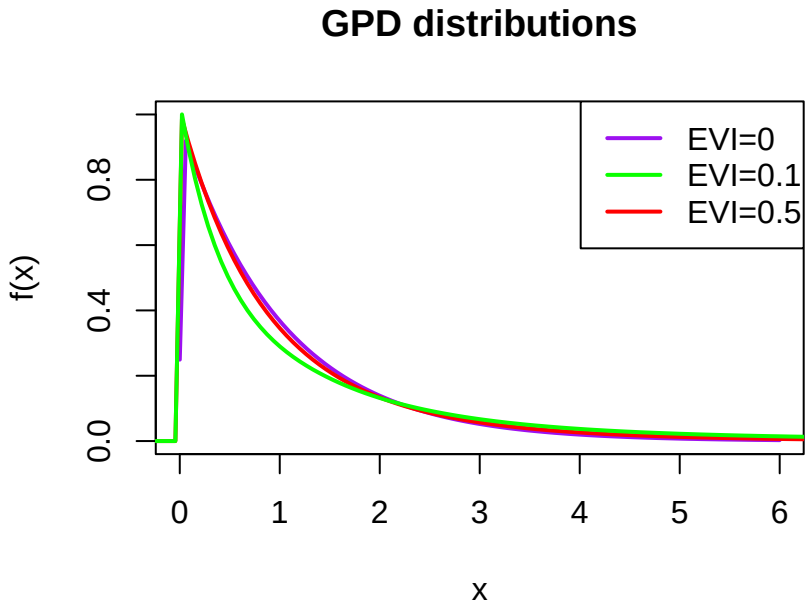


Figure 2.3: GPD with zero and positive tail indexes.

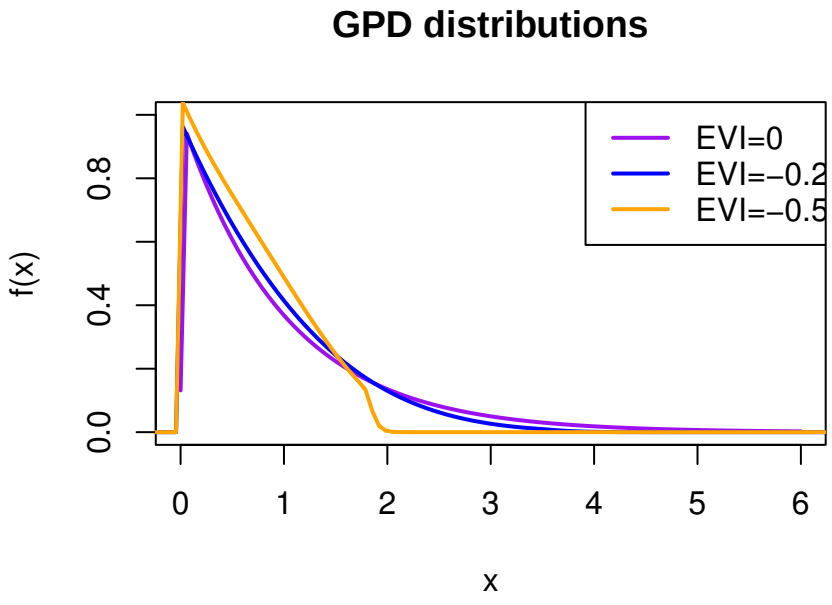


Figure 2.4: GPD with zero and negative tail indexes.

2.4.1 Peaks-Over-Threshold Method

The Peaks-over-threshold (POT) method identifies extremes or large values exceeding a predefined high threshold and models them using the generalised Pareto distribution (GPD); the limiting distribution of these extremes (Beirlant et al., 2004). According to Gilli and Këllezi (2006), the POT method had become the more favoured method in recent applications due to its more efficient use of data for extreme values.

This method aims to estimate the conditional distribution function F_u of the observations of x above a certain threshold u - i.e. F_u describes the distribution of the values exceeding u, given that u is exceeded.

F_u is referred to as the conditional excess distribution function and is defined as

$$F_u(x) = P(X - u \leq x \mid X > u), \tag{2.10}$$

for $0 \leq x \leq u - x_F$, where $x_F < \infty$ is the right endpoint of F and x represents the magnitude of exceedances over u. Furthermore, Equation (2.10) can be written as

$$F_u(x) = \frac{F(x+u) - F(u)}{1 - F(u)} \tag{2.11}$$

with F being the distribution function of X.

Theorem. *Pickands-Balkema-de Haan theorem*

Let $X_1, ..., X_n$ be a sequence of i.i.d. random variables with conditional excess distribution function F_u, then for a large class of underlying distribution functions F, and large enough u, F_u is approximately a GPD,

i.e. for $u \to \infty$,

$$F_u(x) \longrightarrow G_{\gamma,\beta}(x)$$

(Pickands, 1975) and (Balkema and de Haan, 1974).

Ren and Giles (2010) emphasise that this theorem essentially implies that, provided that the underlying distribution F belongs to the maximum domain of attraction of the GEVD, as the threshold u increases, the distribution function of the exeedances over u is approximated by a GPD. Given a sufficiently high threshold u, the following approximation holds from the theorem:

$$F_u(x - u) \approx G_{\gamma,\beta}(x - u), \text{ for } u \to \infty,$$

where $G_{\gamma,\beta}(x-u) = G_{\gamma,u,\beta}(x)$ and the exceedances $x-u \geq 0$.

The GEVD gives the limiting distribution of normalised maxima, while the GPD describes the limiting distribution of the exceedances over some high threshold. Furthermore, the key correlation is that if the maxima follow a GEVD, then the exceedances above a certain high threshold will follow an associated GPD. Accordingly, the parameters of the GPD of the exceedances over threshold u are determined by the parameters of the associated GEVD of the maxima (Ren and Giles, 2010).

This fundamental theorem by Balkema and de Haan (1974) and Pickands (1975) describes the limiting distribution of these exceedances (X) above u as a GPD. Therefore, for a sufficiently large u, F_u is approximately a GPD.

i.e. for $u \to \infty$,

$$F_u(x) \approx G_{\gamma,\beta}(x). \tag{2.12}$$

For $x-u \geq 0$, the excess distribution function Equation (2.11) can therefore be written as:

$$F_u(x-u) = \frac{F(x)-F(u)}{1-F(u)}, \tag{2.13}$$

and thus the expression reduces to

$$F(x)-F(u) = (1-F(u))F_u(x-u),$$

$$F(x) = (1-F(u))F_u(x-u)+F(u),$$

$$F(x) = (1-F(u))G_{\gamma,\beta}(x-u)+F(u). \tag{2.14}$$

The excess distribution yields Equation (2.13), the distribution of the exceedances over u (for $x-u \geq 0$) is exactly known if F is known. However, Longin and Solnik (2001) state that in most financial applications, the distribution of returns is not exactly known, and hence the distribution of exceedances, $F_u(x-u)$, is not known either. The theory of extremes, through its study of the asymptotic behaviour of the returns (X) determines the limiting distribution of exceedances as $u \to x_F < \infty$.

The application of this method involves fitting a GPD to a given data set by first selecting an appropriate threshold, followed by fitting the GPD function to the data. This holds provided that a sufficiently

high threshold has been chosen. To construct a tail estimator we only need to estimate $F(u)$. $F(u)$ can then be estimated using the method of historical simulation or estimation (through the empirical distribution)

$$F(u) = \left(1 - \frac{N_u}{n}\right), \tag{2.15}$$

where n represents the total sample size and N_u is the number of observations above the chosen threshold u. Applying the Maximum Likelihood estimation method produces an approximation of the GPD, which can then estimate $F_u(x-u)$ (Embrechts et al., 1997). This derives the tail estimator

$$\hat{F}(x) = \frac{N_u}{n}\left(1 - \left(1 + \frac{\hat{\gamma}}{\hat{\beta}}(x-u)\right)^{-1/\hat{\gamma}}\right) + \left(1 - \frac{N_u}{n}\right), \tag{2.16}$$

which simplifies to

$$\hat{F(x)} = 1 - \frac{N_u}{n}\left(1 + \hat{\gamma}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\gamma}}. \tag{2.17}$$

The main disadvantage of this method is the selection of threshold. The threshold basically separates the observations that belong to the center of the distribution and those that belong to the tail. If the threshold is too low, we have a large number of data observations and the asymptotic property of the model may be violated. Moreover, we risk biasedness of estimators as observations that do not belong in the tail are included in the estimation process. If it is too high, our estimates are unbiased since we have only a few exceedances to use in our estimation, however the variance (or standard error) is high, which may induce inefficiencies in our parameter estimation (Ren and Giles, 2010; Chan et al., 2022). As Chikobvu and Jakata (2020) state, the BM method can be more efficient than the POT method when:

- the data set is independent but not identically distributed.
- the data set comes in a state that forms block periods inherently, thus the method is easier to apply, e.g. a data sample of quarterly returns.

Nonetheless, the POT method has less disadvantages than the BM method as its efficiency does not rely in the data being in the form blocks. This makes good use of the data and assists in minimising random error when dealing with extremes, as it utilises information from the entire set of data above some threshold. Additionally, it comes with the added flexibility of choosing the threshold to be any convenient value and not a restricted number of blocks. Essentially, it will always perform better than

the BM method, if not the same (Chikobvu and Jakata, 2020). As suggested by Ren and Giles (2010), a basic strategy is to choose a threshold as low as possible, provided that the limiting approximation of the model is able to produce a satisfactory result.

2.4.1.1 Threshold Selection

According to Gilli and Këllezi (2006), no standard rule or automatic algorithm with a satisfactory performance for selecting a threshold exists. Several exploratory data analyses can be executed, in R (R Development Core Team, 2020) for instance, to assist in selecting a threshold. We can analyse a QQ (Quantile-Quantile) -plot against an exponential distribution to try and determine the heavy-tailedness of the data compared to that of the exponential distribution. In order to select an appropriate threshold that is sufficiently high, the empirical Mean Excess (ME) plot can be utilised and examined. The mean excess function is defined as

$$e(u) = E(X - u \mid X > u), \tag{2.18}$$

i.e. the mean or expected value of exceedances over a threshold u, for $u < x_F$ (Embrechts et al., 1997).

If the underlying distribution of $X > u$ follows a GPD, then the corresponding ME is given by

$$e(u) = \frac{\beta + \gamma u}{1 - \gamma}, \tag{2.19}$$

provided that $\gamma < 1$ so that $e(u)$ exists, and $\beta + \gamma u > 0$.

From Equation (2.19) it is clear that the ME function is linear in u. To be precise, $X > u$ follows a GPD if, and only if, the ME function is linear in u (Ren and Giles, 2010). This property then provides a method of selecting an appropriate threshold.

Given a set of data we define the empirical mean excess function as

$$e(u) = \frac{\sum_{i=1}^{n} (X_i - u) I_{(X_i > u)}}{\sum_{i=1}^{n} I_{(X_i > u)}} \tag{2.20}$$

where n represents the total number of observations in the sample, and I is the indicator function on the event $\{X_i > u\}$ (Chinhamu et al., 2017).

The empirical mean excess plot graphically represents $(u, e(u))$ and the random variable X follows a GPD if, and only if, Equation (2.20) is linear in u. Thus, upon examining the plot, we select a u such that the $e(u)$ is approximately linear for $X > u$, (Embrechts et al., 1997). More precisely, if the plot has a positive slope and approximately displays a straight line pattern above a certain level (threshold), we can safely assume that the data follows a GPD above that particular level. If the plot looks rather flat,

then the underlying distribution resembles an exponential distribution, and if it displays a decreasing pattern (negative slope), then a Weibull would seem more appropriate (Panjer, 2006).

However, we may identify several patterns of this kind when analysing the mean excess plot, therefore it is imperative that we test the various levels we identify as thresholds. Wentzel and Mare (2007) also state that choosing a lower threshold will result in a better fit but at the risk of losing the extreme behaviour of the tail, while a higher threshold results in a rather inadequate fit of the model.

Mean residual life (MRL) plots can also assist in threshold selection. They plot the mean excess value (over a preset threshold) for a given set of threshold values. The plots represent the set of points: $\left\{ u \ , \ \frac{1}{N_u} \sum_{i=1}^{N_u} (x_i - u) : u < x_{n,n} \right\}$, where $x_1, \ldots, x_{N_u}$ represent the N_u exceedances. In essence, the MRL plot ought to be approximately linear above a certain threshold for which the GPD model is valid to fit to the excesses.

2.4.1.2 Optimal Threshold Selection: A Diffferent Approach

Longin and Solnik (2001) proposed a different approach to evaluating the optimal number of return exceedances, and thus the optimal threshold, by optimising the trade-off between bias and inefficiency. They implement a Monte Carlo simulation method, which incorporates the mean squared error (MSE) criterion that allows the consideration of the two effects of bias and inefficiency. This method assumes, for convenience, that returns are described by the simple Student's t model with degrees of freedom k. The GPD tail index parameter (γ) is related to k by $\gamma = \frac{1}{k}$.

To compute the optimal positive threshold and optimal number of exceedances, the following simulation procedure is implemented:

1. Let $k = 1$, generate $S = 500$ simulations of N (total sample size) returns from the Student's t distribution with k degrees of freedom.

2. For the chosen value of k, consider a range of exceedance values $n = 0.0005 \cdot N, 0.001 \cdot N, \ldots , 0.2 \cdot N$. For each n, compute S estimates of the tail index, $\hat{\gamma}(n,k)$, using the largest n returns.

3. For each value of n, compute the MSE of the S tail index estimates as

$$MSE\left(\hat{\gamma}_{s=1,\ldots,S}\right) = \left(\bar{\hat{\gamma}} - \gamma\right)^2 + \frac{1}{S} \sum_{s=1}^{S=500} \left(\hat{\gamma}_s - \gamma\right)^2, \tag{2.21}$$

   with $\bar{\hat{\gamma}}$ being the mean of the simulated S observations. The first part of equation (2.21) measures the bias, while the latter expresses the inefficiency.

4. Following that, obtain the optimal number of exceedances $n^*(k)$ as the value of n that minimises the MSE for the preselected value of k.

5. Repeat steps 1-3 for additional values $k = 2, ..., 10$, until the optimal number of exceedances $n^*(k)$ have been obtained for all values $k = 1, ..., 10$. This range of k values is selected to consider different degrees of tail heaviness - since $\gamma = \frac{1}{k}$, the lower the degrees of freedom, the heavier the tail distribution.

6. For each of the optimal number of exceedances, compute the tail index paramter estimate $\hat{\gamma}(n^*(k))$ using the $n^*(k)$ largest returns.

7. The true optimal number of exceedances for the returns distribution, denoted $n*$, is given by the $n^*(k)$ value for which $\hat{\gamma}(n^*(k))$ is closest to $\gamma = \frac{1}{k}$.

8. The associated optimal threshold is then given as the $1 - \left(\frac{n^*}{N}\right)$ quantile of the returns distribution, where the optimal threshold u represents this value as a percentage above or below the mean value (the true returns) in the estimation model.

This process is also applied and outlined in detail by Chan et al. (2022), in their analysis of return and volume exceedances based on the POT method. Once the threshold is selected, the next step is estimating parameters for the GPD model.

2.5 Parameter Estimation Techniques

Several techniques have been developed to estimate the parameters of both the GEVD and GPD, each useful in forecasting the returns or losses and their confidence intervals.

2.5.1 Maximum Likelihood Estimation (MLE) Method

The MLE is a widely used estimator due to its general simplicity and reliability in statistical models (Panjer, 2006). In addition, this estimator is very practical as it is asymptotically normal - allowing approximations for standard errors and confidence intervals (Ren and Giles, 2010). The estimation procedure can be implemented for both the GEVD and GPD.

The likelihood function is defined as

$L(\mathbf{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$ for a random sample $X_1, ..., X_n$ with pdf $f(\mathbf{x}, \theta)$.

The resulting log likelihood function is

$l(\mathbf{x}, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta)$.

The following partial derivatives are solved to determine the value of the parameter $\hat{\theta}$ that maximises the log likelihood:

$\frac{\partial}{\partial \theta_i} l(\mathbf{x}, \theta) = 0.$

2.5.1.1 MLE for GPD

Provided we have a sufficiently high threshold u, and assuming there are N_u observations satisfying $x_i - u \geq 0$ with sub-sample $\{x_1 - u, \ldots, x_{N_u} - u\}$ following a GPD for $\gamma \geq 0$, $0 \leq x_i - u \leq \frac{-\beta}{\gamma}$ for $\gamma < 0$ (Beirlant et al., 2004). The logarithm of the probability density function of x_i can be derived from equation (2.9) as

$$\ln f(x_i - u) = \begin{cases} -\ln(\beta) - \frac{1+\gamma}{\gamma} \ln\left(1 + \gamma\left(\frac{x_i - u}{\beta}\right)\right), & \text{for} \quad \gamma \neq 0, \\ -\ln(\beta) - \frac{1}{\beta}(x_i - u), & \text{for} \quad \gamma = 0. \end{cases}$$

Thus, the log-likelihood function for the GPD is given by the logarithm of the joint density of the N_u observations

$$L(\gamma, \beta \mid \mathbf{x} - u) = \begin{cases} -N_u \ln(\beta) - \frac{1+\gamma}{\gamma} \sum_{i=1}^{N_u} \ln\left(1 + \gamma\left(\frac{x_i - u}{\beta}\right)\right), & \text{for} \quad \gamma \neq 0, \\ -N_u \ln(\beta) - \frac{1}{\beta} \sum_{i=1}^{N_u} (x_i - u), & \text{for} \quad \gamma = 0. \end{cases}$$

We can then maximise $L(\gamma, \beta \mid \mathbf{x} - u)$ for our sample with a suitable threshold u to obtain the estimates of γ and β.

According to Beirlant et al. (2004), the most optimal way to maximise $L(\gamma, \beta \mid \mathbf{x} - u)$ is through the following reparametrisation:

$(\gamma, \beta) \to (\gamma, \tau)$, with $\tau = \frac{\gamma}{\beta}$.

Leading to the pdf

$$g(x_i - u) = \frac{\tau}{\gamma}(1 + \gamma(x_i - u))^{-1-1/\gamma},$$

and

$$L(\gamma, \tau \mid \mathbf{x} - u) = -N_u \ln(\gamma) + N_u \ln(\tau) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{N_u} \ln(1 + \tau(x_i - u)).$$

The MLEs for $\hat{\gamma}$ and $\hat{\tau}$ then follow from solving the partial derivatives:

$$\frac{1}{N_u} \frac{\partial L(\gamma, \tau \mid \mathbf{x} - u)}{\partial \hat{\tau}} = \frac{1}{\hat{\tau}} - \left(\frac{1}{\hat{\gamma}} + 1\right) \frac{1}{N_u} \sum_{i=1}^{N_u} \frac{(x_i - u)}{1 + \hat{\tau}(x_i - u)} = 0,$$

$$\frac{1}{N_u} \frac{\partial L(\gamma, \tau \mid \mathbf{x} - u)}{\partial \hat{\gamma}} = \frac{-1}{\hat{\gamma}} + \frac{1}{\hat{\gamma}^2} \frac{1}{N_u} \sum_{i=1}^{N_u} \ln(1 + \tau(x_i - u)) = 0$$

yielding

$$\hat{\gamma}_{ML} = \hat{\gamma} = \frac{1}{N_u} \sum_{i=1}^{N_u} \ln\left(1 + \hat{\tau}_{ML}\left(x_i - u\right)\right).$$

Beirlant et al. (2004) further highlight that the MLE is only applicable for a GPD model with $\gamma > \frac{-1}{2}$. This is due to the fact that for smaller negative values of γ, the density of the GPD depends on the parameter values.

2.5.1.2 MLE for GEVD

Assuming $X_1, ..., X_m$ are the sample maxima that follow a GEVD, the log-likelihood for the case $\gamma \neq 0$ is given by:

$$L(\gamma, \mu, \sigma) = -m \ln \sigma - \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^{m} \ln\left(1 + \gamma\left(\frac{x_i - \mu}{\sigma}\right)\right) - \sum_{i=1}^{m} \left(1 + \gamma\left(\frac{x_i - \mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}} \tag{2.22}$$

provided $1 + \gamma\left(\frac{x_i - \mu}{\sigma}\right) > 0$, for $i = 1, \ldots, m$.

For parameter combinations in which the above condition is violated, i.e. if at least one of the data observations falls beyond the endpoint of the distribution, then the likelihood takes on the value 0 and the log-likelihood approaches $-\infty$ (Chinhamu et al., 2017), i.e. $L(\gamma, \mu, \sigma) = 0$ and $l(\gamma, \mu, \sigma) \approx -\infty$.

When $\gamma = 0$, the Gumbel limit of the GEVD needs to be incorporated to obtain the log-likelihood

$$L(0, \mu, \sigma) = -m \ln \sigma - \sum_{i=1}^{m} \frac{x_i - \mu}{\sigma} - \sum_{i=1}^{m} \exp\left\{-\left(\frac{x_i - \mu}{\sigma}\right)\right\}, \tag{2.23}$$

to obtain the MLEs $(\hat{\gamma}, \hat{\mu}, \hat{\sigma})$ for the GEVD family (Coles et al., 2001).

Equations (2.22) and (2.23) are maximised with respect to the three parameter vector (γ, μ, σ) in order to attain the Maximum Likelihood Estimates of the GEVD fanily. The MLE technique offers the added advantage of estimating all of the parameters of a distribution simultaneously.

2.5.2 Hill Estimator

The Hill estimator is the most basic and popular tail index estimator which only applies to the Fréchet class with $\gamma > 0$ (Hill, 1975) (i.e. distributions with heavy tails). It is a non-parametric method that does not require assumptions about the underlying distributions. Recall, the sample $X_1, X_2, \ldots, X_n$ with N_u exceedances above the threshold, can be rearranged into order statistics $X_{1,n} \leq \ldots \leq X_{n,n}$.

This estimator is represented by the formula:

$$H_{N_u,n} = \hat{\gamma}_H = \frac{1}{N_u} \sum_{i=1}^{N_u} \log X_{n-i+1,n} - \log X_{n-N_u,n}. \tag{2.24}$$

This form of the Hill estimator is inspired by the quantile plots of Pareto-type distributions. It carries the advantage of simplicity with easy calculation, as well as generating efficient estimates when $\gamma > 0$. Despite its disadvantage of being restricted to the $\gamma > 0$ case, it has been studied extensively and widely used in practice (Beirlant et al., 2004). Its prime disadvantage is in the choice of N_u, as every choice of N_u produces a different estimate for γ. This creates a problem when applying this estimator in practice as specific guidelines on how to choose the value N_u would be necessary.

Furthermore, a severe bias can occur for large values of N_u. Beirlant et al. (2004) elaborates that a large bias results in poor coverage probabilities for confidence levels, and overestimation and underestimation lead to large inefficiencies in practice. The third drawback is that this estimator is not invariant to shifts of the data. However, under the assumption of i.i.d. data, the Hill estimator is asymptotically normal.

The next form of the Hill estimator stems from the probability view. As previously mentioned, this estimator assumes that the data is heavy-tailed (particularly from the Fréchet class). Recall, in Section 2.4.1 it was shown that the GPD can be used to approximte the marginal distribution of the exceedances.

For $u \to \infty$, $F_u \approx G_{\gamma,\beta}(x)$.

An arbitrary approximation of the GPD can be written as

$G(x) \approx 1 - cx^{-1/\gamma}$,

for some constant c. Using Equations (2.14) and (2.15), the constant c can then be derived by assuming the tail estimator

$\hat{F}(u) = F(u)$.

Where $F(u) = 1 - \frac{N_u}{n}$.

This yields an expression similar to Equation (2.17), since $F(u) = 1 - \frac{N_u}{n}$:

$$\hat{F}(x) = 1 - \frac{N_u}{n} \left(\frac{x}{X_{N_u}} \right)^{-1/\hat{\gamma}}. \tag{2.25}$$

This is the estimated distribution of the relative excesses $\frac{x_i}{X_{N_u}}$ above some threshold u, conditional on $x > X_{N_u} = u$. N_u represents the number of observations exceeding the threshold $X_{N_u} = u$ (Beirlant et al., 2004).

Conditional on N_u, the likelihood is

$L\left(\frac{\mathbf{x}}{u}, \gamma\right) = -\prod_{i=1}^{N_u} \log \gamma - \left(1 + \frac{1}{\gamma}\right) \log\left(\frac{x_i}{u}\right)$,

the log-likehood becomes

$l\left(\frac{\mathbf{x}}{u},\gamma\right) = -N_u \log\gamma - \left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{N_u}\log\left(\frac{x_i}{u}\right)$.

Differentiating and equating to zero leads to

$$\begin{aligned}\frac{d}{d\gamma}\log L\left(\frac{\mathbf{x}}{u},\gamma\right) &= \frac{-N_u}{\gamma}+\frac{1}{\gamma^2}\sum_{i=1}^{N_u}\log\left(\frac{x_i}{u}\right),\\ &= 0,\end{aligned}$$

resulting in

$$\begin{aligned}\frac{N_u}{\gamma} &= \frac{1}{\gamma^2}\sum_{i=1}^{N_u}\log\left(\frac{X_{i,n}}{u}\right)\\ \hat{\gamma} &= \frac{1}{N_u}\sum_{i=1}^{N_u}\log\left(\frac{X_{i,n}}{u}\right) = \hat{\gamma}_H.\\ \hat{\gamma}_H &= \frac{1}{N_u}\sum_{i=1}^{N_u}\log\left(X_{i,n}\right)-\log\left(X_{N_u}\right)\\ &= \frac{1}{N_u}\sum_{i=1}^{N_u}\log\left(X_{i,n}\right)-\log\left(u\right).\end{aligned}$$

Thus, selecting an upper order statistic, X_{N_u},as the threshold yields the Hill's estimator for the tail index once more (Beirlant et al., 2004).

2.5.3 Moment Estimator

The moment estimator is an adaptation of the Hill estimator ($\hat{\gamma}_H$) and it is represented as:

$$\hat{\gamma}_M = H_{N_u,n} + 1 - \frac{1}{2}\left(1 - \frac{H^2_{N_u,n}}{H^{(2)}_{N_u,n}}\right)^{-1},$$

where

$$H^{(2)}_{N_u,n} = \frac{1}{k}\sum_{i=1}^{N_u}\left(\ln X_{n-i+1,n} - \ln X_{n-N_u,n}\right)^2.$$

This estimator is very similar to the Hill estimator, except it is applicable to all values of γ, i.e. $\gamma \varepsilon \mathbb{R}$ (Beirlant et al., 2004).

2.6 Tail-Related (extreme) Risk Measures

The events that occur under extreme market conditions are the most alarming for risk analysts as well as investors. Some examples of these events include market crashes in the stock market and currency crises. These events bring forth significant unforeseen losses which can lead to bankruptcy (Chinhamu et al., 2017), or severe financial strain on market participants. For instance, consider extreme changes in the prices of a significant commodity or cryptocurrency prices. EVT is able to model such extreme events and compute their related measures of risk – accentuating it's importance in finance.

Financial risk management serves a very significant motive that seeks to answer pertinent questions relating to the potential losses of future results (Righi and Ceretta, 2015). It entails pricing and managing risks associated with changing market conditions, and the estimation of extreme quantiles and tail-related risk measures. Application in finance often involves large data sets which have attracted a variety of quantitative risk analyses over the years which were previously considered impractical to implement (Park and Kim, 2016), such as estimating large but rare losses existing in the tails of distributions.

In banking, for instance, risk management includes the systems banks use to protect themselves against the risk of financial loss that may be caused by a decline in the prices of the financial assets that they hold or issue out. They then investigate and model the relative differences in the consecutive prices of these assets; and this relative quantity would be considered the returns on this financial asset (Beirlant et al., 2004).

However, Park and Kim (2016) state that obtaining accurate estimates of these extreme quantiles usually requires generating rather large samples of financial losses, and this may be quite time-consuming and the calculations too computationally demanding as they take a large amount of time and memory storage. This means efficiency is thus not achieved as it takes longer to obtain the necessary extremes.

After estimating parameters the next step is to apply a measure of risk. Two of the most prominent

measures of risk in financial markets are Value-at-Risk (VaR) and expected shortfall (ES). These two measures assist in estimating extreme quantiles – i.e. determining the value that a given financial variable exceeds with a given low probability (Chinhamu et al., 2015b). Chinhamu et al. (2017) emphasise that the accuracies of VAR and ES estimates are dependent on how well our chosen model fits the extreme observations.

2.6.1 Generalised Autoregressive Conditional Heteroscedasticity (GARCH) Conditional Models

Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models are popular stochastic volatility models used to model the volatility of returns. These models are used to account for the time varying nature of the mean and the variance of returns (Dicks et al., 2014). When estimating risks, modelling the volatility of extreme returns is very significant as financial markets can be highly volatile. Extreme returns and losses constitute market risk, thus risk models need to be as accurate as possible to ensure efficient predictions.

For a given portfolio with return distribution R_t, a GARCH model specifies two equations:

$R_t = \mu_t + \varepsilon_t$

$\varepsilon_t = Z_t \sigma_t$

R_t represents the conditional mean equation which describes the behaviour of the returns in accordance with past returns, while ε_t represents the conditional variance equation which depicts the evolving volatility behaviour of the returns (Bee et al., 2016). Z_t follows either a standard normal conditional error distribution or a standardised Student's t conditional error distribution. It follows that:

$E(R_t) = \mu_t$ and $Var(R_t) = \sigma_t^2$.

Unfortunately, as Dicks et al. (2014) highlight, the main limitation exhibited in the practical application of these models is that they fail to converge during periods of extreme return observations in the market. They further point out that this issue is not adequately highlighted in the current literature. Fortunately, EVT models are able to solve this convergence issue as they focus specifically on modelling such extreme returns. For this reason, Dicks et al. (2014) proposed a novel approach in their study of VaR and ES forecasting. Their estimation approach augmented various GARCH models with EVT models and subsequently applied it to the JSE Financial Index (J580) to forecast extreme risk over the entire sample period, including periods where the GARCH models don't converge. They concluded that, although it is evident that no single forecasting model is universally optimal, their models provided impressive results for the data under consideration.

2.6.2 Value-at-Risk (VaR)

Value-at-Risk evaluates the maximum possible loss in the market value of a financial portfolio over a specified period of time (given a certain level of confidence: $100(1-p)\%$ level or probability p), and does so by focusing on the tails (Beirlant et al., 2004). Essentially, as Ren and Giles (2010) explain, it is a high quantile on the distribution of the returns of some risky financial instrument or the value below which the portfolio will drop with only a very small probability.

Abad et al. (2014) describe VaR as the worst expected loss under normal market conditions. They explain using an example of a bank that has a trading portfolio with a VaR of \$1 million at the 99% confidence level. This means that, given that the market is operating under normal market conditions, the daily financial loss will exceed \$1 million only 1% of the time. Or the probability that the percentage loss won't exceed the VaR is measured by $100(1-p)\%$.

VaR has become a widely-accepted benchmark for calculating market risks and solvency in finance. Investors utilise it when trying to evaluate and forecast the impact of disadvantageous events that may in reality be worse than what was observed during the period for which relevant data was available (Beirlant et al., 2004).

From a regulatory perspective, Park and Kim (2016) then proceed to make the example of banks being required to hold credit risk capital that ought to be able to absorb unexpected financial losses with at least 99.9% under the the Internal Ratings Based (IRB) approach. This equates to a VaR at a 99.9% confidence level. This kind of regulation endorses the use of VaR as a fundamental market risk management tool in finance. This 99.9% VaR can be used to assess extreme events that occur with 0.0001 probability.

According to Abad et al. (2014), Basel I (also called the Basel Accord) is a very significant agreement that was concluded by several central banks from multiple countries worldwide in 1988. It stipulates recommendations on banking regulations involving market and credit risk. Essentially, it aims to ensure that financial institutions hold sufficient capital to meet their obligations as well as cover unforeseen financial losses. By proposing the Basel I Accord, regulators had realised the substantial risk posed to a banking system by a lack of sufficient capital to finance huge sudden losses in capital markets.

Evaluating market risks is highly imperative as financial institutions are inclined to take interest in measuring the amount of losses they are likely to incur when the prices of their portfolio assets drop. As a result, it is only expected that the existing literature on VaR would multiply in the proposal and development of new approaches to model VaR, as well as the comparison of these models. This risk measurement comes with the following advantages: its a simple and universal concept, and its widely applicable (Righi and Ceretta, 2015). However, its not easy to calculate and a range of methodologies for its calculation have been developed.

VaR can be calculated from financial data using three main approaches: the parametric approach

(which uses the proposed distribution with estimated parameters), the non-parametric approach (which uses historical VaR values) and the semi-parametric approach (which uses EVT and Monte Carlo simulation).

Let X be a random variable (suppose X models the return distribution of some risky financial variable) with cumulative distribution function F over a specific time horizon.

For a chosen small probability p, the $100(1-p)\%$ VaR can be defined as the negative of the p^{th} quantile of F:

Let x_t be some quantile located in the far left of the distribution of X - i.e. x_t has a negative value (a loss),

$$VaR_p = -x_t,$$

$$P(-X > VaR_p) = p$$

Recall that negative returns (losses) can be modelled as $-X$, this means

$P(X < x_t) = p = F(x_t)$

This is when VaR is defined as a positive value, where F^{-1} is the quantile function (the inverse function of the distribution function F). It can also be written as $VaR_p = Q_p(X)$.

The probability that the percentage loss (a positive number) will not exceed the VaR_p is $100(1-p)\%$. And likewise, $100(1-p)\%$ gives the probability that a return will not fall below $-\mid x_t \mid$ (Dicks et al., 2014).

However, defining VaR as a negative value in EVT modelling is more appropriate as it becomes easier to simplify the mathematics. This allows one to work with the far right tail of the return distribution.

Now, VaR is defined as the negative of the p^{th} quantile, i.e.

$$VaR_p = \inf\{x \in \mathbb{R} \mid P(X < -x_t) \leq 1-p\},$$

where x_t is a positive number and

$$VaR_p = -x_t.$$

Thus,

$$\begin{aligned} 1-p &= P(X < -x_t) \\ &= P(X < VaR_p). \end{aligned}$$

In terms of the quantile function, VaR can be written as

$$\begin{aligned} VaR_p &= F^{-1}(1-p) \\ F(VaR_p) &= 1-p \end{aligned}$$

More comprehensively, McNeil and Frey (2000) define VaR (in relation to GARCH models) as:

Let $X_1, ..., X_n$ be i.i.d. random variables representing the financial returns. F is conditional on the set of information Ω_{t-1} available at time $t-1$, and is denoted by $F(x) = P(X \leq x \mid \Omega_{t-1})$.

Assume $\{X_t\}$ follows the stochastic process:

$X_t = \mu + \varepsilon_t$

$\varepsilon_t = Z_t \sigma_t$, $Z_t \sim$ i.i.d. (0, 1)

where μ and σ_t are the conditional mean and variance respectively. Z_t is an i.i.d. process with mean equal to zero and variance equal to 1.

$\sigma_t^2 = E\left(Z_t^2 \mid \Omega_{t-1}\right)$ and Z_t has the conditional distribution function $G(z)$, where

$G(z) = P(Z_t < z \mid \Omega_{t-1})$.

For a given probability p, VaR can be defined as the $100p^{\text{th}}$ quantile of the distribution function of financial returns

$$F(VaR_p) = P(X_t < VaR_p) = p$$

or

$$VaR_p = F^{-1}(p).$$

Abad et al. (2014) state that there are two ways of estimating this quantile:

1. Invert distribution function $F(x)$ to get $VaR_p = F^{-1}(p)$, or
2. invert with respect to $G(z)$ to get $VaR_p = \mu + \sigma_t G^{-1}(p)$.

Consequently, to model VaR, it is necessary to specify $F(x)$ or $G(z)$, as well as estimate σ_t^2. These distribution functions can be estimated using parametric, non-parametric and semi-parametric methods. From the definitions above, one can see that this risk measure is a very theoretical concept. Hence, in order to apply it in the real world, it is necessary to utilise one of the aforementioned estimation methods to compute its values.

2.6.2.1 Parametric Approach: RiskMetrics

Under this approach, risk is measured by fitting probability curves to the data and subsequently drawing inference from them on the VaR (Abad et al., 2014). The most common model to estimate VaR under this approach is the RiskMetrics model.

This method traditionally assumes that the underlying financial returns (X_t) follow a normal distribution. Given this assumption, the VaR of an asset or portfolio at a $(1-p)\%$ confidence level is calculated as

$$VaR_p = \mu + \sigma_t G^{-1}(p),$$

where $G^{-1}(p)$ represents thep^{th} quantile of the standard normal distribution, and σ_t is the conditional standard deviation of the returns (Abad et al., 2014).

Basically, from the assumption $X_t \sim$ i.i.d. $N\left(\mu, \sigma^2\right)$, and since $VaR_p = -x_t$ and $P(X_t < x_t) = p$, it follows that

$$P\left(\frac{X_t - \mu}{\sigma} < \frac{x_t - \mu}{\sigma}\right) = p,$$

this standardisation results in $Z_t \sim N(0,1)$,

$$P\left(Z_t = \frac{X_t - \mu}{\sigma} < \frac{x_t - \mu}{\sigma}\right) = p$$

and

$$\frac{x_t - \mu}{\sigma} = \Phi^{-1}(p)$$

yielding

$$\begin{aligned} x_t &= \sigma\Phi^{-1}(p) + \mu = -\sigma\Phi^{-1}(1-p) + \mu \\ \therefore VaR_p &= -x_t \\ &= -\sigma\Phi^{-1}(p) - \mu = \sigma\Phi^{-1}(1-p) - \mu, \end{aligned}$$

where $\Phi^{-1}()$ is the inverse function of the cdf of the standard normal distribution.

VaR is computed by first fitting a normal distribution to the data, estimating the mean μ and standard

deviation σ, and then analytically calculating

$$VaR_p = \gamma + \sigma \cdot Q_{norm}(p).$$

The biggest disadvantage is the normality assumption for the returns. Unfortunately, as previously discussed, this normality assumption is not realistic, which limits the effectiveness of this approach. As confirmed by empirical evidence, financial returns are not normally distributed due to the heavy-tail, excessive kurtosis and skewness features often present in financial data but not in a normal distribution (which is conveniently symmetric). Therefore, the true size of the losses is greater than that predicted by a normal distribution (Righi and Ceretta, 2015). For this reason, the inference drawn about the future levels of this variable are neither accurate nor reliable for risk management.

The second disadvantage is the assumption of i.i.d. returns. A considerable amount of empirical evidence invalidates this assumption. The third flaw is found in the estimation of the conditional volatility σ_t^2 of the returns. The model used for estimation accounts for cluster volatility and varying volatility but not asymmetry. Moreover, it fails to model the persistence of volatility as adequately as the GARCH family models (Abad et al., 2014).

These shortcomings have instigated more research into the development of more sophisticated and improved volatility models for financial returns, a few are considered in Abad et al. (2014)'s review of VaR methodologies: stochastic volatility, the GARCH and realised volatility-based models. It also includes the investigation of alternative density functions that accommodate the asymmetry, peakness and fat tails of financial returns. The Student's t, which exhibits heavier tails than the normal distribution, has been commonly used in financial risk management for asset returns. Corlu and Corlu (2015) modelled the distribution of several exchange rates using various flexible distributions, including the skewed Student's t distribution.

The empirical results on the performance of this distribution in estimation are inconclusive, as some papers indicate that it performs better than the normal distribution, while some show that it overestimates the proportion of extremes (Abad et al., 2014). The main drawback here is that the Student's t distribution, although it captures leptokurtic behaviour (a higher kurtosis), it does not account for the skewness when modelling the returns. Righi and Ceretta (2015) propose the use of the skewed Student's t distribution that uses a non-centrality parameter in modelling returns.

2.6.2.2 Non-Parametric Approach: Historical Simulation

This approach entails a historical simulation method and evaluates VaR without making stringent assumptions about the distribution of returns. The aim is to let the data speak for itself as much as possible, without assuming some theoretical distribution, in order to obtain a good estimate of VaR (Abad et al., 2014). It is contingent on the notion that specific values can be predicted based on the

estimation of a range for the future values, after which the VaR can be calculated using these values (Ren and Giles, 2010).

In essence, Abad et al. (2014) explains that for us to able to predict the size of risk in the near future using data from the recent past, we will need to assume that the near future is adequately similar to the most recent past. Historical simulation (HS) utilises the empirical distribution of the returns to approximate $F(x)$, and hence VaR_p is measured by the p^{th} quantile of this distribution.

i.e. $VaR_p = F^{-1}(p)$.

Nonetheless, this approach is hindered by its own limitations. The major downside of this method is that the results are entirely dependent on our data set. Abad et al. (2014) state that if our data period is unusually quiet, this method is likely to underestimate risk, if it's unusually volatile, it's likely to overestimate it. Furthermore, HS is often inefficient in reflecting and reacting to extreme events and volatility changes. For instance, the rise in financial risk caused by an unanticipated market turbulence. These drawbacks result in large errors in our risk estimates (Righi and Ceretta, 2015).

However, it is conveniently easy to implement and its lack of parametric assumptions means it allows for skewness, fat tails and any other non-normal features in financial observations. Nevertheless, according to Righi and Ceretta (2015), this method has been reported to estimate VaR inaccurately. This is due to the more recently developed VaR methodologies that incorporate advanced and volatility models. These kind of models offer improved volatility forecasts and more accurate estimations of tail-related risk measures (Bee et al., 2016).

Ren and Giles (2010) report that the VaR estimation method based on the extreme value approach is more valuable and useful in identifying extreme risk than the two above mentioned traditional methods. This method is discussed under the next approach.

2.6.2.3 Semi-Parametric Approach

These methods comprise of a combination of the parametric and non-parametric approaches. Two important methods will be considered: the Extreme Value Theory approach and the Monte Carlo (MC) simulation method, as well as the Hill estimator method.

Extreme Value Theory approach

Consider the previously explained GPD model associated with the Peaks-over-Threshold method as well as the GEVD model associated with the Block Maxima method.

For a small upper tail probability p (with $p > \hat{F}(u)$ and corresponding tail quantile $q = 1 - p$), by inverting the tail estimator in Equation (2.17) we can obtain the following quantile estimator or tail quantile:

$$\hat{x}_p = u + \frac{\hat{\beta}}{\hat{\gamma}}\left(\left(\frac{n}{N_u}p\right)^{-\hat{\gamma}} - 1\right)$$

where $\hat{\beta}$ and $\hat{\gamma}$ are the MLE estimates under the GPD model. And since VaR is precisely the extreme quantile, we can estimate it by

$$\hat{VaR}_p = \begin{cases} u + \frac{\hat{\beta}}{\hat{\gamma}}\left(\left(\frac{n}{N_u}(p)\right)^{-\hat{\gamma}} - 1\right), & \text{for} \quad \gamma \neq 0, \\ u - \hat{\beta}\log\left(\frac{n}{N_u}(1-p)\right), & \text{for} \quad \gamma = 0. \end{cases} \tag{2.26}$$

Similarly, for the GEVD model we obtain the approximation (Chinhamu et al., 2015b):

$$\hat{VaR}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}\left(1 - (-n\ln(1-p))^{-\hat{\gamma}}\right), & \text{for} \quad \gamma \neq 0, \\ \hat{\mu} - \hat{\sigma}\ln(-n\ln(1-p)), & \text{for} \quad \gamma = 0. \end{cases}$$

This EVT approach is called an unconditional method since neither of the above VaR models account for and reflect the current volatility state in their estimates of VaR. In particular, as emphasised by Park and Kim (2016), tail risk measures are highly sensitive to the estimates of the EVT distribution's parameters (GEVD or GPD), and the more extreme the quantile of interest gets, the larger the volatility of the risk measures.

In order to incorporate volatility models in this approach, McNeil and Frey (2000) proposed an improved methodology for VaR estimation that integrates EVT with volatility models. With EVT estimating the distribution of the tails, they proposed using GARCH models for volatility estimation in view of the fact that most financial return series display conditional heteroscedasticity. Their method is known as the conditional EVT method.

Provided that the financial return time series is strictly stationary and $\varepsilon \sim GPD = G_{\gamma,\beta}(\varepsilon)$, the conditional p^{th} quantile of the returns is estimated by the VaR:

$$\hat{VaR}_p = \mu + \sigma_t G^{-1}_{\gamma,\beta}(p),$$

where σ_t^2 is the conditional variance of the financial returns (Abad et al., 2014). The quantile function of the GPD is evaluated equivalent to the above extreme quantile for the GPD model in Equation (2.26)

$$G^{-1}_{\gamma,\beta}(p) = u + \frac{\hat{\beta}}{\hat{\gamma}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\gamma}} - 1\right).$$

Hill estimator method

Recall that the tail index, γ , depicts the features of the tail distribution; under Hill's estimation technique the tail index is estimated as follows:

$$\hat{\gamma}_H = \left[\frac{1}{u} \left(\sum_{i=1}^{u} log\,(r_i) - log\,(r_{u+1}) \right) \right]^{-1},$$

where r_u is the threshold return and u is the number of observations equal to or less than the threshold return (Abad et al., 2014).

The quantile estimator is provided by:

$$VaR_p = r_{u+1} \left(\frac{1-p}{\frac{u}{\hat{\gamma}_H}} \right)^{-1/\hat{\gamma}_H}.$$

Abad et al. (2014) state that the ultimate issue with this estimator is the lack of any analytical means to select a threshold value of u in an efficient manner. Therefore, an alternative procedure involves calculating and representing the different values of the Hill estimator tail index for different u values. The graphical representation displays the values of $\hat{\gamma}_H$ based on u, and is known as Hill graphics and can be a chart or a graphic. The value of u is chosen from the region where the Hill estimators are relatively stable, that is, where the chart is leaning slightly horizontally. The central idea is: as u increases, the Hill estimator variance decreases, thus increasing the bias. Essentially, we are likely to predict a balance between both trends, and once this level is observed, the estimator will remain constant (Abad et al., 2014).

Monte Carlo simulation method

This approach describes a simple way of estimating the VaR on a certain day t on a one-day time period at a 99% confidence level. Generate N simulations from the returns distribution on day $t+1$ then sort these N simulations and read off the $\left(\frac{N}{100}\right)^{th}$element as the estimate of the VaR. In more simple terms, VaR is estimated empirically as the VaR quantile of the simulated return distribution (Abad et al., 2014).

A significant feature of high-frequency intra-daily returns is the volatility clustering. One can then use GARCH models to generate variance estimates at a certain time t $(\hat{\sigma}_t)$ using the simulated returns. These models provide better predictions and tail estimation (Bee et al., 2016). The key disadvantage of this simulation method is its substantial computational expense due to the large number of simulations

executed. This has also limited its application in solving problems related to financial risk in the real world.

Ren and Giles (2010) state that VaR is a good risk measure, however it does not encapsulate all the aspects of market risks; such as subadditivity. This means that the VaR of a given portfolio is not necessarily less than the VaR of the individual assets that make up that portfolio - classifying it as an incoherent measure of risk (Righi and Ceretta, 2015). In addition, VaR does not take into account the the potential losses that may arise beyond the quantile of interest, so it essentially tells us nothing more than the probability of expecting a loss greater than the quantile (or VaR) itself (Abad et al., 2014).

Ren and Giles (2010) further suggest that a more suitable methodology for measuring risk would be a combination of VaR and other complementary measure mechanisms – such as expected shortfall. Expected shortfall, on the other hand, cautions us about the losses to expect in a bad state as well as their magnitude and not just the quantile we are interested in (McNeil and Frey, 2000; Righi and Ceretta, 2015).

2.6.3 Expected Shortfall (ES)

Expected shortfall estimates the expected size of losses that exceed a corresponding VaR level, thus it is conditional on the loss being greater than the VaR level. This risk measure also offers the advantage of informing us about the likely magnitude of the potential returns/losses, given that they exceed the VaR (Chinhamu et al., 2017).

Righi and Ceretta (2015) categorise the various risk models used to predict ES under unconditional and conditional models. The unconditional models include, from the estimation approaches discussed under VaR: the RiskMetrics model, the EVT approach and the Historical simulation methods. Righi and Ceretta (2015) refer to them as unconditional because they are not influenced by past information. More precisely, the risk forecasts for a future date are evaluated directly from the most recent return observations without any filter.

For a given probability level p, the ES for a risky X is given by

$$ES_p = E\left(X \mid X > VaR_p\right).$$

It is also a negative number and represents the loss as a percentage. That is, the expected percentage loss of X, when X exceeds the p^{th} quantile (VaR_p) - an extreme loss.

Equivalently, we can express it as

$$ES_p = VaR_p + E\left(X - VaR_p \mid X > VaR_p\right). \tag{2.27}$$

Here, the second term represents the mean of the excess distribution $F_{VaR_p}(x)$ above the threshold VaR_p. Applying the Pickands-Balkema-de Haan theorem as we did previously: if the threshold VaR_p, for $1-p > F(u)$, is sufficiently large, then $F_{VaR_p}(x)$ (the excess distribution above this threshold) is a GPD,

$$F_{VaR_p}(x) = G_{\gamma,\beta\ +\gamma\left(VaR_p-u\right)}(x).$$

Thus, the mean of the excess distribution $F_{VaR_p}(x)$ can be calculated as

$$\frac{\beta+\gamma(VaR_p-u)}{1-\gamma},$$

where $\gamma < 1$, and substituting into Equation (2.27) yields the expression

$$\hat{ES}_p = \frac{\hat{VaR}_p}{1-\hat{\gamma}} + \frac{\hat{\beta}-\hat{\gamma}u}{1-\hat{\gamma}},$$

and explicitly we can write

$$\hat{ES}_p = \frac{1}{1-\hat{\gamma}}\left[u+\frac{\hat{\beta}}{\hat{\gamma}}\left(\left(\frac{n}{N_u}p\right)^{-\hat{\gamma}}-1\right)\right]+\frac{\hat{\beta}-\hat{\gamma}u}{1-\hat{\gamma}}. \tag{2.28}$$

The ES formula in Equation (2.28) displays the significance of the shape parameter in tail estimation. As stated by Ren and Giles (2010), this parameter also determines how VaR and ES differ as measures of risk.

The literature of financial risk management bears a wide range of risk models and estimation techniques for VaR, as well as the comparison of these models to establish the most efficient one. However, as Righi and Ceretta (2015) state, ES has not been adequately studied to determine the best or the most effective risk estimator as compared to the others. This lack of consideration can be attributed to the fact that the use of ES in industry is not as common as that of VaR. They further state that one of the main factors is that backtesting ES models is more complicated than for VaR models.

Thus, Righi and Ceretta (2015) compared several distinct estimation methods and models for ES and investigated if there is clear advantage of certain models over others. They applied them to distinct asset classes and considered various estimation windows and significance levels. In order to assess the performance of these models, they performed some standard backtesting procedures and even proposed some amendments to the usual tests. In their study of ES and VaR models they highlight that a risk model with the best VaR estimation does not necessarily produce the best ES estimate - recall that

VaR ignores loss information beyond its value, while ES makes use of this information by definition. Nevertheless, these two are significantly linked since estimating VaR is necessary to predict ES.

The accuracy and efficiency of VaR and ES estimates can be assessed using various backtesting procedures, such as the Kupiec Likelihood Ratio test (Chinhamu et al., 2017). Abad et al. (2014) carried out a theoretical review of the existing literature on VaR methodologies, with a focus on the development of new approaches for it's estimation. They highlighted the relative strengths and weaknesses of each method and reviewed some backtesting procedures to evaluate the performances of these methods. Park and Kim (2016) proposed a new GPD parameter estimator, under the POT framework, to estimate tail risk measures VaR and CTE (Conditional Tail Expectation) for heavy-tailed losses.

McNeil and Frey (2000) proposed a method for estimating VaR and related risk measures to describe the tail of the conditional distribution of a heteroscedastic financial return series. Their approach combines fitting of GARCH models to estimate current volatility and EVT for estimating the tail of the innovation distribution of the GARCH model. Bee et al. (2016) applied realised volatility forecasting to EVT and compared their approach to the conditional EVT model of McNeil and Frey (2000). They assessed both approaches' ability to filter dependence in the extremes and produce stable VaR and ES estimates for one-day and ten-day time periods. They found that GARCH-type models perform well in filtering dependence, while the realised EVT approach seems preferable in forecasting. Dicks et al. (2014) proposed a novel approach that augmented GARCH models and EVT forecasts during periods where GARCH models do not converge. They investigated and applied various combinations to the JSE Financials (J580) and judged the performance of their models by the number of VaR and ES violations compared to the expected number.

Chinhamu et al. (2017) investigated the performance of the lambda distribution (GLD), the generalised Pareto distribution (GPD) and the generalised extreme value distribution (GEVD) in modelling daily platinum, gold and silver price log-returns and concluded that GLD is a suitable model for extreme risk in precious metal prices, followed by GPD. Chikobvu and Jakata (2020) modelled the probabilistic behaviour of unusually large financial losses and gains of the South African financial index, J580, and carried out a comparative analysis of the GEVD and GPD. Chinhamu et al. (2015b) fitted the GEVD and GPD models to the negative tail of the Johannesburg Stock Exchange mining index returns and compared the predictive ability of their VaR estimates.

Gilli and Këllezi (2006) applied EVT to 6 major stock market indices and computed their tail risk measures (ES, VaR and Return level) and the related confidence intervals. Dicks et al. (2014) applied EVT to GARCH (Generalised Autoregressive Conditional Heteroscedasticity) stochastic volatility models to model the extreme market returns of the JSE Financial Index (J580), and assessed the accuracy of these models using their VaR and ES estimates. Corlu and Corlu (2015) investigated the performance of the GLD when fitted to the skewed and heavy-tailed returns of exchange rate, and compared it to that of the skewed t distribution and the NIG (normal inverse Gaussian) distribution.

Ren and Giles (2010) applied EVT to the daily returns of crude oil prices in the Canadian spot market between 1998 and 2006 using the GPD. They estimated tail risk measures (VaR and ES) under various high quantile levels. They modelled both the losses and gains, and found that the GPD fit the positve returns slightly better than the negative returns. Wentzel and Mare (2007) utilised the POT method in analysing 10 years of daily FTSE/JSE TOP40 closing index data from 1996 to 2006. They sought to model the risk involved in investing in the South African equity market.

Andreeva et al. (2012) utilised the POT model of EVT to a series of losses of RTS index (RTSI) over a 15-year period (1995-2009) with a focus on tail-related risk as a modelling tool for modern risk management. Sigauke et al. (2013) used a GPD to model extreme daily price increases in the peak electricity demand in South Africa. Their study helps in quantifying the amount of electricity which can be shifted from the grid to off peak periods. Wong and Li (2010) applied the POT approach to model some insurance data and demonstrate the merit of having the full parametric model for the entire data set.

2.6.4 Backtesting Procedures

Backtesting procedures are used to assess the efficiency and effectiveness of VaR and ES estimates - i.e. how adequate and accurate our estimates are in forecasting risk. We can test risk (VarR and ES) models by performing statistical accuracy tests like, for instance, the Kupiec likelihood ratio test (Kupiec LR test) (Kupiec et al., 1995). This test utilises the fact that a good model ought to have its proportion of violations of VaR estimates close to the corresponding tail probability level (Chinhamu et al., 2015b). This is to achieve model robustness.

2.6.4.1 Kupiec Likelihood Ratio Test

This method entails evaluating the number of times, s_p, the observed returns fall below or above the VaR estimate at the level p (Chinhamu et al., 2017). That is, suppose returns at time t are modelled as x_t:

The test seeks to verify whether the VaR estimate is violated significantly more (or a fewer number of times) compared to p. We're interested in the number of times x_t falls below or above the VaR , i.e. $x_t > VaR_p$ for negative returns (losses).

We then compare the corresponding failure rates to tail probability p. The null hypothesis states that the expected proportion of violations is equivalent to p. That is, $\frac{\hat{s_p}}{N} = p$ with Kupiec LR statistic:

$$LR_{uc} = 2\ln\left(\left(\frac{s_p}{N}\right)^{s_p}\left(1-\frac{s_p}{N}\right)^{N-s_p}\right) - 2\ln\left(p^{s_p}(1-p)^{N-s_p}\right),$$

p can also be interpreted as the desired coverage rate. This test statistic is denoted LR_{uc} because the Kupiec LR test is an unconditional coverage test. If the proportion of violations is not significantly

different from level of significance $(100p)\%$, the test statistic takes on the value zero and the overall effectiveness of the model is confirmed as there is no evidence of inadequacy in the model. If it is significantly different, the test statistic indicates evidence that the VaR model either significantly underestimates or overestimates risk. It's asymptotic distribution is a chi-square with one degree of freedom (Abad et al., 2014). The p-value of this test is the probability that a $\chi^2(1)$ exceeds the LR_{uc}, i.e. $1-F(LR_{uc})$.

The null hypothesis is rejected if the test statistic is greater than the critical value of the $\chi^2(1)$ and the p-value is smaller than the level of significance $(100p)\%$. This implies the VaR model is not producing accurate estimates. A risk model that produces too many or too few violations is rejected, and in that way risk analysts and managers are able to identify why the model failed.

2.6.4.2 Christofferson Test

The Christoffersen (1998) conditional coverage test is an extension of the Kupiec LR test. It accounts for clustering of extremes and simultaneously evaluates if the proportion of violations is statistically equal to the expected one, and if the series of VaR violations is independently distributed. Christoffersen (1998) basically proposed a more elaborate test that includes an independence test which seeks to reject VaR models with clustered violations. The likelihood ratio statistic is expressed as

$LR_{cc} = LR_{uc} + LR_{ind}$,

where the LR_{ind} statistic corresponds to the LR statistic for the hypothesis of independence of violations. The Christofferson test statistic is given by

$$LR_{cc} = LR_{uc} + 2\ln\left(\frac{(1-\pi_0)^{N_{00}}(\pi_0)^{N_{01}}(1-\pi_1)^{N_{10}}(\pi_1)^{N_{11}}}{(1-\pi)^{N_{00}+N_{10}}(\pi)^{N_{01}+N_{11}}}\right),$$

where $i = 0,1;\ j = 0,1$ and N_{ij} denotes the number of returns in state j, upon being in state i in the previous period. That is, state $i = 0, j = 1$ means that the VaR estimated is violated, and the associated conditional probability of being in state $j = 1$ from previously being in state 0 is π_0. And state $j = 0$ means it is not violated. π_1 is the conditional probability associated with state $j = 1$ occuring if $i = 1$.

$\pi_0 = \frac{N_{01}}{N_{00}+N_{01}}$,

$\pi_1 = \frac{N_{11}}{N_{10}+N_{11}}$,

$\pi = \pi_0 + \pi_1$.

$LR_{ind} \sim \chi^2(1)$ and $LR_{cc} \sim \chi^2(2)$.

Under the null hypothesis, the violations are independent across the time periods, thus the conditional probabilities should be approximately equal ($\pi_0 = \pi_1$). This means the probability of a violation occuring after a period with no violation is the same as that of a violation occuring after a period with a

violation.

This test also bears the advantage of identifying which model failed based on rejecting the one that generates too many or too few clustered violations. The p-value of this test is the probability that a $\chi^2(2)$ exceeds the LR_{cc}, i.e. $1 - F(LR_{cc})$. However, the observed p-values of either of the LR_{uc} and LR_{cc} alone do not provide enough basis to rank the models (Abad et al., 2014).

2.6.4.3 Back-testing Criterion Statistic

Another similar test statistic or accuracy test for the significance of deviation of $\frac{\hat{s_p}}{N}$ from p is the back-testing criterion statistic (Abad et al., 2014), given as

$$Z = \frac{N\left(\frac{\hat{s_p}}{N}\right) - Np}{\sqrt{Np(1-p)}},$$

which is asymptotically distributed by a standard normal distribution.

2.6.4.4 McNeil and Frey Test

A bit more complicated, on the other hand, are the backtesting procedures for ES. The McNeil and Frey test, according to Righi and Ceretta (2015), is probably the most successful ES backtest in the literature. We are interested in the size of the variation or discrepancy between X_t and ES in the event that a (quantile) VaR violation occurs (McNeil and Frey, 2000). It is based on series r, which represents the residual exceedances over the VaR, i.e. the violations (X) standardised by the ES and the standard deviation of the returns X.

Given a level of significance p, r_i is given by

$$r_i = \begin{cases} \frac{X_t - ES_p^t}{\sigma_t}, & \text{for} \quad X_t > VaR_p^t, \\ 0, & \text{for} \quad X_t \leq VaR_p^t. \end{cases}$$

Here, backtesting is implemented on days t. Therefore, r_i represents the standard deviations of X from the ES estimate in the event of a violation and the residuals r_i are i.i.d.

H_0 : r has mean equal to zero ($\mu_0 = 0$).

H_A : r has mean less than zero ($\mu_0 < 0$).

Under the null hypothesis, r has zero mean and the residual exceedances are i.i.d. Under the alternative hypothesis, the risk measure conditional ES is underestimated. Notably, overestimation of this risk measure, in this case, is limited to the difference between the ES and VaR since when violation does not occur r takes on the value 0 (Righi and Ceretta, 2015). We can also note that underestimation is theoretically unbounded once the distribution of X is continuous in $\mathbb{R}$. The test statistic is given by:

$$T = \frac{\bar{r} - \mu_0}{\frac{\bar{\sigma}}{\sqrt{m}}},$$

where $\bar{r}$ and $\bar{\sigma}$ are the mean and standard deviation of the exceedance residuals $\{r_1, r_2, \ldots, r_m\}$. This is a one sided t-test where the null hypothesis is not rejected for p-values that exceed the chosen level of confidence p. Which means the validity of the ES model is approvable at $(100)p\%$.

To test the hypothesis, a bootstrap simulation technique that makes no assumptions about the underlying distribution of the residuals can also be used. This is owing to the fact that residuals derived under the normality assumption always fail the bootstrap test, producing p-values much lower than 0.01 with biased results (McNeil and Frey, 2000) - thus it alleviates any bias assumptions. The technique generates a sample without replacement from the shifted residuals $\tilde{r}_i = r_i - \bar{r} + \mu_0$ - i.e. sample = $\{\tilde{r}_1^*, \tilde{r}_2^*, \ldots, \tilde{r}_m^*\}$, to obtain the test statistic:

$$T_j^* = \frac{\bar{\tilde{r}}^* - \mu_0}{\frac{\bar{\tilde{\sigma}}}{\sqrt{m}}},$$

for each bootstrap sample.

In Righi and Ceretta (2015)'s comparison of ES models, on the other hand, they proposed an adaptation to this McNeil and Frey (2000) test. Additionally, they carried out a Monte Carlo experiment and their proposed test proved to be more successful than the usual tests.

2.7 Cryptocurrencies

2.7.1 Overview

Cryptocurrencies are a subset of digital currencies or assets designed to function as mediums of exchange that are created and managed through the use of advanced encryption techniques (known as cryptography) (Osterrieder et al., 2016). These techniques, along with blockchain technology, provide secure anonymous transactions for their users as well as the ability to create additional units of currency (Chan et al., 2022). They utilise data encryption (a standard method that secures data by converting it from a readable format into an encoded form) which authenticates users in order to protect all information (including personal information and payments) and communication to ensure that only the sender and intended recipient are able to read and process the data.

Essentially, blockchains are defined to be secure decentralised public ledgers that facilitate the recording of all the transactions and electronic exchange of assets performed by users in a business network. They are distributed amongst the network and linked together to form an unbroken chain of blocks that store and share data, and maintain its transparency as the ledgers are unalterable - users would

require agreement from their network to modify, or even delete, these blockchains. This preserves transparency and prevents the entry of unauthorised transactions.

Furthermore, as opposed to physical money or banking systems, cryptocurrencies are decentralised as they are not issued or controlled by the state or any other financial institution (Osterrieder et al., 2016). The most prominent way of purchasing or selling them is through platforms called crypto exchanges (widely known as crypto apps) such as Kraken, Coinbase and Binance, to name a few. Chan et al. (2022) state that they were initially developed mainly for speed and economic efficiency when transferring money globally. They have since gradually developed and revolutionarily grown into more versatile assets that are widely available.

As of December 2022, CoinMarketCap (2022) reports that the cryptomarket is comprised of over 21, 000 different cryptocurrencies. Bitcoin and Ethereum are the largest cryptocurrencies in terms of market capitalisation and trading volume (with Bitcoin representing 42.3% of the total market and Ethereum 19.2%, according to market information obtained from CoinMarketCap (2023) (January 2023). Their main objective is to provide, among other things, faster and more efficient monetary payments and decentralised finance (Chan et al., 2022). Bitcoin utilises advanced cryptography to create a platform where its users use a pseudonymous address to identify one another, and a pass code or some private key to update the digitally distributed public ledger when transferring between their accounts (Islam and Das, 2021).

Table 2.1: Market capitalisations of top 10 cryptocurrencies (CoinMarketCap, January 2023).

Cryptocurrency	Market capitalisation (USD billions)	Market capitalisation (relative %)
Bitcoin	438.38	42.31
Ethereum	197.7	19.2
Tether	66.76	6.41
BNB	47.98	4.56
USD Coin	43.64	4.18
XRP	21.26	1.98
Binance USD	16.03	1.56
Cardano	12.97	1.23
Dogecoin	11.67	1.09
Solana	8.96	0.88

2.7.2 Bitcoin

The cryptocurrency market has witnessed tremendous growth on a global scale over the past decade, with Bitcoin being the largest and most popular cryptocurrency. Bitcoin was the first launched cryptocurrency (in 2009) and despite being a decentralised currency, free from government intervention or

manipulation, its market capitalisation and trading volume continue to outperform all other cryptocurrencies (Osterrieder et al., 2016). This also explains why a lot of the crypto academic research has been conducted mainly on Bitcoin.

The cryptocurrency industry is booming as the phenomenon of cryptocurrencies is spreading worldwide and constantly drawing the attention of investors, the media, as well as financial institutions. Unfortunately, this complete decentralisation bears the downside of no central authority to ensure that things are operating smoothly in the market, or to intervene and control inflation, or even back the value of these digital assets (Chan et al., 2022). Even though the foreign exchange market exhibits unstable pricing, the volume of trade and the frequency with which the trading occurs remains high, which explains why many perceive it as the most liquid market worldwide.

The Bitcoin market and its financial characteristics have been studied considerably over the past few years. Bitcoin remains the most influential cryptocurrency despite inheriting financial traits far different from traditional financial assets, or facing criticism for being perceived as highly speculative (Hussain et al., 2021). Its price changes, and thus its returns, exhibit extreme behaviour due to its highly volatile and risky nature. Therefore, examining this extreme behaviour in the tail of its return distribution as well as its negative financial implications is an imperative study.

2.8 Extreme Value Theory Applied in Cryptocurrencies

Investing in the cryptocurrency industry does not only have the benefit of excess returns but also huge risks, due to the unique traits of these emerging financial assets (Liu et al., 2020). Cryptocurrencies are extremely volatile with leptokurtic behaviour, and their financial returns much more riskier than traditional fiat/paper currencies. Their extreme price changes contradict the normality assumption; which complicates modelling the returns for prediction and risk management purposes (Gkillas and Katsiampa, 2018).

Analysing the tail behaviour of returns is highly important for risk management, benefiting both investors and policy makers. Modelling the losses and gains present in the fat tails assists investors to fully understand the underlying risks of their investment choices, so as not to make uninformed trade decisions in this unstable and highly volatile market environment (Gkillas and Katsiampa, 2018), as well as predict 'boom' and 'bust' trends. Hussain et al. (2021) further justifies the usefulness of heavy tails in determining the likelihood of a market crash, which could ultimately aid in providing insightful information for forecasting the state of the economy.

The relationship between trading volume and return levels in financial markets has been a subject of interest for quite some decades now. This dynamic relationship conveys useful information on understanding market structures, as well as future price movements, especially in times of market stress such as market turbulence or irrational market exuberance. Market participants are able to gain

insight into assessing prediction of price returns, market efficiency and how to incorporate this valuable information into their trading strategies (Chan et al., 2022). Katsiampa et al. (2018) 's views were also in line with this observation and suggests that traders who utililise such information perform better than those who do not.

Chan et al. (2022), in their extreme value analysis of the tail relationships between returns and volume, emphasise that the stylised facts from empirical studies on the volume-return relation suggest a positive correlation between these two variables when studying interest rates, market indices, commodities and equities. They further note that speculative trading poses an asymmetric effect on different types of assets, resulting in a variation in their return-volume dynamics. Assessing and understanding the extreme returns in Bitcoin and their properties, as well as the dynamic relationship between returns and volume, using EVT is the main motivation of this research. It's important to note that extreme returns may also be positive - extreme price escalations can provide users with exceptional profits, even up to 100%, in a very short time period.

Applying an EVT approach to the analysis of the return-volume dynamic at the extreme tails is a crucial and worthwhile study in financial markets. Particularly in light of their inevitable features of unexpected and large price movements, an extreme value analysis can assist traders and investors in understanding the market during the occurences of extreme events. The cryptocurrency market does inherit qualities similar to financial markets, thus studying this volume-return relationship using EVT is very significant.

When investigating the return-volume relation, examining the extreme dependence and correlation between the two variables is critical as these statistical tools serve as technical indicators for market participants who rely on such when making trade decisions in times of extreme market conditions. Traders are always interested in how they can make more profits or minimise losses, thus knowing how trading volume can affect their returns especially at times when they are uncertain about the direction of the market - due to extreme events that initiate a change in the liquidity conditions, market trends, volatility and the value of their assets - can be of great use. As Balcilar et al. (2017) state, cryptocurrencies are characterised by sharp price fluctuations which are associated with increased transaction volume. This means that price movements and transaction volume levels exhibit a strong dependence.

The risk of losing a lot, if not all, is very high in this market due to the nature of the investment in cryptocurrencies being very speculative since market values are always changing. Thus, the extreme value behaviour of cryptocurrencies is an interesting study. Rare events or events with a very low probability of occuring, such as severe losses, market crashes or even market illiquidity, contribute to risk exposure in this market. Notably, EVT has been applied to a range of cryptocurrencies in the recent years but there is still not enough research that involves EVT in crypto analysis.

Despite the numerous studies of the return-volume relation as well as their dependence and correlation

structures in financial markets, research of this kind in the cryptocurrency market is rather scarce or limited (Chan et al., 2022). Longin and Pagliardi (2016) applied EVT to study the tail relationship between trading volume and returns during the extraordinary market conditions (booms and crashes) of the US stock market. They found that the extreme correlation between return and volume decreases as we move towards the tails of the return distribution where extreme events exist. This implies that trading volume and returns are not strongly correlated during booms and crashes in the stock market. Lin (2013) and Gunduz and Hatemi-J (2005) investigated the dynamic stock return-volume relation in emerging markets. Even earlier academic studies include Jansen and De Vries (1991) 's extreme values analysis of the tail behaviour of stock returns, Balduzzi et al. (1996) 's study of returns and the price-volume relation, and Saatcioglu and Starks (1998) and Chen et al. (2001) 's analysis of the stock price-volume relation. Marsh and Wagner (2004) investigated this relation for several large international equity markets in the course of market stress and found that trading volume is crucial in comprehending large price fluctuations, expected equity returns, as well as their variability. Do et al. (2014) examined how trading volume affects financial returns in different regional locations including Asia, Latin America and Western Europe. An important finding by Balduzzi et al. (1996) showed that trade misinterpretation in financial markets can lead to unfavourable market conditions such as market illiquidity and even extreme price shifts.

Although the crypto industry is still in its developing stages, it has already attracted quite a lot of academic attention. However, there still remains a research gap on the volume-return relationship in cryptocurrencies. Zhang et al. (2018) carried out an analysis of the volume-return relation of the Bitcoin market predicated on multifractal detrended cross-correlation analysis. Their results revealed that Bitcoin return-volume exhibits a non-linear dependent relationship. While more recently, Naeem et al. (2020) used copulas to look at the dependence between extreme returns and trading volume among a variety of cryptocurrencies and discovered that the lower tail exhibits a stronger dependence than the upper tail. Balcilar et al. (2017) explored the relationship between Bitcoin transaction volume and returns and volatility. They implemented a non-parametric causality-in-quantiles test, and found that volume can predict returns, except during periods of bull and bear in the Bitcoin market. This result stresses the importance of considering tail behaviour when assessing the dependence between volume and returns. Chan et al. (2022) 's contribution to the literature has been the most profound, applying an EVT approach (POT method) to investigate the tail relation between the transaction volume and returns of Bitcoin and Ethereum at high-frequency, with a focus on the extreme dependence and correlation between the two variables. Their results show that there exists a weak positive correlation between returns and volume when considering extreme events (booms and busts) in the left and right tails of the distribution.

Osterrieder and Lorenz (2017) studied the extreme behaviour of Bitcoin returns and compared it to traditional G10 exchange rate currencies (versus USD) and found that Bitcoin exhibits higher volatility and heavier tails. Gkillas and Katsiampa (2018) studied the tail behaviour of 5 major cryptocurren-

cies (Bitcoin, Ethereum, Ripple, BitcoinCash amd Litecoin) by employing an extreme value analysis. They fitted a GPD model to the returns of each cryptocurrency by extracting the extremes using POT approach. They estimated tail-risk measures Value-at-Risk and Expected Shortfall, and their VaR estimates (at different percentiles) for both negative and positive returns showed that Bitcoin Cash has the highest potential gain and loss since it exhibited the highest value of VaR and ES - making it the riskiest cryptocurrency compared to the others, while Bitcoin and Litecoin were found to be the least risky. It is very important for investors to be cognisant of the riskiest asset. On the other hand, Zhang et al. (2019) investigated the tail risk behaviour of high frequency (hourly) log returns of four most popular cryptocurrencies (Bitcoin, Ethereum, Litecoin, and Ripple verus the US Dollar). They modelled the returns using the GPD and estimated tail risk measures at different thresholds and found that Bitcoin was the least risky cryptocurrency (it exhibited the largest potential gain or loss for both positive and negative log returns at every percentile and threshold) with Ripple being the most risky.

Hussain et al. (2021) used EVT for financial risk modelling for Bitcoin and employed the GPD to model daily extreme returns in the Bitcoin market from 2017 to 2019 and found that Bitcoin had the highest return level over all periods. The tail of the market retains useful information to understand the extreme volatility movements in Bitcoin prices, thus justifying their application of EVT modelling in this very volatile period in the crypto industry. In an attempt to discover and better understand the severe characterisitcs of the extreme negative returns in Bitcoin, they used the POT method in their analysis and also determined return-level probabilities for these extremes based on the fitted GPD model. These probabilities can be used to deduce the probability of unpredictable Bitcoin returns in extreme events or cases. The equation of daily returns being: $R_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$, where R_t is the return on the Bitcoin for period t. P_t denotes the Bitcoin at the end of period t, with P_{t-1} being the Bitcoin at the end period $t-1$. They utilised daily data in the US Dollar and estimated potential maximum losses from the most extreme return levels for various return periods. Their results displayed sufficient agreement with the theoretical EVT models and depicted that informative knowledge that's realistic in the financial sectors can be obtained from the models.

Osterrieder et al. (2016) executed a detailed extreme value analysis of the most important cryptocurrencies, their correlations, volatilities and tail dependencies as well as their statistical properties. They analysed the returns of data from six of the ten largest cryptocurrencies: Bitcoin, Ripple, Monero, MaidSafeCoin, Dash and Litecoin. The data consisted of historical global price indices from cryptocurrencies - aggregated cryptocurrency prices from multiple exchanges which provide a weighted average cryptocurrency price. They fitted both the GPD and GEVD models, with a particular focus on the left tail of the distribution, and computed risk measures VaR and ES. They used copula theory to investigate the dependence between the various currencies and compute tail dependence coefficients, as well as the extremal index to measure clustering of extreme returns. They found that cryptocurrencies exhibit strong non-normal characteristics, large tail dependencies and heavy tails, and that Bitcoin was one of the less risky cryptocurrencies.

Katsiampa et al. (2018) examined the extreme dependence between returns and trading volume in both the left and right tails of eight major cryptocurrencies in extremely volatile periods using the POT approach. They used bivariate EVT to examine the joint distribution of extreme price returns and trading volumes. They analysed data that comprised of both prices and trading volumes of cryptocurrencies which had been existing for more than a year at that time, with a market capitalisation of more than 5 billion dollars. Katsiampa (2017) investigated the ability of several competing GARCH models to explain the Bitcoin price volatility. Cheah and Fry (2015) performed econometric modelling of Bitcoin prices to show that Bitcoin markets are substantially vulnerable to speculative bubbles. Liu et al. (2020) investigated whether cryptocurrencies investors' decisions can rely on the pragmatic and parsimonious approaches for VaR forecasting under RiskMetrics type models.

Islam and Das (2021) investigated the extreme nature of cryptocurrency returns in their prediction of Bitcoin returns using EVT approaches: the Block maxima and the Peaks-over-Threshold. They first ran several statistical tests to confirm the extreme nature of Bitcoin returns before fitting the models. They then used the fitted models to study and predict long term Bitcoin return levels for various return periods with a 95% confidence interval, by estimating extreme quantiles.

They used daily adjusted closing prices of Bitcoin in USD from Yahoo Finance (2019). The standard daily rate of return was calculated as

$$\text{Daily rate of return} = \frac{\text{Adjusted closing prices - Opening price}}{\text{Opening price}} .$$

They concluded that the GEVD under the BM method fit the Bitcoin returns better than the GPD under the POT method.

2.8.1 High-frequency Data

The study of High-frequency trading (HFT) is becoming more and more significant, as Chan et al. (2022) note, with the increasing popularity and the unprecedented growth of the crypto industry, more online exchange platforms are now providing their users with tools to trade at a very high speed over the course of minutes or even seconds. Examining the volume-return dynamic in a high-frequency setting, especially during extreme events in the market, is beneficial to traders and investors who have a strong reliance on models that are based on this volume-return relation. Chan et al. (2022) are the first and so far the only authors to analyse high-frequency extreme dependence of cryptocurrency volume and returns at the extreme tails.

The tail relationship between trading volume and financial returns for Bitcoin at high frequency will be analysed, in particular the extreme dependence and extreme correlation between these two variables.

Understanding the extreme tail dependency in Bitcoin is useful for recognising and evaluating the risks associated with returns when making trading decisions, so as to prevent and minimise the possible negative implications of extreme events.

High frequency cryptocurrency data, such as intraday or hourly prices or returns, will be analysed using EVT models in order to investigate and understand (not only the dependence and correlation structure) the severe or extreme returns in Bitcoin. According to Osterrieder et al. (2016), the Bitcoin market displays the highest liquidity compared to all the other cryptocurrencies. This contributes to the growing popularity of this cryptocurrency, and the shift of the cryptocurrency industry as a whole towards HFT. This has stimulated the focus on Bitcoin for this research.

2.9 Extreme Value Theory Applied in Gold

All over the world gold has dominated as one of the best assets to invest in due to its stability in value and rate of return, its influence on other precious commodities, as well as the impact it has on an economy at large. This valuable metal is widely used in a range of industries and can be converted to paper money. It is very liquid and remains a good investment in both the long run and short run, which explains why it continuously captures the attention of many investors globally. In the past, according to Som and Kayal (2022), gold was considered a safe haven by investors who wanted to diversify their portfolios.

Nonetheless, it is also perceived as a speculative investment by some - since it does not generate any income on its own and positive returns are induced by a rise in price. The price of gold changes considerably over time - and that alone prompts a level of risk associated with gold. A major concern in finance are the returns from gold prices because, just like crypto traders, investors in gold are always interested in predicting the magnitude of the profits and losses they can expect.

With the rising development and growing popularity of digital assets, as well as the evolution of technology, an interesting debate has emerged in the world of finance and investment: are cryptocurrencies the new 'digital gold'? Modelling and comparing the risk in the tails of these two assets (cryptocurrencies and gold) has been an interesting study for researchers and financial risk managers in the past few years. Risk measures such as Value-at-Risk and Expected Shortfall can be estimated in order to determine the asset with the lowest risk - which makes it more appealing to potential investors.

Chinhamu et al. (2015a) used GPD models, in comparsison with the normal distribution and Student's t, to model gold returns and estimated risk measures VaR and ES. Their results concluded that the GPD is an appropriate model to describe the conditional excess distributions of a heteroscedastic gold log return series and provides adequate estimations for VaR and ES. Pratiwi et al. (2019) used the GEVD's

Block Maxima approach to model gold returns and estimated gold investment risk using the popular VaR measure. Ali et al. (2020) used GEVD models to model the negative log returns of Malaysian gold prices and applied different parameter estimation methods to estimate the VaR risk measure.

Khan et al. (2021) analysed the risk of gold investment by applying the EVT to historical daily data for extreme daily losses and gains in the price of gold in the Pakistan Bullion Market. They applied both the BM and POT methods and concluded that their point and interval estimates of VaR and ES for different high quantile levels show strong stability through a range of the selected thresholds - suggesting an adequate accuracy and reliability of the estimated quantile-based risk measures. Som and Kayal (2022) investigated the effect of the inclusion of cryptocurrencies in a portfolio with and without gold, to see if Bitcoin is indeed the 'digital gold'. They evaluated risk measures VaR and ES in their portfolio optimisation technique and their results were in favour of the inclusion of Bitcoin in the portfolio.

2.10 Bivariate Extreme Value Theory and Copula Theory

Simply modelling the returns and trading volume variables using univariate tools does not capture adequately the extreme conditions of the crypto market, especially for the purpose of making predictions and more informed decisions. In order to study the volume-return relationship in the extreme cases, the joint distribution of extreme returns and trading volume is an optimal statistical starting point. Bivariate EVT assists in modelling this relationship using copula theory to evaluate and carry out inferences on the dependence structure between the extreme values of returns and trading volume in the tails of their joint distribution, as well as to compute tail dependence coefficients and correlation coefficients.

Bivariate extreme value models focus on the joint distribution of the extreme values in two variables of interest. The models can be used to evaluate the likelihood of extreme events occuring for both variables simultaneously. Coles et al. (2001) defined a multivariate extreme distribution as follows:

Definition 2.1. Multivariate domain of attraction

If there exists a sequence $\mathbf{a_n} > 0$ and $\mathbf{b_n}$ such that $\mathbf{a_n^{-1}}(\mathbf{M_n} - \mathbf{b_n})$ converges in distribution to a non-degenerate p-variate distribution function G with non-degenerate margins such that:

$$P\left(\mathbf{X} \leq \mathbf{a_n^{-1}}(\mathbf{M_n} - \mathbf{b_n})\right) = F^n(\mathbf{a_n x} + \mathbf{b_n}) \to G(\mathbf{x}),$$

and $n \to \infty$, then F is in the domain of attraction of a multivariate extreme value distribution G (i.e. $F \in D(G)$). For the bivariate case, the multivariate extreme distribution with $p = 2$ is being considered.

The financial literature shows that the comovement between financial returns has often been described or modelled using the method of linear correlation (Shamiri et al., 2011). However, this method is suitable for measuring dependence if the returns follow a multivariate normal distribution. That said, as was previously established, financial returns exhibit non-normal features such as skewness and heavy tails, thus the normal distribution models them inaccurately, even though easy to implement. Thus, linear correlation is not appropriate for measuring dependency for heavy-tailed distributions.

Now, this initiates a search for a dependence structure that is able to accomodate EVT's extreme or special features of asymmetry, fat tails and tail dependence. This structure will allow for non-linear dependency and will not be restricted to a single numeric measure of dependency (Longin and Solnik, 2001). This initiated the use of the concept of copulas in finance, as copulas provide a flexible methodology that allows the modelling of multivariate dependence in a non-Gaussian structure.

In addition, assessing the financial risk associated with extreme events requires an estimation of the tail dependence between the extreme or less probable values of variables, i.e. the probability of extreme events (and thus values) for the two random variables occurring simultaneously. Ultimately, an analysis of the dependence in the tails of the joint distribution of returns should then be carried out - substantiating the use of copulas for dependence modelling in an extreme value analysis.

2.10.1 Copulas

Copulas are multivariate cumulative distribution functions that are commonly used to describe the dependencies and correlations among variables, independent of their marginal distributions (Panjer, 2006). The most significant theorem for the theory of copulas was the formal definition summarised by Sklar (1959):

Theorem 2.2. *Sklar's Theorem*

An n-dimensional joint distribution can be decomposed into its n univariate marginal distributions and an n-dimensional copula. For an n-dimensional random vector $X = (X_1, \dots, X_n)^T$ *with cumulative distribution* F*, let* F_i *denote the marginal distributions of* X_i*. Then there exists an associated copula* C *such that for all* $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ *:*

$$F(x) = F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)).$$

Therefore a copula C *maps the marginal distributions* $F_1, \dots, F_n$ *to the joint distribution* F *and the following can be deduced*

$$X \sim F = C(F_1, \dots, F_n).$$

Note: for a continuous distribution F*, the copula is unique. It is any joint distribution function with*

uniform marginl distributions on the interval $[0,1]$.

i.e. $F_i(X_i) = U_i$ *, where* $U_i \sim U(0,1)$*,*

which results in the copula

$U = (U_1, \ldots, U_n)^T \sim C$.

This copula is defined as the joint distribution of the probability integral transformations of the original variables (U_i) (Osterrieder et al., 2016; Panjer, 2006) .

Let $F_1$and F_2 be the marginal distribution functions of (X_1, X_2), then their copula C is expressed as

$$C(u_1,u_2) = P(U_1 \leq u_1, U_2 \leq u_2),$$

$$C(u_1,u_2) = C(F_{X_1}(X_1), F_{X_2}(X_2)),$$

where $X_1, X_2 \in \mathbb{R}$ and $u_1, u_2 \in (0,1)$.

The density function of the copula can be defined as

$$c(u_1,u_2) = \frac{\partial^2 C}{\partial u_1 \partial u_2} = \frac{\partial^2 C}{\partial u_2 \partial u_1},$$

assuming both partial derivatives exist and are continuous.

Implementing the chain rule yields the joint density

$$\begin{aligned} f_{1,2}(x_1,x_2) &= \frac{\partial^2 F_{1,2}(x_1,x_2)}{\partial x_1 \partial x_2} \\ &= \frac{\partial^2 C(u_1,u_2)}{\partial u_1 \partial u_2} \frac{\partial F_1(x_1)}{\partial x_1} \frac{\partial F_2(x_2)}{\partial x_2} \\ &= c(u_1,u_2) f(x_1) f(x_2), \end{aligned} \tag{2.29}$$

where $f(x_1)$ and $f(x_2)$ represent the marginal densities of $X_1$and X_2, respectively.

Moreover, from Equation (2.29) it can be deduced that it is always possible to express a bivariate density by specifying the marginal densities and a copula density (Shamiri et al., 2011). Meaning, the copula has complete information about the dependence structure of the two variables. This is the predominant characterisitic of copulas - their ability to decompose a joint distribution into the following: marginal distributions and a dependence structure. Additionally, as Shamiri et al. (2011) further state, different dependence structures are able to merge the same marginal distributions into different joint distributions. And similarly, different marginal distributions under the same dependence structure can also result in different joint distributions.

In statistical terms, Costinot et al. (2000) state that copulas can in fact be interpreted as the dependence function of the random variables in question. There exists many copula functions that fall under different classes. A particular class of copulae is called the Archimedean copulae. The copulas under the Archimedean family are described as simple since they are described by one parameter only, but they do also share the convenient and fundamental properties of symmetry and associativity. They enable a considerable number of dependence structures and most of them have closed form expression - which is convenient for the parameter estimation procedure. Three prime examples of 2-dimensional copulas under this class are:

Example 2.3. The Clayton copula, which models lower tail dependence only, is defined as

$$C(u_1,u_2,\alpha) = \left(u_1^{-\alpha}+u_2^{-\alpha}-1\right)^{-\frac{1}{\alpha}},\ u_1,u_2 \in (0,1)\ ,$$

for $\alpha > 0$.

Example 2.4. The Gumbel copula, which models upper tail dependence only,

$$C(u_1,u_2,\alpha) = \exp\left[-\left\{(-\log u_1)^{\alpha}+(-\log u_2)^{\alpha}\right\}^{\frac{1}{\alpha}}\right],\ u_1,u_2 \in (0,1)\ ,$$

for $\alpha \in [1,\infty)$.

Example 2.5. The Frank copula,

$$C(u_1,u_2,\alpha) = -\frac{1}{\alpha}\log\left[1+\frac{(\exp(-\alpha u_1)-1)(\exp(-\alpha u_2)-1)}{(\exp(-\alpha)-1)}\right],$$

for $\alpha > 0$. This copula does not model tail dependence.

Consider the variable X, the cryptocurrency returns, and $F_X(x)$, the cdf of X. All observations of X greater than the threshold u_x denote positive exceedances, and recall that by symmetry, negative return exceedances can be defined as well. As was discussed in the earlier sections of this chapter, extreme values exceed a sufficiently high threshold and occur with a small probability p (tail probability). Thus, an extreme return X is greater than u_x with probability p and lower than u_x with probability $1-p$. The following relation is then deduced:

$$p = P(X \geq u_x),$$

and

$$\begin{aligned} 1-p &= P(X \leq u_x) = F_X(u_x) \\ p &= 1-F_X(u_x). \end{aligned}$$

Now, using the same reasoning, Ledford and Tawn (1996) derived a bivariate model of the tails of two variables. In this case we are considering positive returns since $X > u$ pertains to the right tail of the distribution.

In order to quantify the dependence structure between the extreme values of returns (X) and trading volume volume (Y) (with appropriately large thresholds $u_x$and u_y, respectively), a 2-dimensional Gumbel copula function (also called an extreme value copula) is used to model their joint cumulative distribution (Chan et al., 2022).

Considering the return exceedances from X and the volume exceedances from Y in the Bitcoin market, the tail of the distribution of each variable is modelled as

$$\begin{aligned} F_X(x) &= (1-P(X>u_x))+P(X>u_x)\cdot G_{\gamma,u,\beta}(x) \\ &= 1-P(X>u_x)\cdot\left(1+\gamma_1\frac{(x-u_x)}{\beta_1}\right)^{\frac{-1}{\gamma_1}}, \end{aligned} \tag{2.30}$$

and

$$\begin{aligned} F_Y(y) &= (1-P(Y>u_y))+P(Y>u_y)\cdot G_{\gamma,u,\beta}(y) \\ &= 1-P(Y>u_y)\cdot\left(1+\gamma_2\frac{(y-u_y)}{\beta_2}\right)^{\frac{-1}{\gamma_2}}, \end{aligned} \tag{2.31}$$

respectively. Equation (2.30) basically indicates that an observed positive return X does not belong to the tail with probability $1-P(X>u_x)$, or it is drawn from a univariate GPD (with a threshold u_x) with probability $P(X>u_x)$ (Longin and Solnik, 2001). Furthermore, a return observation that does not exceed the threshold u_x is irrelevant in modelling the distribution of the tail, thus its actual value is not included. Similarly, the same explanation applies to positive volume exceedances in Equation (2.31).

A Gumbel copula is a special type of Archimedean copula that is extensively used in financial modelling. This copula is based on the logistic model and is used to model (positive) upper tail dependence - at times it is referred to as the logistic dependence function (Costinot et al., 2000).

The logistic model is a common bivariate parametric model that provides a bivariate extreme value distribution (Gumbel, 1958) using a logistic dependence function D_G. Recall, for the univariate EVT case, the aymptotic behaviour of returns or volumes are considered and the limiting nondegenerate distribution then defined as a GPD. The dependence function is defined as

$$D(r_1,r_2)=\left(r_1^{-1/\alpha}+r_2^{-1/\alpha}\right)^{\alpha},$$

where $r_i=-1/\log G_{\gamma,u,\beta}(x)$.

By definition (Longin and Solnik, 2001), the dependence function maps the univariate marginal distributions of return variables X and Y that each follow a GPD, to obtain the limiting nondegenerate bivariate extreme value distribution which satisfies the limit condition

$$G_{X,Y}(x,y) = \exp\left\{-D_{G_{X,Y}}\left(-1/\log G_{\gamma,u,\beta}(x), -1/\log G_{\gamma,u,\beta}(y)\right)\right\}$$

and

$$\begin{aligned} D_{G_{X,Y}}(r_1,r_2) &= \left[\left(-1/\log G_{\gamma,u,\beta}(x)\right)^{-1/\alpha} + \left(-1/\log G_{\gamma,u,\beta}(y)\right)^{-1/\alpha}\right]^{\alpha} \\ &= \left[\left(-\log F_X(x)\right)^{1/\alpha} + \left(-\log F_Y(y)\right)^{1/\alpha}\right]^{\alpha}. \end{aligned}$$

The POT method is applied to analyse extreme returns and volume, and the Gumbel copula function to model the dependence between bivariate exceedances is defined as follows:

$$\begin{aligned} F_{X,Y}(x,y) &= C_\alpha\left(F_X(x),\ F_Y(y)\right),\ x,y \in \mathbb{R} \\ &= \exp\left\{-\left[\{-\log(F_X(x))\}^{\alpha} + \{-\log(F_Y(y))\}^{\alpha}\right]^{\frac{1}{\alpha}}\right\},\ F_X(x),\ F_Y(y) \in \mathbb{R} \\ &= \exp\left\{\left[\left\{-\log\left(1 - P(X > u_x)\left(1 + \gamma_1\frac{(x-u_x)}{\beta_1}\right)^{\frac{-1}{\gamma_1}}\right)\right\}^{\frac{1}{\alpha}}\right.\right. \\ &\left.\left. + \left\{-\log\left(1 - P(Y > u_y)\left(1 + \gamma_2\frac{(y-u_y)}{\beta_2}\right)^{-\frac{1}{\gamma_2}}\right)\right\}^{\frac{1}{\alpha}}\right]^{\alpha}\right\}, \end{aligned} \tag{2.32}$$

for $x > u_x$ and $y > u_y$, where u_x and u_y are thresholds associated with returns and volumes, respectively. $\alpha \in [1,\infty)$ represents the dependence parameter which controls the level of dependence between extremes, (γ_1,β_1) and (γ_2,β_2) are parameters describing the marginal distributions of X and Y, respectively. $\alpha = 1$ corresponds to independence and $\alpha = \infty$ corresponds to total dependence - the strength of dependence increases as α increases. This Gumbel copula models the bivariate distribution of returns and trading volumes exceedances. The copula, along with the other parameter estimates (u,γ,β) for both returns and volume and (α,ρ) can be fitted to data using the maximum likelihood estimation technique in R (R Development Core Team, 2020).

2.10.2 Maximum Likelihood Estimation with Censored Data

The reliable MLE method can also be used to estimate the parameters for the Gumbel copula. The copula is fitted to data using the censored-likelihood method which is based on a censoring assumption. Here, the likelihood function is constructed in such a way that a return observation X_i that falls below threshold u_x is considered as censored at the threshold. The same assumption applies for trading volume observations Y_i.

To begin with, this method relies on the following assumptions: returns are independent and so are volumes. Secondly, the thresholds u_x and u_y selected to determine exceedances (or equivalently the tail probabilities $P(X > u_x)$ and $P(Y > u_y)$) are independent of the returns (or volumes) and time (Longin and Solnik, 2001).

If we let $(x_1, \dots, x_n), \dots, (y_1, \dots, y_n)$ be independent observations of a random vector (X,Y) with some arbitrary cdf $F(x,y)$, for each variable X and Y, the observations exceeding thresholds u_x and u_y , respectively, follow a GPD. That is, $F(x,y)$ corresponds to the region where $x > u_x$ and $y > u_y$. However, the problem is that $F(x,y)$ cannot be used for the other regions outside $\{x > u_x, y > u_y\}$.

So, making inference on this bivariate model for exceedances is not simple because it could happen that for a given observation point (x_i, y_i), only one value exceeds the corresponding threshold. Thus, taking into consideration both thresholds $u_x$and u_y, the space of returns and volumes is divided into four regions by:

$\{A_{jk} \; ; \; j = I(X > u_x) \; , \; k = I(Y > u_y)\}$,

where I denotes the indicator function. The regions are

$$\begin{aligned} A_{00} &= (-\infty, u_x) \times (-\infty, u_y) \\ A_{01} &= (-\infty, u_x) \times [u_y, \infty) \\ A_{10} &= [u_x, \infty) \times (-\infty, u_y) \\ A_{11} &= [u_x, \infty) \times [u_y, \infty) . \end{aligned}$$

All observations that do not exceed a threshold are considered censored data.

For example, a likelihood for the points (x_i, y_i) in region A_{11} would be simply constructed by using the density of $F(x,y)$. However, for a point in region A_{10}, the likelihood would be

$$P(X = x_i, Y \leq u_y) = \frac{\partial}{\partial x} F(x,y) ,$$

which means that $F(x,y)$ only contains information corresponding to x, since $x_i > u_x$ and $y_i < u_y$. This is why the method of censored data is necessary.

Following Longin and Solnik (2001), the likelihood is constructed according to the observations of

returns and volumes at time t - i.e. observations $(X, Y = X_t, Y_t)$ that fall in region A_{jk}. It is denoted as $L_{jk}(X_t, Y_t)$:

Let p_i denote the tail probabilities, where $p_1 = P(X > u_x)$ and $p_2 = P(Y > u_y)$.

$$\begin{aligned} L_{00}(X_t, Y_t) &= F(X_t, Y_t) \\ &= \exp\{-D(W_1, W_2)\}, \end{aligned}$$

$$\begin{aligned} L_{01}(X_t, Y_t) &= \frac{\partial}{\partial Y_t} F(X_t, Y_t) \\ &= \exp\{-D(W_1, Z_2)\} \cdot \frac{\partial}{\partial Y_t} D(W_1, Z_2) \cdot K_2, \end{aligned}$$

$$\begin{aligned} L_{10}(X_t, Y_t) &= \frac{\partial}{\partial X_t} F(X_t, Y_t) \\ &= \exp\{-D(Z_1, W_2)\} \cdot \frac{\partial}{\partial X_t} D(Z_1, W_2) \cdot K_1, \end{aligned}$$

$$\begin{aligned} L_{11}(X_t, Y_t) &= \frac{\partial^2}{\partial X_t \partial Y_t} F(X_t, Y_t) \\ &= \exp\{-D(Z_1, Z_2)\} \cdot \left[\frac{\partial}{\partial X_t} D(Z_1, Z_2) \cdot \frac{\partial}{\partial Y_t} D(Z_1, Z_2) - \frac{\partial^2}{\partial X_t \partial Y_t} D(Z_1, Z_2) \right] \cdot K_1 \cdot K_2. \end{aligned}$$

The variables W_i, Z_i and K_i are defined as follows:

$$W_1 = -1/\log F_X(u_x),$$

$$W_2 = -1/\log F_Y(u_y),$$

$$Z_1 = -1/\log F_X(x),$$

$$Z_2 = -1/\log F_Y(y),$$

$$K_1 = \frac{-p_1}{\beta_1}\left(1+\gamma_1\frac{(x-u_x)}{\beta_1}\right)^{\frac{-1-\gamma_1}{\gamma_1}}\cdot Z_1^2\cdot\exp\left(\frac{1}{Z_1}\right),$$

$$K_2 = \frac{-p_2}{\beta_2}\left(1+\gamma_2\frac{(y-u_y)}{\beta_2}\right)^{\frac{-1-\gamma_2}{\gamma_2}}\cdot Z_2^2\cdot\exp\left(\frac{1}{Z_2}\right).$$

These likelihood contributions for the bivariate distribution of exceedances are described by the set of parameters:

$\Phi = \{p_1, p_2, \gamma_1, \gamma_2, \beta_1, \beta_2, \alpha\}$.

This leads to the following likelihood function

$$L((x_1,\ldots,x_n),(y_1,\ldots,y_n),\Phi) = \sum_{j,k\in\{0,1\}} L_{jk}(X_t,Y_t)\cdot I_{jk}(X_t,Y_t),$$

where $I_{jk}(X_t,Y_t) = I\left\{(X_t,Y_t)\in A_{jk}\right\}$ (Longin and Solnik, 2001). Therefore the likelihood function for a sample of n independent observations of returns and trading volumes is expressed by

$$L((X_t,Y_t),\Phi) = \prod_{i=1}^{n} L(X_t,Y_t,\Phi).$$

2.10.3 Tail Dependence

Following the above introduction into copula theory for modelling dependence of extremes, the tail dependence coefficient (TDC) can now be calculated for these copulas. In the bivariate context, TDC can be described as simply the probability that a random variable exceeds a particular threshold, given that another associated random variable has already exceeded that threshold.

Let's consider X and Y with marginal distributions F_X and F_Y, and a threshold value u. The coefficient of upper tail dependence, λ_U, is defined as the limit, as the threshold tends to one, of the probability that the cdf of X exceeds u, given that the cdf of Y exceeds u,

$$\lambda_U = \lim_{u\to 1^-} P(F_X(X) > u \mid F_Y(Y) > u) \tag{2.33}$$

$$= \lim_{u\to 1} P\left(X > F_X^{-1}(u) \mid Y > F_Y^{-1}(u)\right)$$

$$= \lim_{u\to 1} \frac{P\left(X > F_X^{-1}(u), Y > F_Y^{-1}(u)\right)}{P\left(Y > F_Y^{-1}(u)\right)}, \tag{2.34}$$

provided that the limit exists. Note, from the definition of copulas, $F_X(X)$ and $F_Y(Y)$ have uniform distributions - i.e. $u \in [0,1]$.

Similarly, the lower TDC of X (denoted λ_L) can be defined as

$$\lambda_L = \lim_{u \to 0} P\left(X \leq F_x^{-1}(u) \mid Y \leq F_y^{-1}(u)\right),$$

provided again that the limit exists (Nelsen, 2006; Shamiri et al., 2011).

If, and only if, $\lambda_U > 0$, it is said that X and Y have upper tail dependence. If $\lambda_U = 0$, then they have upper tail asymptotic independence. Analogously, for the lower tail, the same values apply to indicate lower tail dependence and asymptotic independence, respectively.

Now, in terms of a bivariate copula function, the coefficient of upper tail dependence can be written as

$$\lambda_U = \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1-u},$$

where $C(u,u)$ exists and is the copula of the random vector (X,Y). Then, C has upper tail dependence if $\lambda_U \in (0,1]$, and upper tail independence if $\lambda_U = 0$.

In the same way, the lower tail dependence takes the form

$$\lambda_L = \lim_{u \to 0} \frac{C(u,u)}{u},$$

and copula C has the same values as λ_U apply for its lower tail dependence and independence, respectively (Nelsen, 2006; Haugh, 2016).

For the Gumbel copula, in particular, the tail dependence coefficients are estimated as follows:

$$\begin{aligned} C_\alpha(u,u) &= \exp\left[-\left\{(-\log u)^\alpha + (-\log u)^\alpha\right\}^{\frac{1}{\alpha}}\right] \\ &= \exp\left(2^{\frac{1}{\alpha}} \log u\right) \\ &= u^{2^{\frac{1}{\alpha}}} \end{aligned}$$

thus,

$$\begin{aligned}\lambda_U &= \lim_{u\to 1}\frac{1-2u+C(u,u)}{1-u}\\ &= \lim_{u\to 1}\frac{-2+2^{\frac{1}{\alpha}}u^{2^{\frac{1}{\alpha}}}}{-1}\\ &= 2-2^{\frac{1}{\alpha}}.\end{aligned}$$

Since the Gumbel copula does not model lower tail dependence, one can observe the following result to confirm:

$$\begin{aligned}\lambda_L &= \lim_{u\to 0}\frac{C(u,u)}{u},\\ &= \lim_{u\to 0}\frac{2^{\frac{1}{\alpha}}u^{2^{\frac{1}{\alpha}}}}{1}\\ &= 0.\end{aligned}$$

2.10.3.1 Extreme Correlation

The modelling of the tail dependence structure then allows for a measurement of the extreme correlation between the variables.

The TDC is related to the extreme correlation coefficient, ρ, between the extreme values of X and Y by

$$\rho = 1-\lambda_U^2,$$

ρ ranges from -1 (strong negative correlation) to +1 (strong positive correlation), with special cases of $\lambda_U = 0$ and $\lambda_U = 1$ which indicate asymptotic independence ($\alpha = 1$) and perfect dependence ($\alpha = \infty$), respectively (Longin and Solnik, 2001).

The single downside of the logistic model is that it is symmetric in the two variables, as a result, it may not be suitable in some scenarios (Coles et al., 2001). Nonetheless, it is appropriate for studying situations where both variables tend to exhibit extreme values at the same time. And extreme crypto returns and extreme volumes of trade tend to follow one another very closely.

According to Osterrieder et al. (2016), dependence modelling using copulas is widely used in applications of financial risk evaluations as well as actuarial analysis. It is particularly well-suited for modelling extreme events since EVT requires a focus on the behaviour of the tail of the distribution rather than the entire distribution.

2.11 Fitting the Models and Assessing their Performance

In order to estimate the tail distribution of a set of financial series of returns, we need to plot and analyse the return time series data. When modelling the returns from a price variable one might be interested in calculating the log returns (r_t) as the relative differences between the log prices of this financial variable for time t :

$$r_t = \log(P_t) - \log(P_{t-1}),$$

where P_t is the price at time t.

When evaluating financial risk, we're particularly interested in the losses, i.e. the negative returns. Recall the important relation for calculating the minima: $\min(X_1, X_2, \dots, X_n) = -\max(-X_1, -X_2, \dots, -X_n)$ which allows us to fit the GEVD and GPD models for extreme losses.

We can then observe the positive or negative return series for extreme observations and perform some preliminary tests for stationarity, normality, heteroscedasticity as well as volatility clustering. The Augmented Dickey-Fuller (ADF) test is commonly used to test for stationarity of a time series. The Anderson-Darling test can be used to test for normality, while the Jarque-Bera test implemented in Ren and Giles (2010) 's analysis of Canadian Crude oil prices uses sample skewness and kurtosis to measure the deviation of a distribution from normality.

Descriptive statistics, including skewness and excess kurtosis assist in analysing the tail behaviour of both the negative (losses) and positive (gains) returns. Plotting a histogram of the negative returns (loss values) also aids in displaying the tail behaviour. For instance, a histogram that depicts a heavy-tail on the right tail implies a heavy-tail on the right tail of the loss distribution (Chikobvu and Jakata, 2020). Meaning, the fat-tailed Fréchet class will most likely fit this tail of the return distribution.

Once we've established that asymmetric and heavy tail characteristics exist in our data, we have strong evidence to fit either of the EVT models to the returns. We may then be interested in modelling both the left and right tails of the return distribution separately to capture their distinct properties.

2.11.1 Assessing the 'Goodness of Fit' of the EVT Models

To conduct exploratory data analyses it is possible to assess diagnostic plots which include QQ plots, probability plots, residual plots and density plots to test for the goodness of fit of the GPD or GEVD model.

The QQ plot is a graphical technique that checks whether the sample data is consistent with some known distribution, hence its usefulness in the assessment of the goodness-of-fit (Ren and Giles, 2010). For any class of distributions, the theoretical quantiles are linearly related to the corresponding quantiles of a random sample of data from that class distribution (Wong and Li, 2010). Hence, if the empirical data comes from the underlying distribution, the data points will not deviate significantly from the straight line and the plot will be approximately linear. This can very easily be checked through visual inspection.

For the residual plots, the residuals ought to be approximately exponentially distributed with majority of the points lying on the straight line, indicating a good fit of the model. For these plots, data points are converted to unit exponential residuals and the straight line represents perfectly exponential behaviour (Chinhamu et al., 2017). The residuals are defined as

$$residuals = \begin{cases} \left(1+\gamma\left(\frac{x-u}{\beta}\right)\right)^{-1/\gamma}, & \text{for} \quad \gamma = 0, \\ \exp\left\{-\exp\left[-\left(\frac{x-u}{\beta}\right)\right]\right\}, & \text{for} \quad \gamma \neq 0. \end{cases}$$

Recall that the empirical mean excess plot is required to be linear in u for the GPD model to fit. To assess the goodness-of-fit of the GPD we can analyse the plot of the excess distribution and QQ plot of residuals relative to a certain threshold u (Chinhamu et al., 2015b). When the distribution of the returns exhibits a heavier tail than the exponential distribution, the mean excess function increases, while the opposite (decrease) is observed for a light-tailed returns distribution (Beirlant et al., 2004). This reasoning on the shape of the ME function is conditioned on the incorporation of the exponential distribution due to its memoryless property.

Beirlant et al. (2004) note that for heavy-tailed distributions, the exponential QQ plot displays a convex shape for the much larger data observations with an increasing gradient close to the higher observations, which induces an increasing ME function. This suggests that the tail is heavier than that of an exponential distribution, and that a GPD with $\gamma > 0$ is a good fit. The converse holds for lighter tails since the exponential QQ plot shows a concave shape - suggesting that a GPD model with $\gamma < 0$ would be a good fit. If the data points on the QQ plot display a straight diagonal line, then an exponential distribution would be a very good fit to the tail.

Plotting a fitted density plot (obtained using the estimated parameters) against the observed frequencies of the (log) returns is also a useful method of observing how well the distribution performs in modelling the data. Probability plots, on the other hand, plots values from the empirical distribution function against those of the model distribution. The observations are ordered first, thereafter a point that corresponds to each value is plotted in order to create this plot (Panjer, 2006). To assess how the model fits, the plotted points need to be close to the 45° line that conjoins the origin with coordinates $(1,1)$ in order to conclude a good fit.

2.11.2 Tests for goodness-of-fit

In order to test how adequately the EVT models fit the data, as well as identify an optimal model, several statistical tests for the 'goodness-of-fit' of the models can be implemented in addition to the above graphical tools. These assist in assessing how well the asymptotic distributions describe the extremes. They include the Kolmogorov-Smirnov (K-S) test and the Akaike information criterion (AIC).

2.11.2.1 Kolmogorov-Smirnov (K-S) Test

The Kolmogorov-Smirnov test is a non-parametric test that enables the comparison and measurement of the equality between an empirical distribution of a sample of data and a given reference distribution. This test is especially suitable for a large data samples. For this two-sided one-sample K-S test, the null hypothesis specifies that the true distribution of the data is equal to the reference or hypothesised distribution. The null hypothesis is rejected for small p-values (< 0.05) and a test statistic greater than the critical value, meaning that there is sufficient evidence to conclude that the data does not come from the reference distribution. This test will be applied to test for normality of the data, as well as to check if the fitted EVT distribution models provide a satisfactory fit to the tails of the returns and volumes.

2.11.2.2 Akaike Information Criterion (AIC)

The Akaike information criterion is a criterion commonly used for model selection. It is a likelihood method that estimates prediction error and assesses how well a statistical model fits the data it was generated from. It is frequently used to compare several different models for the data by evaluating the quality of each, relative to one another, in order to determine which is the most optimal model. The model with the lowest AIC value is selected as the one with the best fit. This AIC value is calculated as

$$AIC \quad = \quad 2k - 2\ln\left(\hat{L}\right),$$

where k represents the number of the parameters and $\hat{L}$ represents the maximum value of the likelihood function for the particular model. For this statistical analysis, the most optimal model is one that most accurately fits the returns and volume data with minimum error.

Chapter 3

Data Analysis and Results

Nowadays, data has become more easily accessible, and this has stimulated a growing research interest in higher frequency data. Higher frequency leads to quite a large increase in the number of data observations - which, fortunately for EVT applications, is an added benefit as extreme events or values are rare. For this thesis, historical high frequency data obtained from Cryptocompare (2023) will be utilised.

3.1 An Application to Hourly BTC/USD

The data comprises of hourly observations of both prices (returns) and trading volumes for Bitcoin (BTC) listed on the Kraken cryptocurrrency exchange platform, specifically spanning from 1st July 2017 12 am to 20th April 2020 12 pm. The data inherently includes effects from the COVID-19 pandemic on the crypto market, thus extreme values in crypto prices as well as corresponding trade volumes are bound to be observed due to this extreme event. The Kraken exchange is based in the US and is a trusted and popular exchange all over the world, including in South Africa. The prices are for the currency pair exchange: Bitcoin versus the US Dollar (BTC/USD), while the trading volume is recorded as the volume of Bitcoin units exchanged as well as the USD-volume of Bitcoin exchanged. There are 24 337 observations (n) in total for both variables.

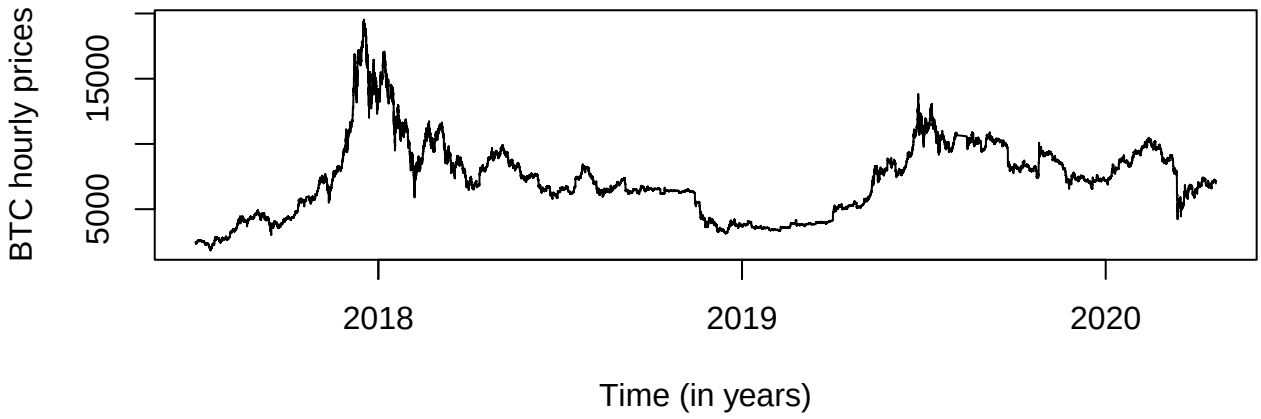


Figure 3.1: Time series of hourly closing Bitcoin prices (in US Dollars).

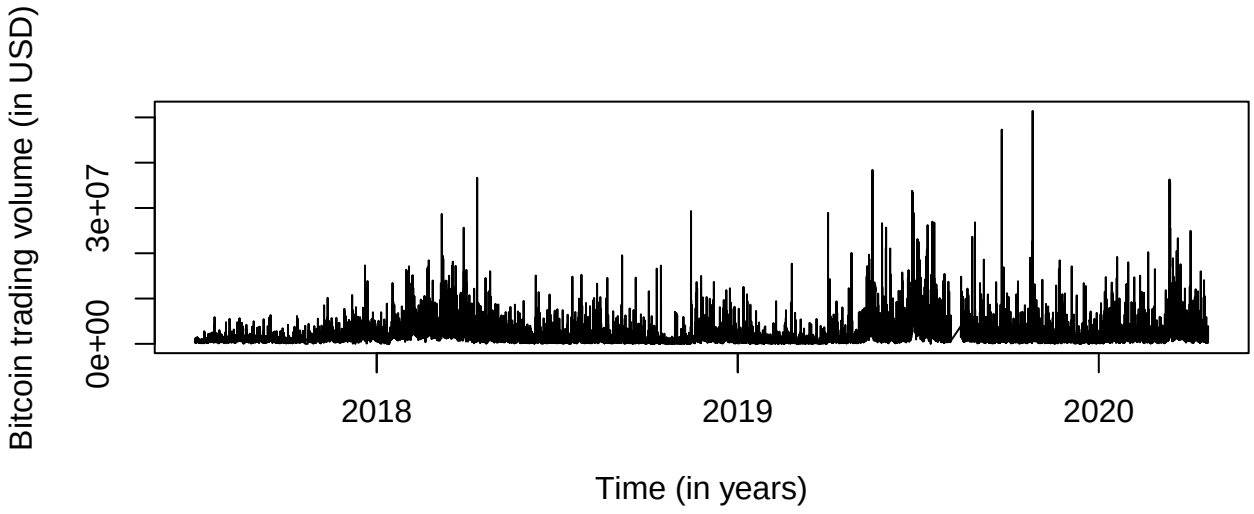


Figure 3.2: Time series of hourly Bitcoin trading volume.

Figure 3.1 is a plot of the hourly Bitcoin closing prices for the three year period, and Figure 3.2 shows the time series plot of hourly Bitcoin trading volume (in USD). Figure 3.1 displays a number of periodic ups-and-downs, and in particular a substantial increase in Bitcoin prices. Large movements are present in Figure 3.2 as well. From a purely visual assessment of these two figures, one can deduce that these variables are non-stationary as there exists trends (long-term increases or decreases in data) and patterns of seasonality (a pattern or cycle that reoccurs at a known frequency, e.g. a month or year) in the two series.

Heteroscedasticity is a state where data points are scattered unequally and there exists large differences in the size of the data observations. This means that the standard deviations of the variable are not equal - thus variance over the data is non-constant. Volatility clustering, on the other hand, refers to the idea that large changes (e.g. price changes) tend to be followed by large changes (of either sign), and small changes tend to be followed by small changes. Essentially, large changes tend to cluster together, and small changes cluster together. This concept is regarded as an important stylised fact in financial markets.

Stationarity in a time series means that the variance is constant over time, with no presence of trends. This concept is vital as the overall behaviour of the data should be the same (even though the data values are not the same) so as to preserve the statistical properties of the data over time.

Over the years, the field of finance has become more accustomed to using log data as opposed to raw data in its applications. This is owing to the favourable statistical properties that result from applying a log transformation. In order to account for and eliminate volatility clustering and heteroscedasticity and ultimately achieve stationarity, the raw data needs to be transformed using a data adjustment procedure described by Gallant et al. (1992), and further used by Chan et al. (2022).

The raw data is transformed by backward differencing the logarithmic prices and logarithmic trade volumes; i.e. the hourly Bitcoin returns at hour t (X_t) are calculated as

$$X_t = \ln(P_t) - \ln(P_{t-1}),$$

where P_t is the Bitcoin closing price at the current hour t and P_{t-1} is the price at the previous hour $t-1$. The hourly volumes are detrended as follows:

$$Y_t = \ln(V_t) - \ln(V_{t-1}),$$

where V_t represents the hourly trading volumes (in USD) at hour t. Thus, Y_t is the detrended trading volume variable. Detrending entails the elimination of the effects of trend from data, enabling the identification of cyclical fluctuations and any other potential subtrends or patterns that could distort the analysis of financial data.

The adjusted log returns are plotted in Figure 3.3. Figure 3.3 shows that the adjusted returns exhibit

much less seasonality and trends, which is owing to the difference and logarithmic transformations. Still, one cannot deny that there is some evidence of a few turbulent periods visible in Figure 3.3, especially the striking fluctuation that stands out during 2020 - which may be owing to the effects of the COVID-19 pandemic.

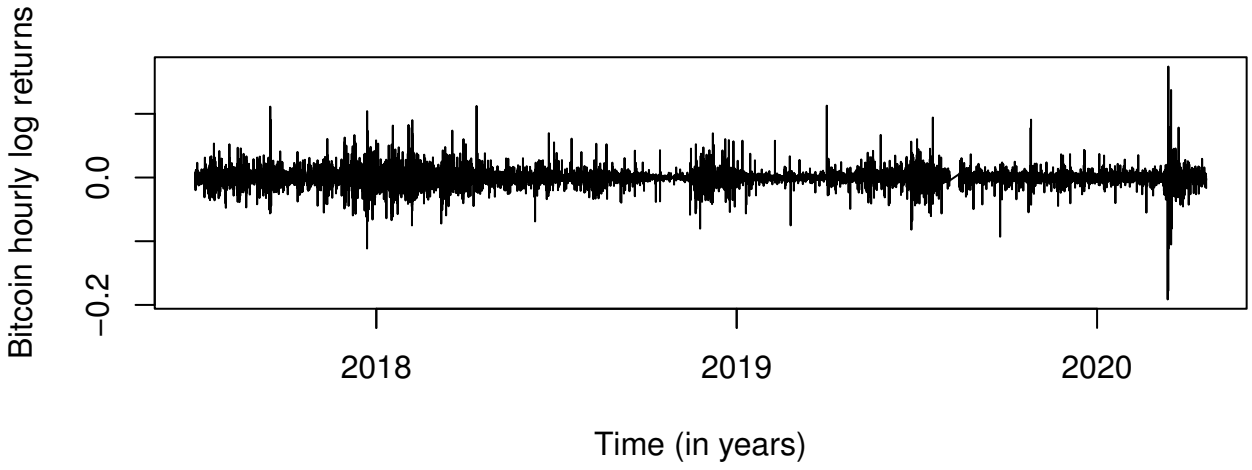


Figure 3.3: Hourly adjusted Bitcoin log returns.

The Augmented Dickey-Fuller (ADF) test is then performed on the adjusted returns and volumes in order to test for stationarity. For this statistical significance test, under the null hypothesis, the time series has a unit root (i.e. it is stationary) (Dickey and Fuller, 1979). The more negative the test statistic, the lower the p-value (Chikobvu and Jakata, 2020). The test results confirm stationarity at the 1% level of significance for both returns and volumes.

```
data:  diffvol Dickey-Fuller = -39.483, Lag order = 28,
    p-value = 0.01 alternative hypothesis: stationary
Warning message: In adf.test(diffvol) : p-value smaller than printed p-value
```

Figure 3.4: ADF test results.

A descriptive summary of log price returns and detrended trade volumes for Bitcoin against the USD for the time period 11am 01 July 2017 to 11am 20 April 2020 is presented on Table 3.1. The statistics include, among many, the minimum, maximum mean, kurtosis, skewness, and standard deviation.

From this table, a large kurtosis value (significantly greater than 3) is observed; which indicates a much heavier tail and a peaked density for the distribution of returns compared to a normal distribution.

Notably, the returns are negatively skewed. The standard deviation (SD) equal to 0.01 gives an estimate of volatility equal to 1%, in terms of variation in returns. The detrended volume, on the other hand, is positively skewed with a very small kurtosis value lower than 3 - suggesting a much lighter tail for the distribution of volumes.

Table 3.1: Summary statistics of the log returns and the detrended volume of BTC/USD.

Statistics	Log returns	Detrended volume
Observations	$24,336$	$24,336$
Minimum	-0.191	-6.051
Q1	-0.003	-0.51
Median	0.000	-0.044
Mean	0.000	0.000
Q3	0.003	0.464
Maximum	0.174	5.644
Skewness	-0.12	0.35
Kurtosis	31.07	1.77
Standard deviation	0.01	0.81
Variance	0.00	0.656
Range	0.37	11.7
IQR	0.006	0.974

3.1.1 Model Fitting and Extreme Correlation and Dependence

Parameter estimates for the GPD model of return and volume exceedances are provided on Tables 3.2 and 3.3 and interpreted. A series of different fixed threshold levels are considered for both negative and positive returns. The threshold values, u, are selected according to the distance of exceedances from the empirical mean obtained from the summary statistics (as presented on Table 3.1). That is, the threshold values are selected as $\pm 0\%, \pm 1\%, \pm 2\%, \pm 3\%, \pm 4\%$ and $\pm 5\%$ above or below the mean, which corresponds to log return exceedances $\pm 0.01, \pm 0.02, \ldots, \pm 0.05$ above or below the mean of the returns variable. The last row of each table computes the optimal threshold level - labelled with a $*$. These particular threshold values were simulated using the optimal threshold selection procedure in Section 2.4.1.2.

The negative returns are simply computed as

$$X_t = -X_t$$

Table 3.2 summarises estimation results of negative return exceedances (i.e. returns falling below the threshold u^{ret}) and positive volume exceedances (i.e. volumes that exceed threshold u^{vol}), and Table 3.3 summarises those of positive return exceedances (i.e. returns above the threshold u^{ret}) and positive volume exceedances. The GPD parameter estimates were obtained using the MLE method in R (R Development Core Team, 2020). The rows denoted (a) give the parameter estimates and the ones denoted (b) give standard errors.

Table 3.2: Parameter estimates for negative return exceedances and positive volume exceedances for BTC/USD.

u^{ret}		p^{ret}	β^{ret}	γ^{ret}	u^{vol}	p^{vol}	β^{vol}	γ^{vol}	$\lambda^{ret/vol}$	$\rho^{ret/vol}$
-5%	(a)	0.002	0.015	0.299	3.138	0.002	0.6671	0.0469	0.9981	0.004
	(b)		0.004	0.187			0.18	0.2188	0.015	
-4%	(a)	0.004	0.009	0.397	2.657	0.004	0.4888	0.1321	0.996	0.008
	(b)		0.002	0.142			0.0814	0.1336	0.023	
-3%	(a)	0.010	0.011	0.215	2.227	0.01	0.4693	0.1011	0.99	0.02
	(b)		0.001	0.070			0.0466	0.075	0.023	
-2%	(a)	0.026	0.01	0.177	1.744	0.026	0.4694	0.064	0.8215	0.325
	(b)		0.001	0.043			0.0274	0.0425	0.017	
-1%	(a)	0.076	0.009	0.148	1.168	0.076	0.5363	-0.0143	0.919	0.155
	(b)		0.00	0.026			0.0168	0.0208	0.008	
-0%	(a)	0.488	0.006	0.000	-0.035	0.488	0.7228	-0.1009	0.5207	0.728
	(b)		0.00	0.003			0.0079	0.0061	0.00	
–1.975%*	(a)	0.027	0.009	0.177	**1.727***	0.027	0.4686	0.0636	0.9724	0.054
	(b)		0.001	0.043			0.0268	0.0417	0.012	

Table 3.3: Parameter estimates for positive return exceedances and positive volume exceedances for BTC/USD.

u^{ret}		p^{ret}	β^{ret}	γ^{ret}	u^{vol}	p^{vol}	β^{vol}	γ^{vol}	$\lambda^{ret/vol}$	$\rho^{ret/vol}$
$+0\%$	(a)	0.505	0.006	0.000	-0.061	0.505	0.7301	-0.103	0.4868	0.763
	(b)		0.000	0.003			0.0078	0.0059	0.00	
$+1\%$	(a)	0.076	0.008	0.209	1.168	0.076	0.536	-0.0143	0.9187	0.156
	(b)		0.000	0.027			0.0168	0.0208	0.011	
$+2\%$	(a)	0.024	0.009	0.237	1.775	0.024	0.4881	0.0463	0.8425	0.29
	(b)		0.001	0.05			0.029	0.0424	0.016	
$+3\%$	(a)	0.009	0.012	0.266	2.278	0.009	0.4732	0.1029	0.9316	0.132
	(b)		0.001	0.087			0.0498	0.08	0.015	
$+4\%$	(a)	0.004	0.016	0.161	2.726	0.004	0.4393	0.2138	0.9568	0.085
	(b)		0.003	0.215			0.0852	0.164	0.018	
$+5\%$	(a)	0.002	0.018	0.141	3.138	0.002	0.6671	-0.0469	0.9561	0.086
	(b)		0.004	0.164			0.180	0.2188	0.025	
+1.85%*	(a)	0.028	0.009	0.210	**1.71***	0.028	0.469	-0.0621	0.9713	0.057
	(b)		0.001	0.045			0.026	0.0409	0.014	

Most of the EVI estimates (γ) in Tables 3.2 and 3.3 are positive values - which gives a good indication that the extreme returns and volumes are well described by a heavy-tailed distribution (i.e. a GPD). In Tables 3.2 and 3.3, consider moving from the 0% threshold to the extreme left and right tails of the distributions (by increasing the threshold) the tail dependence coefficient (λ) between returns and trading volumes gradually increases, while the extreme correlation gradually declines.

The simulated optimal thresholds are 1.85% and 1.71 for positive return exceedances and volumes respectively, and -1.975% and 1.718 for negative return exceedances and volumes repectively. These optimal thresholds correspond to tail probabilities 0.028 and 0.027 respectively - the first tail probaility can be interpreted as the number of observations in the tail of positive returns divided by the total number of observations (n). Similarly, 0.027 corresponds to the number of observations in the tail of negative returns divided by n.

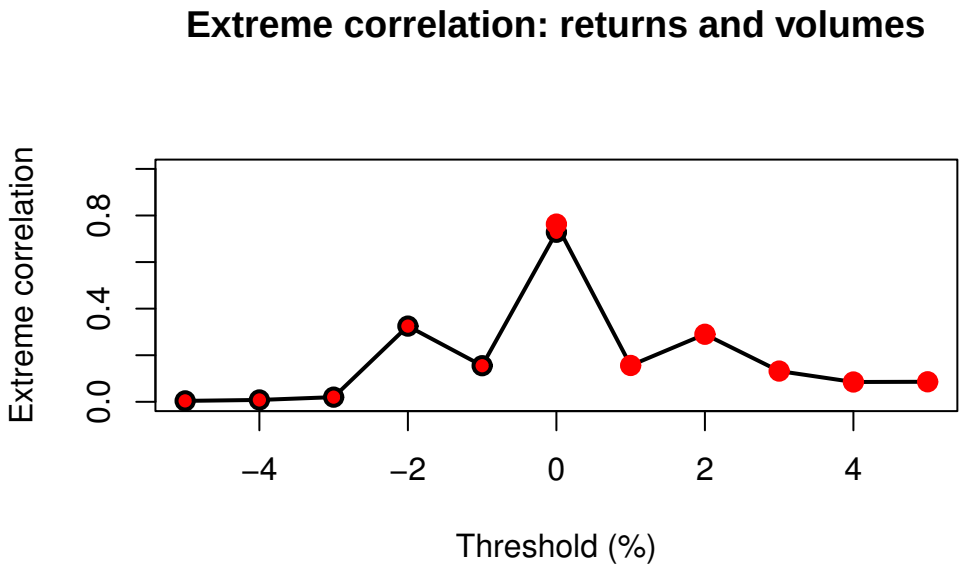


Figure 3.5: A plot of the extreme correlation (ρ) between the returns and volume exceedances for varying threshold values (u).

Figure 3.5 plots a visual comparison of the extreme correlations derived from the extreme bivariate models for BTC/USD returns and volumes. The plot shows that Bitcoin returns and trading volume exhibit a very strong positive extreme correlation at the 0% threshold, with the extreme correlation between positive returns and volumes being slightly higher than that between negative returns and volumes. As the one moves to the left and right tails for the range of fixed thresholds, it is noticeable that the correlation decreases and becomes very weak but still positive. Overall, a symmetric structure can be deduced for the positive and negative return exceedances.

3.2 Assessing Model Robustness

In order to test the robustness of the results obtained from the univariate and bivariate models, an application is implemented on a different kind of high frequency data - daily BTC/USD. This will assist in assessing whether similar results or features can be obtained, and whether the models can adjust to daily data. The daily data spans the period 6 October 2013 - 29 November 2020 from the Coinbase exchange. Coinbase is one of the world's largest trading platforms in terms of trading volume across a wide range of crypto assets, boasting a huge presence in South Africa as well. This has motivated the choice of using BTC/USD from the Coinbase exchange to verify the results presented in Section 3.1.

The time series of daily Bitcoin returns and trading volumes are plotted in Figures 3.6 and 3.7. The daily Bitcoin prices depict a steady pattern of low prices for years 2013 to around 2017. From 2018 onward, the prices fluctuate in a very irregular and unstable manner as the Bitcoin market experienced some very turbulent conditions. The same inference can be deduced for trading volumes in Figure 3.7. The adjusted log returns are plotted in Figure 3.8.

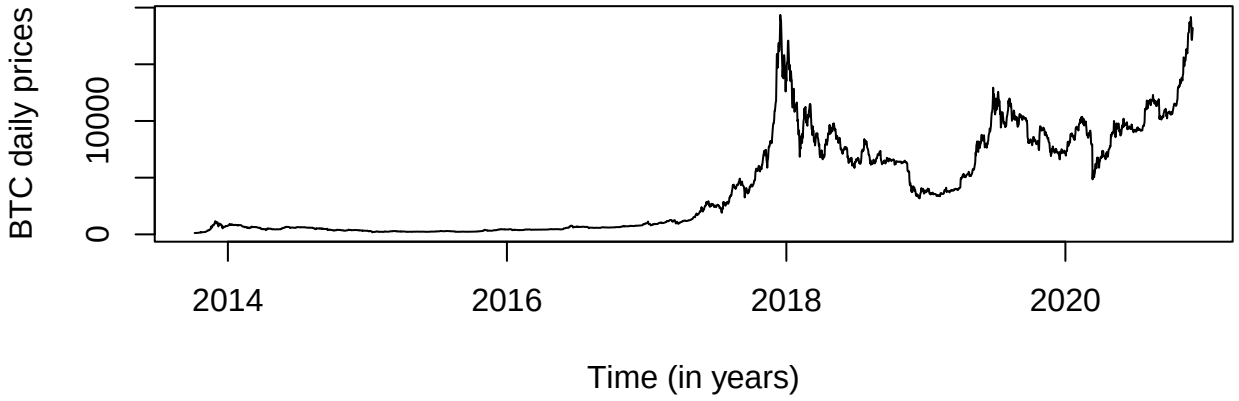


Figure 3.6: Time series of daily Bitcoin prices.

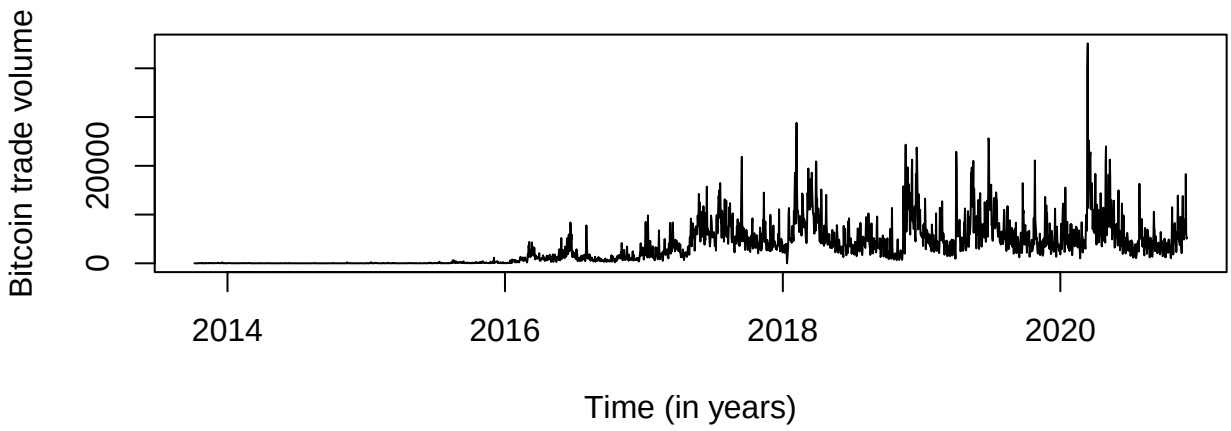


Figure 3.7: Time series of daily Bitcoin trade volumes (in BTC units).

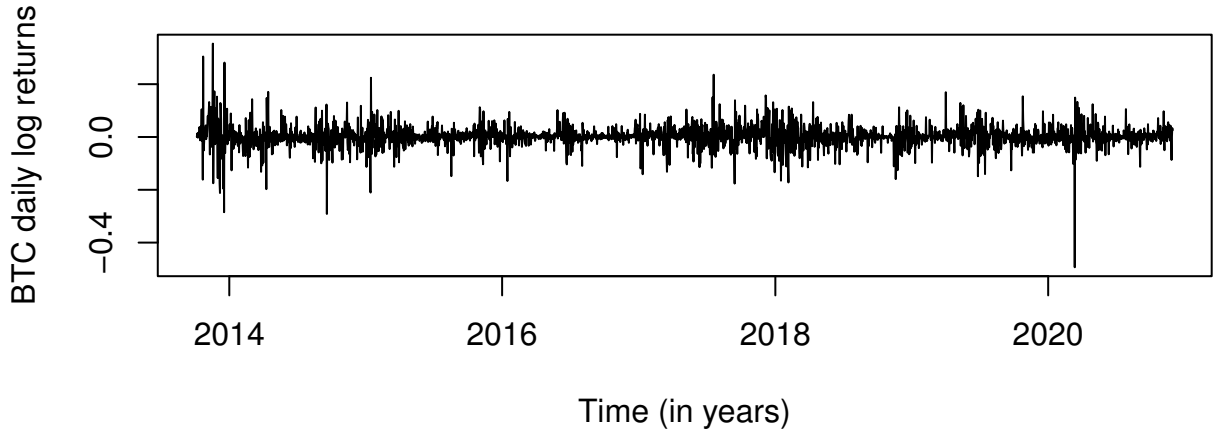


Figure 3.8: Daily adjusted log returns of BTC/USD.

Table 3.4 records the MLE estimation results of negative return exceedances and volume exceedances, and Table 3.5 records those of positive return exceedances and volume exceedances.

Table 3.4: MLE Parameter estimates for daily BTC/USD negative return exceedances and positive volume exceedances.

u^{ret}		p^{ret}	β^{ret}	γ^{ret}	u^{vol}	p^{vol}	β^{vol}	γ^{vol}	$\lambda^{ret/vol}$	$\rho^{ret/vol}$
-5%	(a)	0.082	0.03	0.186	1.13	0.082	0.83	0.221	0.956	0.086
	(b)		0.003	0.077			0.089	0.084	0.017	
-4%	(a)	0.106	0.033	0.119	0.952	0.106	0.745	0.249	0.955	0.089
	(b)		0.003	0.059			0.072	0.078	0.016	
-3%	(a)	0.1398	0.034	0.099	0.762	0.1398	0.671	0.273	0.937	0.123
	(b)		0.003	0.05			0.058	0.071	0.017	
-2%	(a)	0.195	0.031	0.119	0.574	0.195	0.563	0.317	0.785	0.383
	(b)		0.002	0.047			0.042	0.062	0.016	
-1%	(a)	0.288	0.027	0.17	0.345	0.288	0.538	0.276	0.952	0.144
	(b)		0.002	0.044			0.031	0.047	0.014	
-0%	(a)	0.456	0.022	0.226	0.038	0.456	0.592	0.17	0.532	0.718
	(b)		0.001	0.038			0.025	0.031	0.013	

Table 3.5: MLE Parameter estimates for daily BTC/USD positive return exceedances and positive volume exceedances.

u^{ret}		p^{ret}	β^{ret}	γ^{ret}	u^{vol}	p^{vol}	β^{vol}	γ^{vol}	$\lambda^{ret/vol}$	$\rho^{ret/vol}$
$+0\%$	(a)	0.535	0.0237	0.146	-0.068	0.535	0.597	0.153	0.485	0.765
	(b)		0.001	0.031			0.023	0.028	0.00	
$+1\%$	(a)	0.353	0.025	0.142	0.22	0.353	0.545	0.24	0.91	0.172
	(b)		0.001	0.039			0.028	0.04	0.014	
$+2\%$	(a)	0.238	0.028	0.117	0.457	0.238	0.547	0.297	0.902	0.187
	(b)		0.002	0.045			0.036	0.053	0.016	
$+3\%$	(a)	0.17	0.029	0.126	0.648	0.17	0.6	0.305	0.816	0.334
	(b)		0.002	0.053			0.048	0.066	0.017	
$+4\%$	(a)	0.123	0.028	0.165	0.848	0.123	0.755	0.226	0.915	0.164
	(b)		0.002	0.07			0.067	0.07	0.018	
$+5\%$	(a)	0.082	0.035	0.072	1.13	0.082	0.83	0.221	0.916	0.16
	(b)		0.003	0.066			0.089	0.084	0.021	

The extreme correlation structure of the daily BTC exceedances of returns and volumes is plotted in Figure 3.9. The plot shows a fairly symmetric structure - which confirms the observation concluded for the hourly Bitcoin data in Section 3.1.1. At the 0% threshold a very strong positive correlation can be seen and, moving towards the left and right tails the correlation gradually decreases for the extreme values of both negative and positive returns. These extreme values are induced by extreme events in the Bitcoin market, i.e. negative returns and trading volumes have a very low but positive correlation during extreme events, and the same can be said for the relationship between positive returns and volumes.

Essentially, similar results for the dependence and extreme correlation structures are obtained for both an hourly and daily frequency of observations in Bitcoin. Thus, the univariate and bivariate models produce robust results.

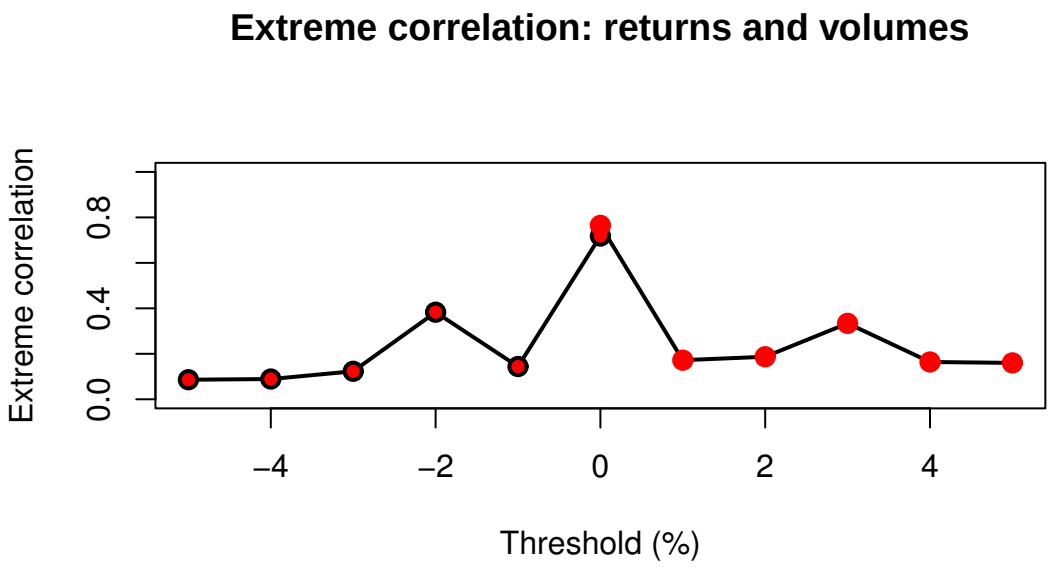
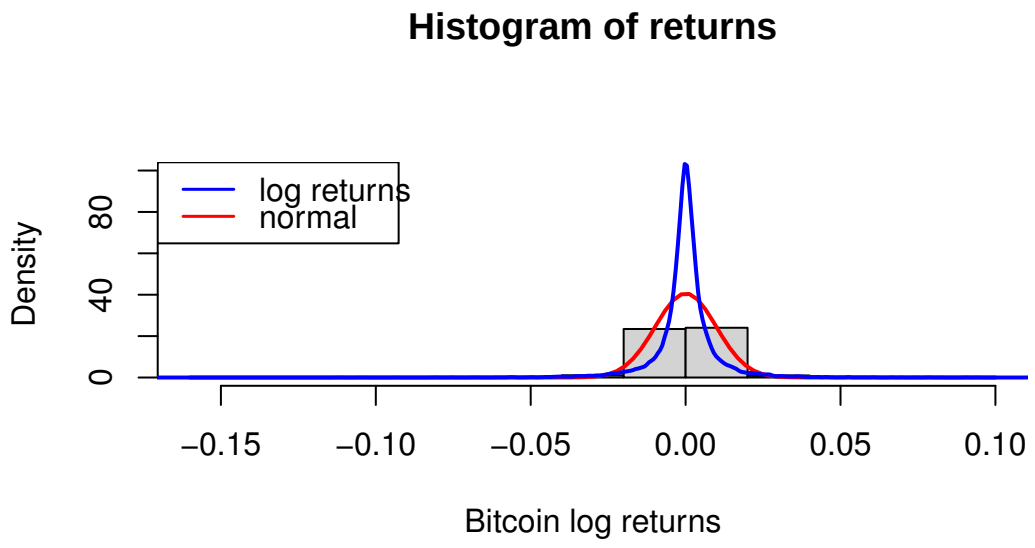


Figure 3.9: Extreme correlation structure between the returns and volume exceedances for daily BTC/USD.

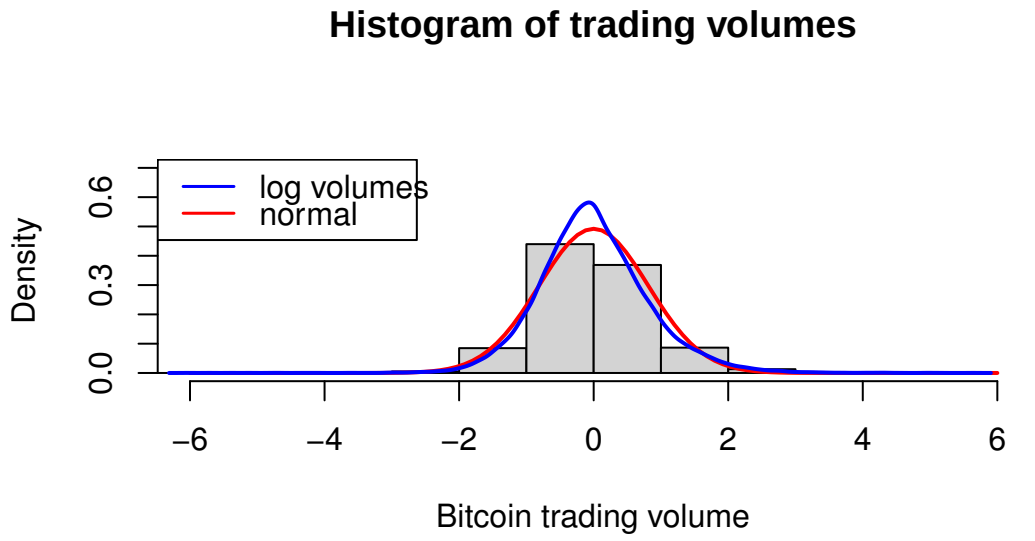
3.3 Exploring a Different Threshold Selection Method

3.3.1 An Application to BTC/USD

Histograms of both hourly returns and trading volumes are plotted against the normal distribution in Figure 3.10 in order to determine whether the data comes from a heavy-tailed distribution. The red density curve represents the normal distribution and a clear deviation from the normal distribution can be seen as the peaks of the normal curves and the histograms appear to differ. Figure 3.11 shows the QQ plots of residuals against the normal distribution. Most of the quantile points do not lie on the theoretical normal line, especially at the far ends, thus the assumption of normality for both returns and volumes is unrealistic, and a heavy-tailed distribution would be suitable for the data.



(a) Histogram of returns.



(b) Histogram of trading volumes.

Figure 3.10: Histograms of returns (a) and trading volumes (b) compared with a normal distribution.

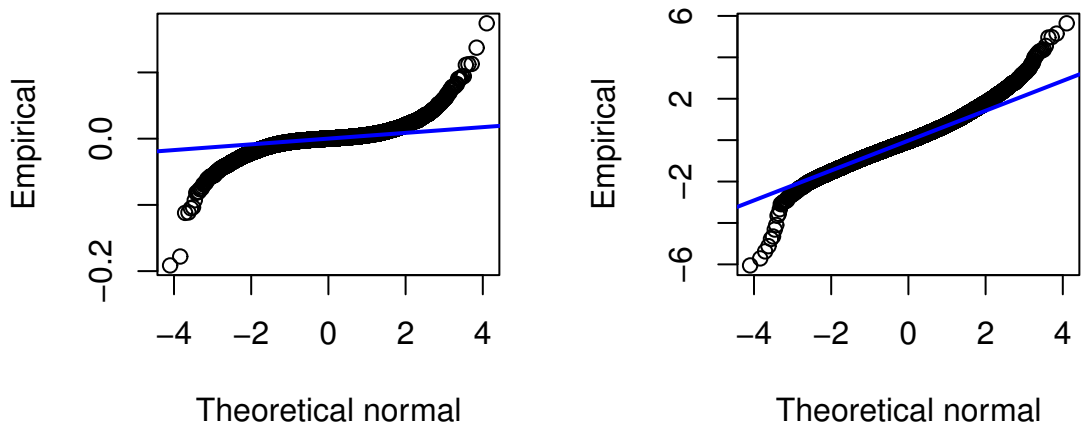


Figure 3.11: Normal QQ plots for returns (left) and trading volumes (right).

The Anderson-Darling test was conducted to see if the hourly returns and volumes are not normally distributed. Under this test, the null hypothesis specifies that the data follows a normal distribution, and the observed p-values in Table 3.6 are significantly low and confirm that the data is not non-normal thus rejecting the null hypothesis at the 5% level of significance. Essentially, this suggests that the quantiles of the empirical distributions are very far from the quantiles of the normal distribution, at the tails of the distributions - implying a leptokurtic (heavy) tail behaviour in the distributions of the returns and volumes.

The Kolmogorov-Smirnov Test for normailty was also conducted on both returns and volumes to confirm confidence in the non-normailty of the crypto data. This test is used to test whether or not a sample of data comes from a certain distribution. It calculates the distance from the cdf of the given reference distribution and the empirical df of the data. Here, the test is used to test for normality thus it has the same null hypothesis as the one above - which is rejected for p-values less than 0.05. From Table 3.6, the p-values of both variables are significantly lower than 0.05, indicating that neither returns nor volumes are normally distributed.

Table 3.6: Results for the Anderson-Darling and Kolmogorov-Smirnov Tests for normality for BTC/USD returns and volumes.

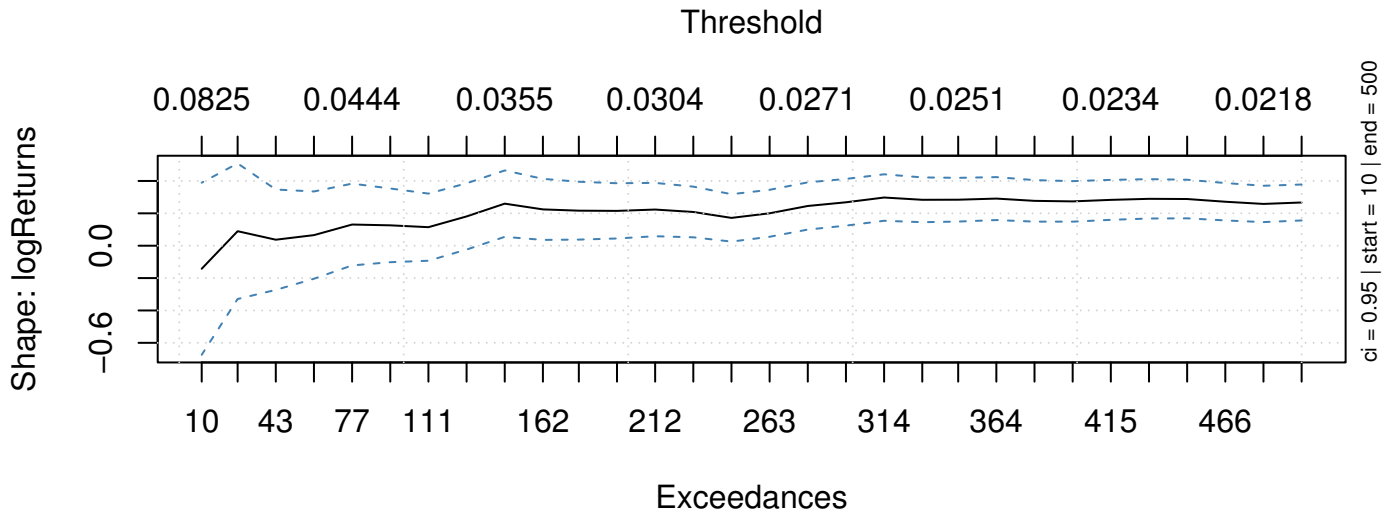
	Anderson-darling test		**K-S test**	
	Test statistic	**p-value**	**Test statistic**	**p-value**
returns	1223.9	< 0.00001	0.48031	< 0.00001
trading volumes	69.778	< 0.00001	0.073506	< 0.00001

In order to select a sufficient threshold to fit a GPD model, the mean excess plot can be used. Figures 3.13 and 3.14 display mean excess plots of returns and volumes. Both figures depict a steep downward trend followed by an upward trend at the end. Thus, for the high threshold (u) values, the mean excess function is increasing or linear in u. In figure 3.13 it is observable that a threshold within 0 and 0.05 seems to be a suitable choice for positive returns, while for negative returns, a threshold greater than or 0, and up to 0.07 for a sufficiently high threshold would be reasonable.

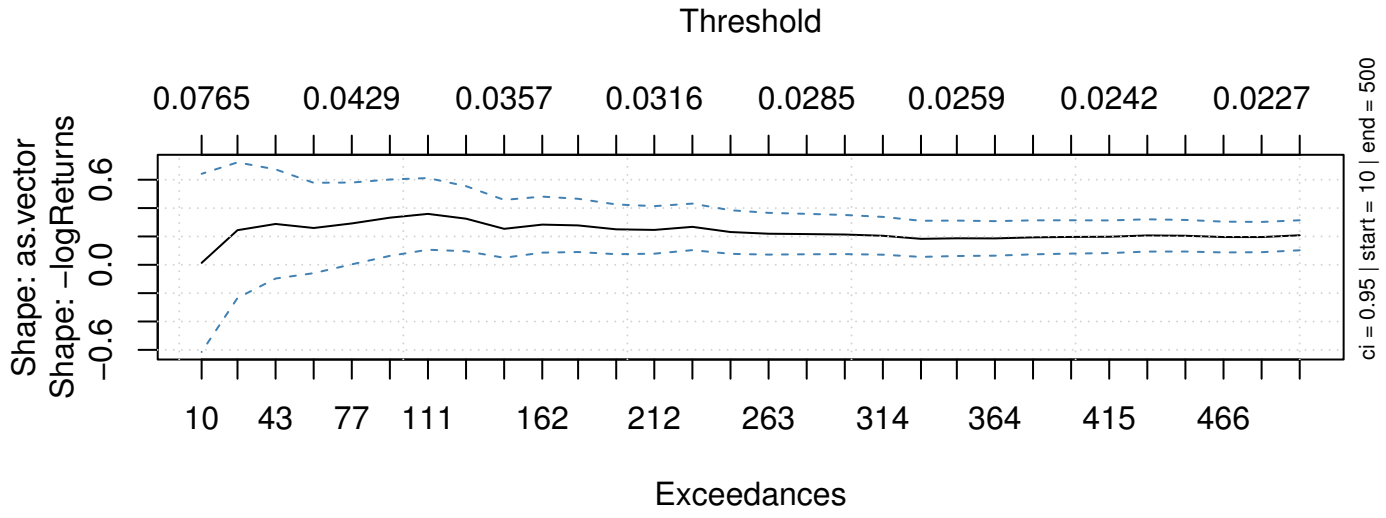
For the volumes, the mean excess plot on the left in Figure 3.14 does not depict a clear linear pattern, thus the 20 highest trading volume observations can then be emitted (temporarily) to obtain the plot on the right. A linear pattern is now more visible on the plot on the right and a suitable threshold would be between the values 2.3 and 3.5. Figure 3.12 displays plots of the estimates of the shape parameter (EVI) against varying threshold values for each variable. The lower and upper blue dashed lines represent confidence intervals at the 95% level. The horizontal axis also includes a count of the number of exceedances above the corresponding threshold. This plot aids in determining a suitable threshold as it depicts for which threshold the shape parameter appears to be stable. It is sometimes referred to as a shape plot.

Selecting an appropriate threshold is not a simple task and involves a considerable level of subjective thinking or judgment. For this reason, one cannot rely on a single method or graphical tool to select a threshold. Several techniques should be used in order to boost confidence in the results obtained and the subsequent conclusions deduced.

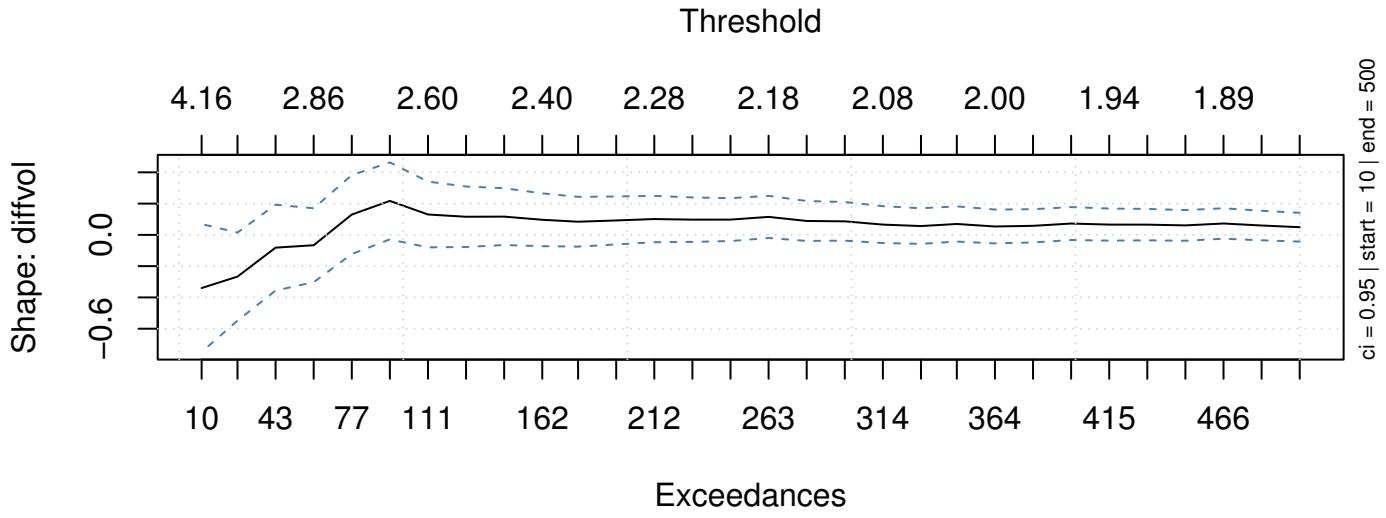
It can be observed from Figure 3.12 (a) that the estimated EVI parameter stabilises at the thresholds 0.029, 0.036 and 0.075 for positive returns, while for negative returns this seems to appear at thresholds 0.03, 0.04, and 0.07, and for trading volumes, thresholds 2.18, 2.67 and 3.3 seem stable. It is good to note, however, that these are not the only threshold values at which the shape parameter appears stable, but for this analysis they will suffice. These threshold selection ranges are then used to identify the most suitable thresholds to fit GPD model to the data. Their GPD MLE parameter estimates are given in Table 3.7.



(a) Positive returns.

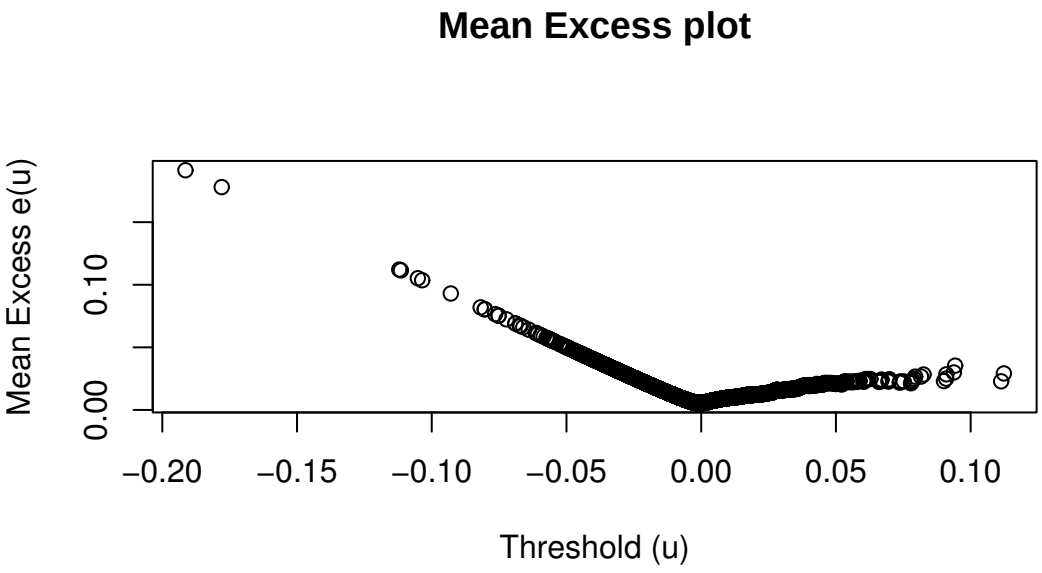


(b) Negative returns.

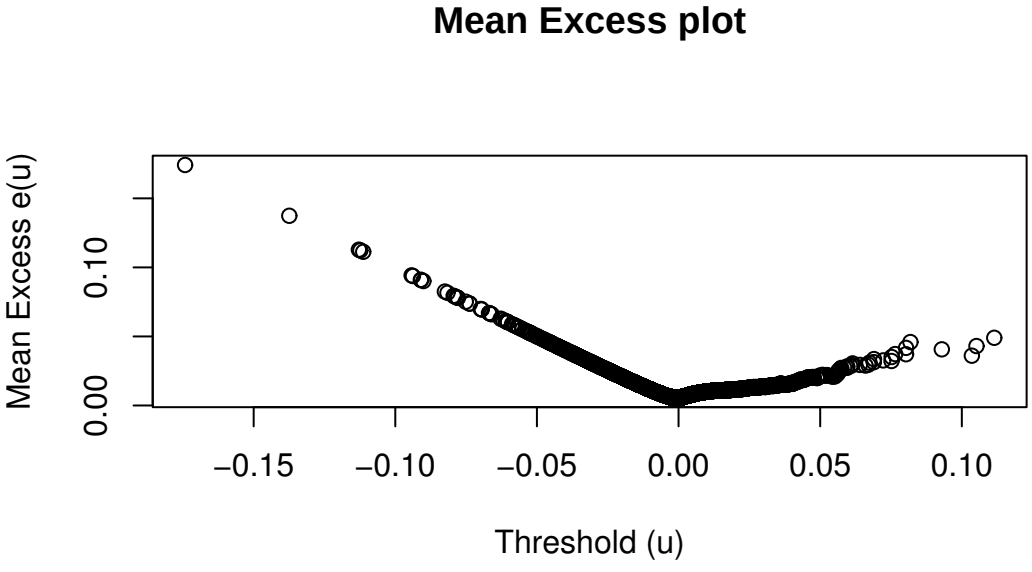


(c) Trading volumes.

Figure 3.12: Plots of estimates of the shape parameter (EVI) against varying thresholds for positive returns (a), negative returns (b) and trading volumes (c).



(a) Mean excess for positive returns.



(b) Mean excess for negative returns.

Figure 3.13: Plots of the mean excess for positive and negative returns of Bitcoin.

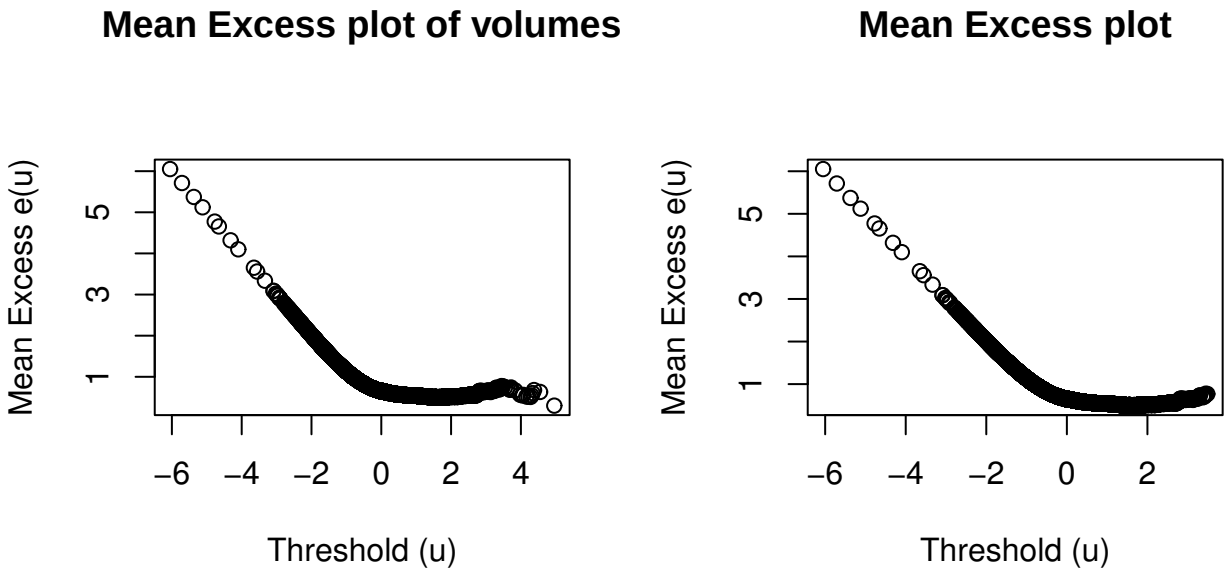


Figure 3.14: Mean excess plots for trading volume.

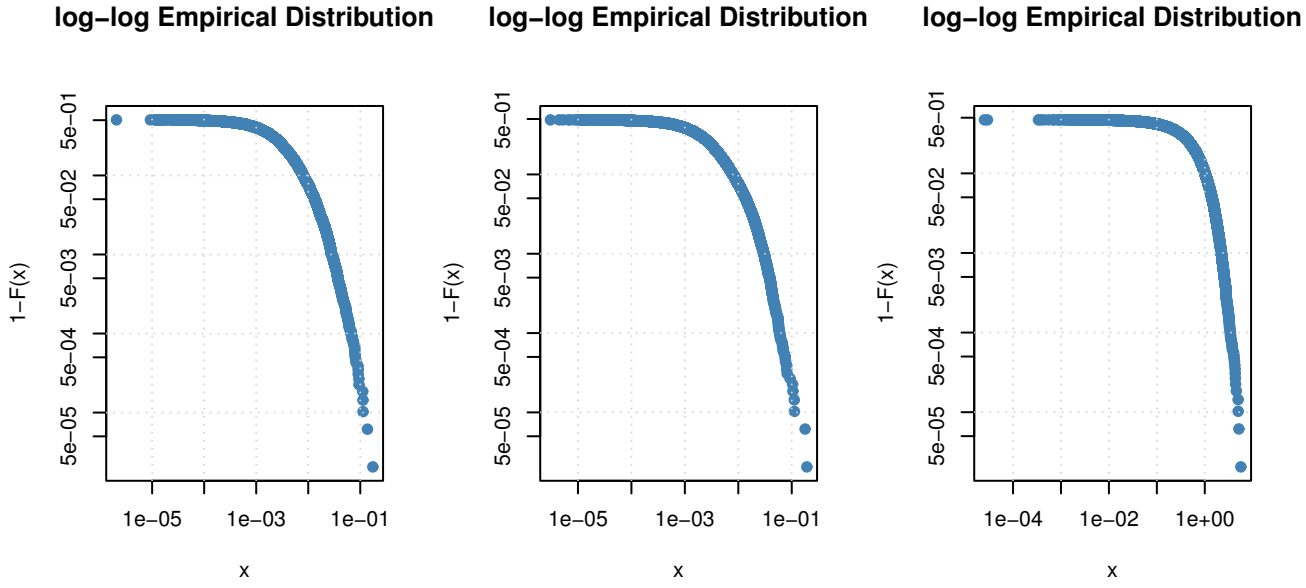


Figure 3.15: Empirical distribution plots of positive and negative returns, and volumes, respectively.

The empirical distribution plots of the variables, positive returns, negative returns and volumes shown in Figure 3.15 indicate that the tails of these variables are approximately linear, suggesting a Pareto

behaviour. These graphical and numerical analyses justify the fitting of the GPD as they provide strong evidence of fat tails and non-normality.

The mean residual life plots in Figure 3.16 show and confirm the results from the mean excess plots. The red lines display 95% confidence intervals. The plots display a steep decline initially, following that, an approximately linear pattern can be seen above thresholds thresholds $u = 0$ up to $u = 0.05$, $u = 0$ up to $u = 0.07$, and $u = 2.2$ up to $u = 3.3$, for positive returns, negative returns and volumes respectively. At these regions the GPD model is valid for fitting. The GPD models are now fitted to the data.

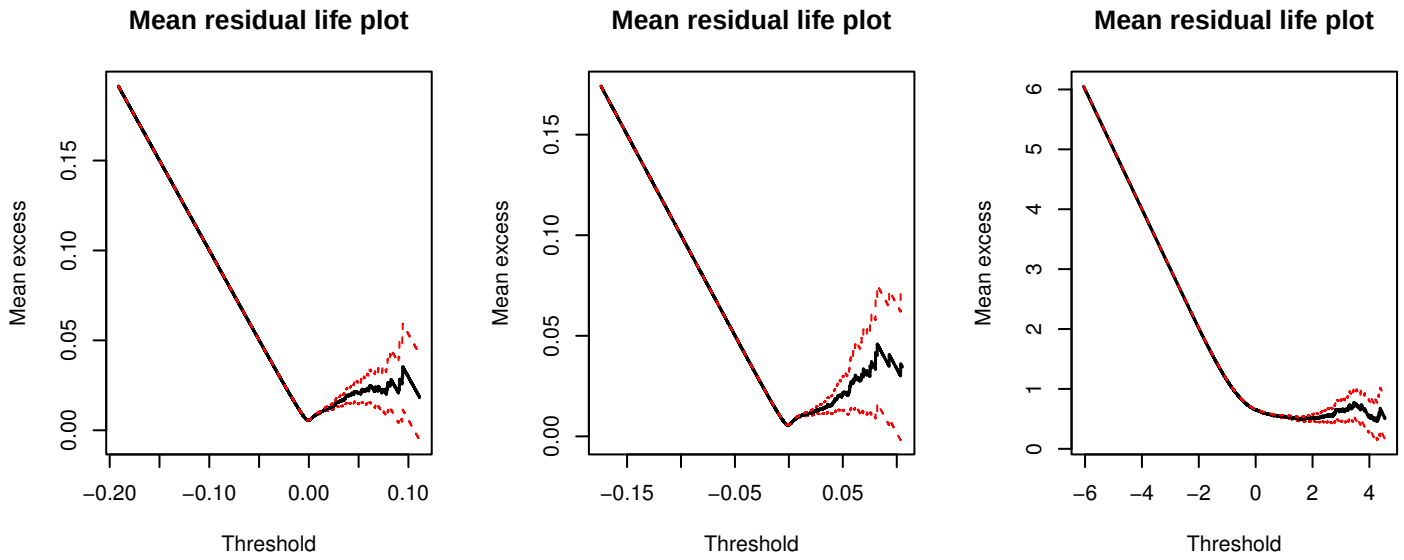


Figure 3.16: Mean residual life plots for positive and negative returns, and volumes, respectively.

Table 3.7: MLE Parameter estimates of GPD models fitted using the selected threshold.

	u	N_u	p	$\hat{\gamma}$	$se(\hat{\gamma})$	$\hat{\beta}$	$se\left(\hat{\beta}\right)$
positive returns	0.029	235	0.01	0.203	0.08	0.012	0.001
	0.036	139	0.006	0.251	0.121	0.013	0.002
	0.075	18	0.001	0.2	0.328	0.018	0.007
negative returns	-0.03	245	0.01	0.215	0.07	0.011	0.001
	-0.04	108	0.004	0.397	0.142	0.01	0.002
	-0.07	14	0.001	0.181	0.374	0.027	0.012
volumes	2.18	264	0.011	0.119	0.073	0.449	0.043
	2.67	97	0.004	0.143	0.138	0.482	0.082
	3.3	28	0.001	-0.28	0.198	0.921	0.248

To assess the fitness of the GPD models to returns and volumes, diagnostic plots are plotted in Figures 3.17 to 3.19. For this exploratory data analysis, thresholds $u = 0.029$, $u = 0.03$, and $u = 2.67$ (for

positive returns, negative returns and volumes respectively) are selected as they are not too high that the number of exceedances is too low and variance is high due to very few extremes being observed, but also not too low to ensure that only extreme values are considered in our estimation and bias is minimised.

For the positive returns variable, the tail index estimates do not vary much for the proposed threshold values, while the trading volume variable exhibits lower tail index estimates. This indicates lighter tails for the GPD, especially for the negative EVI at the threshold $u = 3.3$. For positive returns with threshold $u = 0.029$ and EVI estimate $\hat{\gamma} = 0.203$, the GPD model is fitted to 235 exceedances, for negative returns with $u = 0.03$ and EVI estimate $\hat{\gamma} = 0.215$, the GPD model is fitted to 245 exceedances, and for trading volumes with $u = 2.67$ and EVI estimate $\hat{\gamma} = 0.143$, the GPD model is fitted to 264 exceedances.

Figure 3.17 displays a good fit of the GPD model to the positive returns at the threshold $u = 0.029$, the points on the probability plot are linear or very close to the 45-degree line. Most of the points in the QQ plot lie quite close to the straight line, with only a few departures from the line, creating an approximately linear pattern - which indicates a good fit in the tails.

The fitted density plot appears to be consistent with the empirical one - confirming what was observed in the probability plot. These same observations can be inferred from Figures 3.18 and 3.19 as well. Clearly, the diagnostic plots are in favour of the fitted GPD models as they perform well.

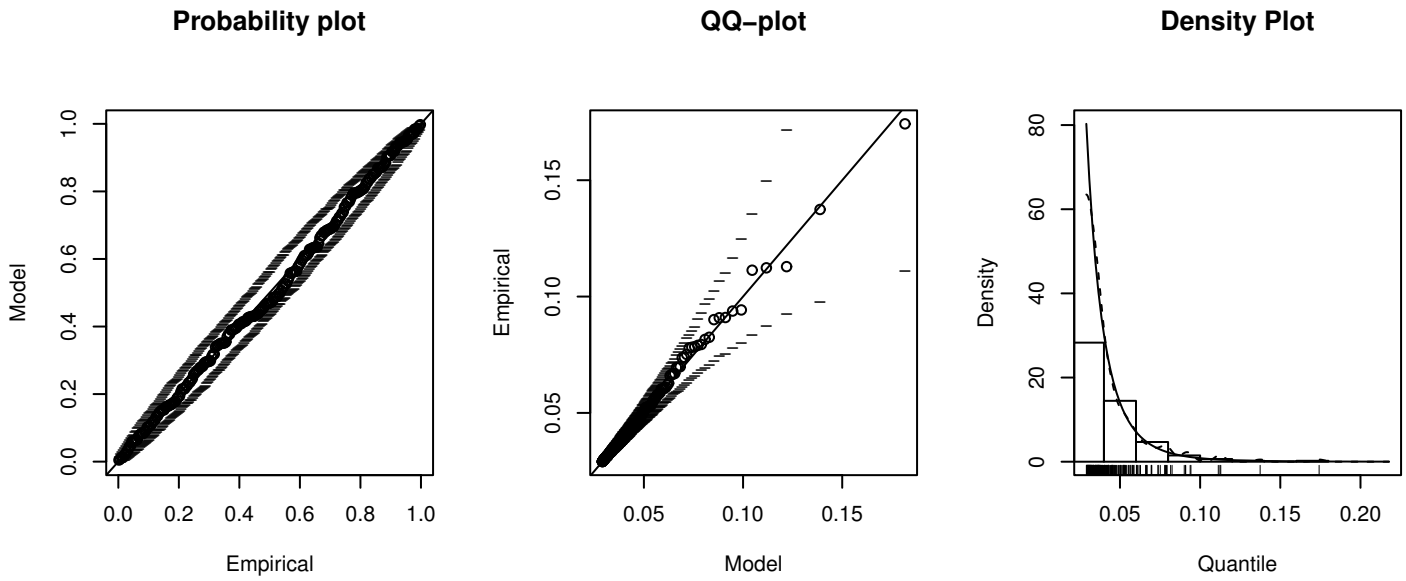


Figure 3.17: Diagnostic plots for positive returns modelled by a GPD with threshold $u = 0.029$.

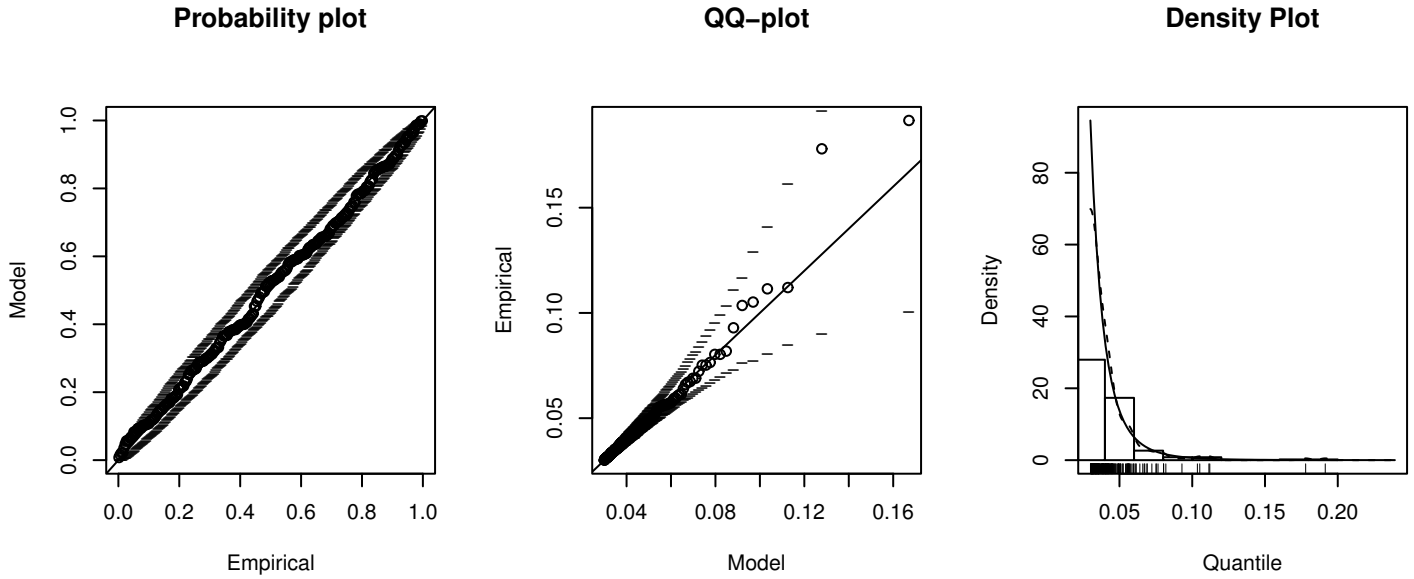


Figure 3.18: Diagnostic plots for negative returns modelled by a GPD with threshold $u = -0.03$.

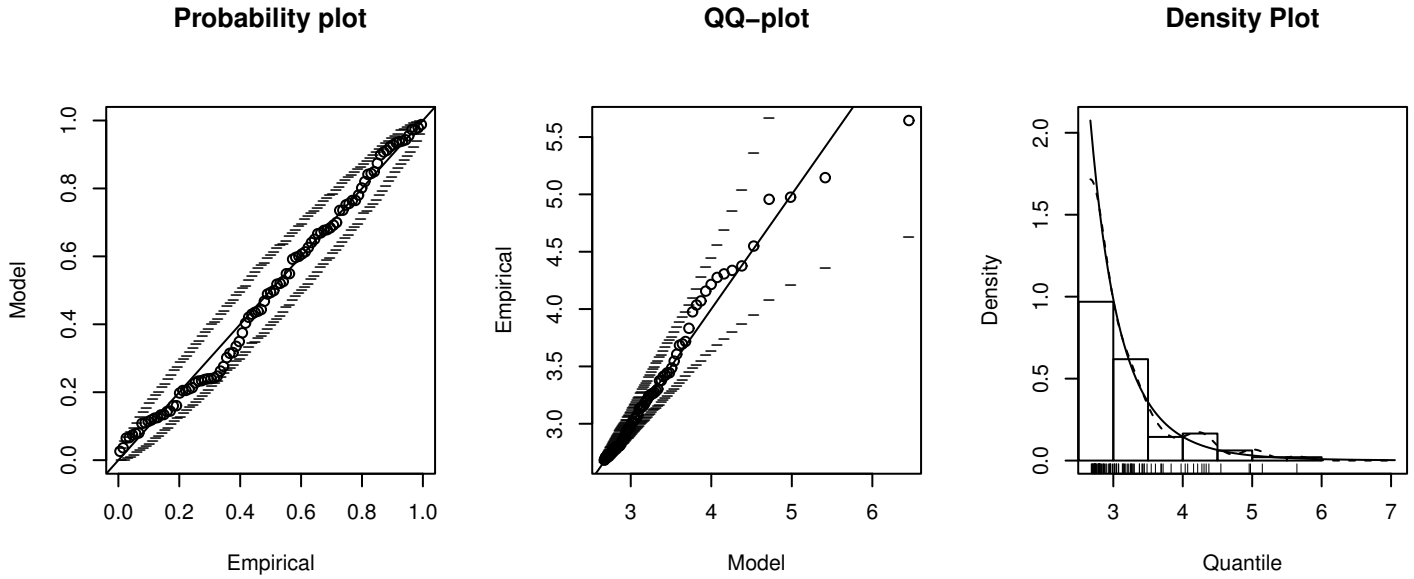


Figure 3.19: Diagnostic plots for volumes modelled by a GPD with threshold $u = 2.67$.

Graphs of empirical excess distribution are plotted for the three variables according to different threshold levels are in Figure 3.20. For all returns and volumes, the excess distribution plots follow the trace of the corresponding GPD very closely. This proves that the GPD models the return and volume exeedances quite well. The positive return exceedances seem to fit the GPD more suitably than the other

variables.

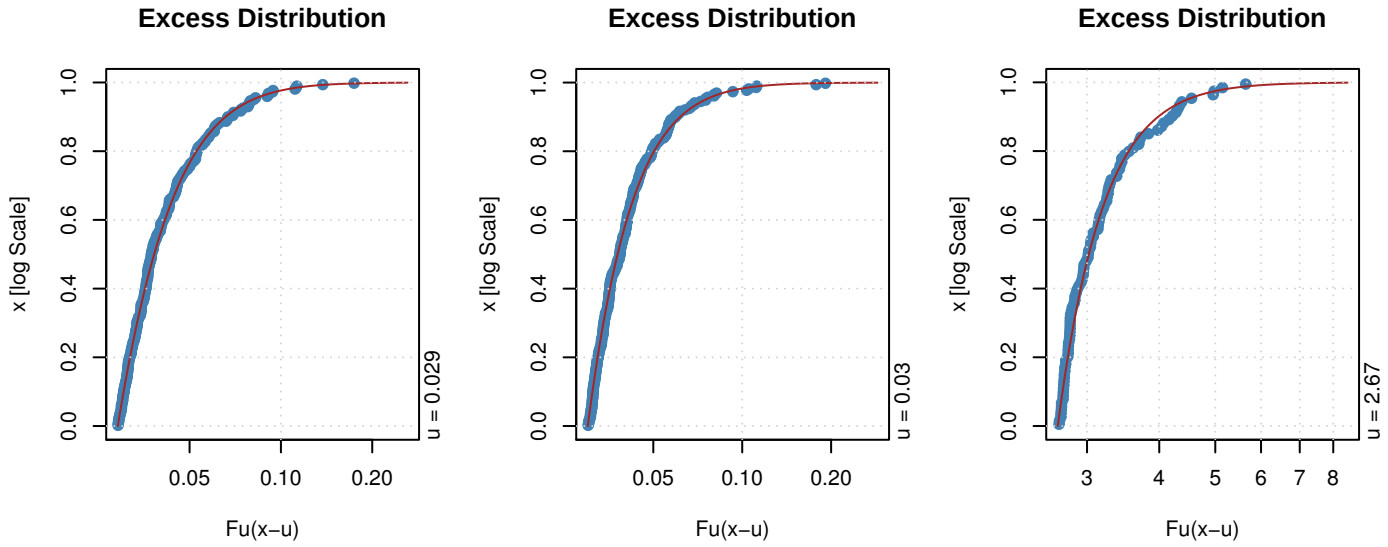


Figure 3.20: Excess distribution plots of positive and negative returns, and volumes, respectively.

Additional diagnostic plots are provided in Figures 3.21 to 3.23. Figure 3.21 displays a very good fit of the GPD model to the positive returns; the graph of the tail of the distribution shows that the extremes are approximately GPD, with majority of the points lying very close to the red line on the residual plot. The QQ plot against the exponential shows quantiles approximately above the exponential quantile, which implies a heavy tail. Similar results can be observed for the negative returns in Figure 3.22. The diagnostic plots for volume in Figure 3.23, however, display a slight deviation as the residuals are not as congested on the straight line and the QQ plot against the exponential distribution shows a few quantiles below the exponential quantile - indicating a lighter tail.

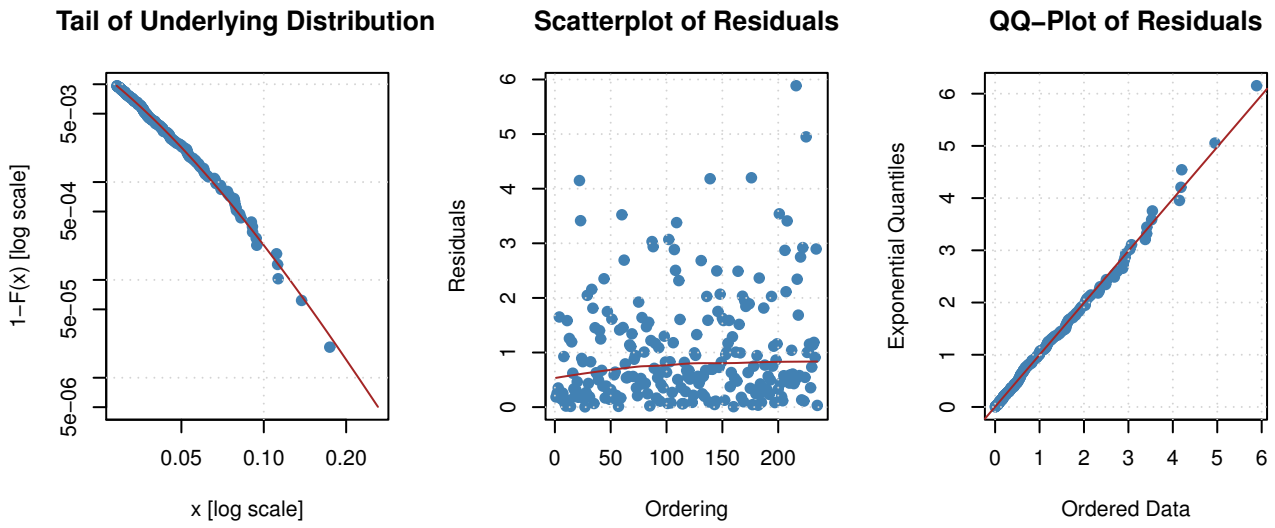


Figure 3.21: More diagnostic plots for positive returns modelled by a GPD with threshold $u = 0.029$.

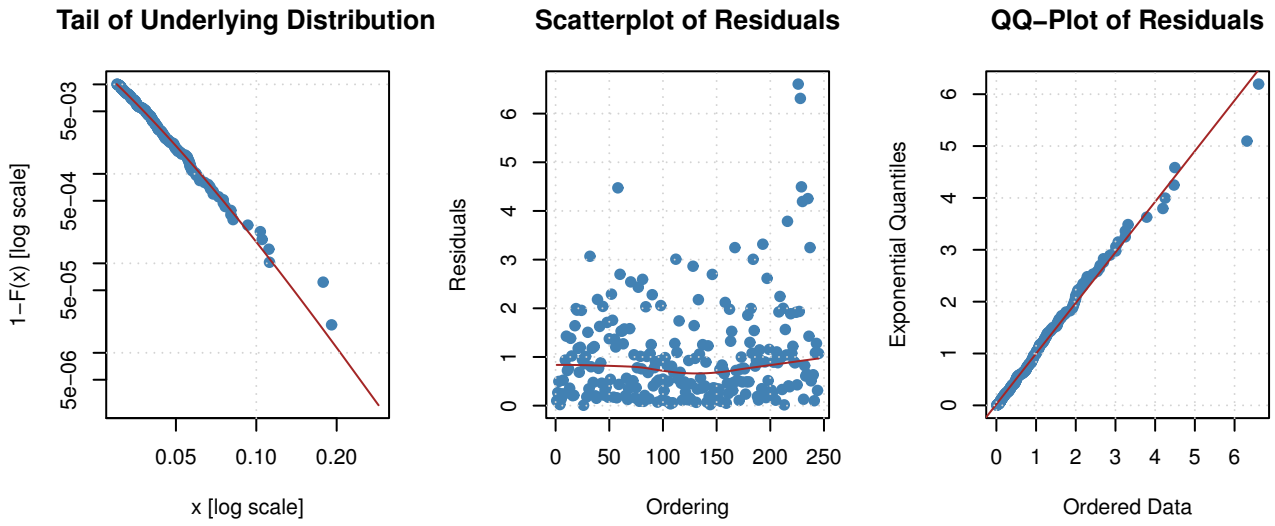


Figure 3.22: More diagnostic plots for negative returns modelled by a GPD with threshold $u = -0.03$.

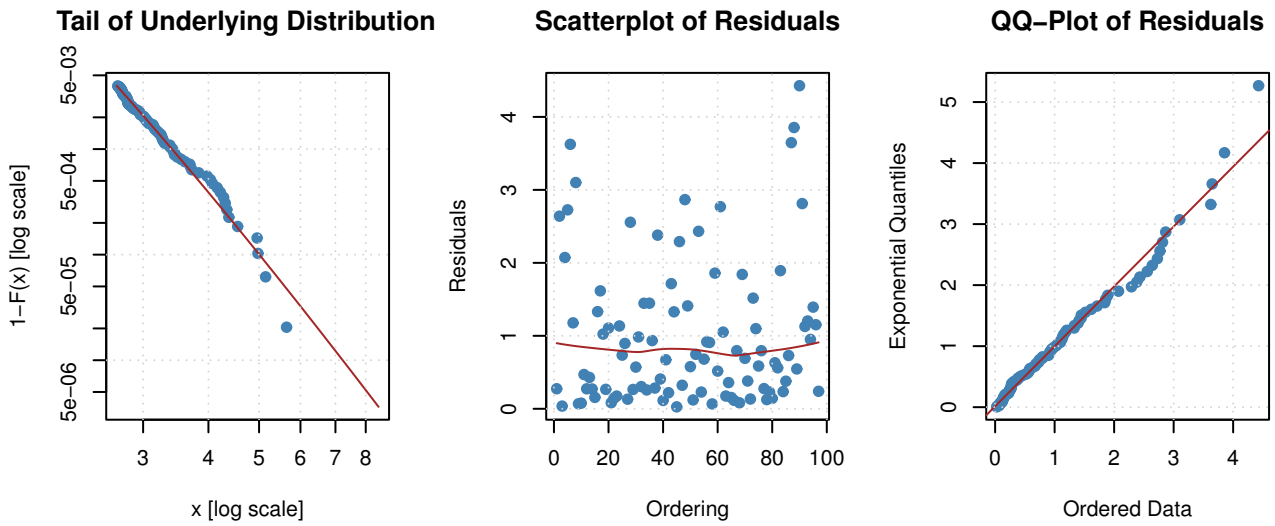


Figure 3.23: More diagnostic plots for volumes modelled by a GPD with threshold $u = 2.67$.

Return level plots for the selected models for positive returns and negative returns are plotted in Figure 3.24. Both plots indicate a good fit of the GPD models to the empirical data since the points appear to be close to the return level curve (the solid line), with very few deviations. The points on the return level plot for the positive returns seems to be more accurate as the points are very close to the line. The plots depict a concave shape since the estimated EVI parameter is a positive value for both variables.

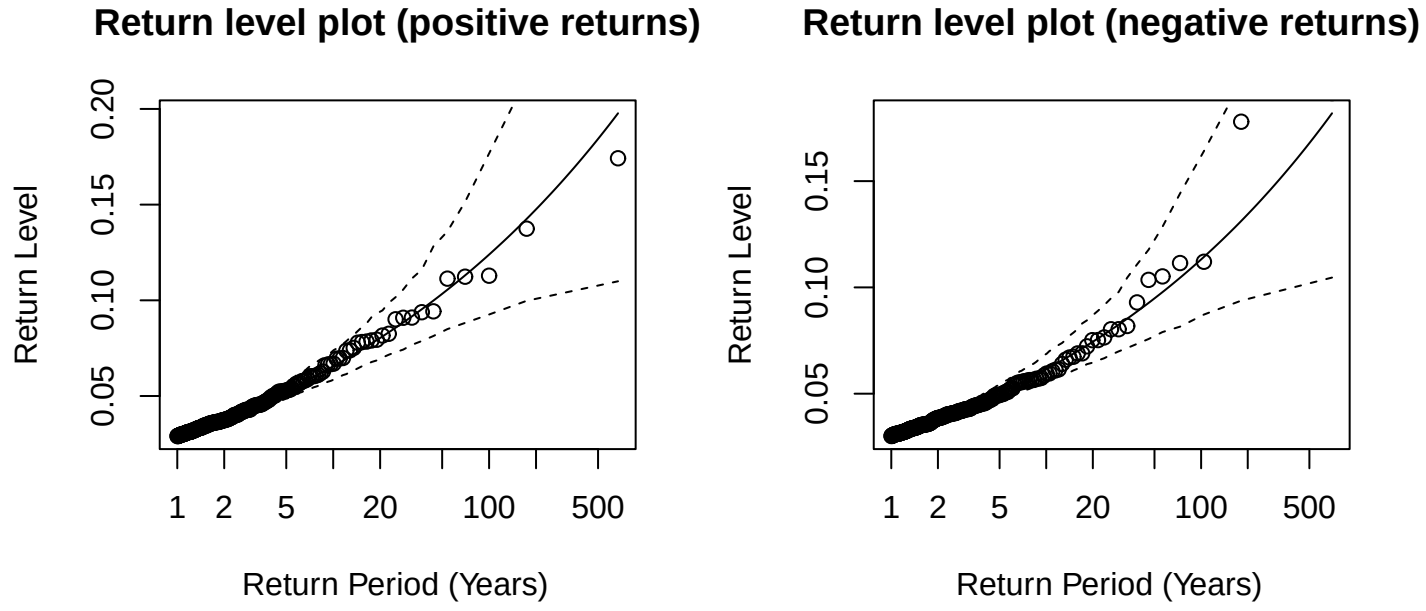


Figure 3.24: Return level plots for positive return and negative return GPD models with thresholds $u = 0.029$ (left) and $u = -0.03$ (right).

Profiled confidence intervals can be computed for both the scale and shape parameters from the fitted GPD models. These confidence intervals are generated using 'profile likelihood' - a useful procedure that can be applied to draw inference on some combination of unknown parameters. For the above GPD models, numerically assessing the profile likelihood of the shape parameter (γ) involves maximising the log-likelihood with respect to the scale parameter (β) by first fixing γ to a particular value. A range of other values for γ are then considered and maximisation executed for each. The profile log-likelihood for γ is then comprised of all of these maximised values of the log-likelihood, resulting in an approximation of confidence intervals at a particular level of significance (Coles et al., 2001). Similarly, the profile likelihood can be computed for the scale parameter.

Table 3.8 records the lower and upper confidence limits for the MLE paramters of selected GPD models for positive returns, negative returns, and trading volumes. The confidence interval for the shape parameter for volumes includes the value zero. Nonetheless, at the 95% level of confidence, it is evident that all the point estimates from Table 3.7 fall within the range of the profiled confidence intervals. Figures 3.25 to 3.27 display graphs of the profile likelihood functions for the shape and scale parameters for the selected GPD models for the different variables. In each figure, the area within the two horizontal and two vertical lines depicts the likelihood based confidence regions for the parameters.

Model	Parameter	Lower bound	MLE Point estimate	Upper bound
Positive returns				
$u = 0.029$	β	0.01	0.012	0.015
$u = 0.029$	γ	0.072	0.203	0.395
Negative returns				
$u = -0.03$	β	0.01	0.011	0.013
$u = -0.03$	γ	0.104	0.215	0.293
Trading volume				
$u = 2.67$	β	0.342	0.482	0.667
$u = 2.67$	γ	-0.089	0.143	0.206

Table 3.8: 95% Confidence intervals for the MLE parameters of hourly BTC.

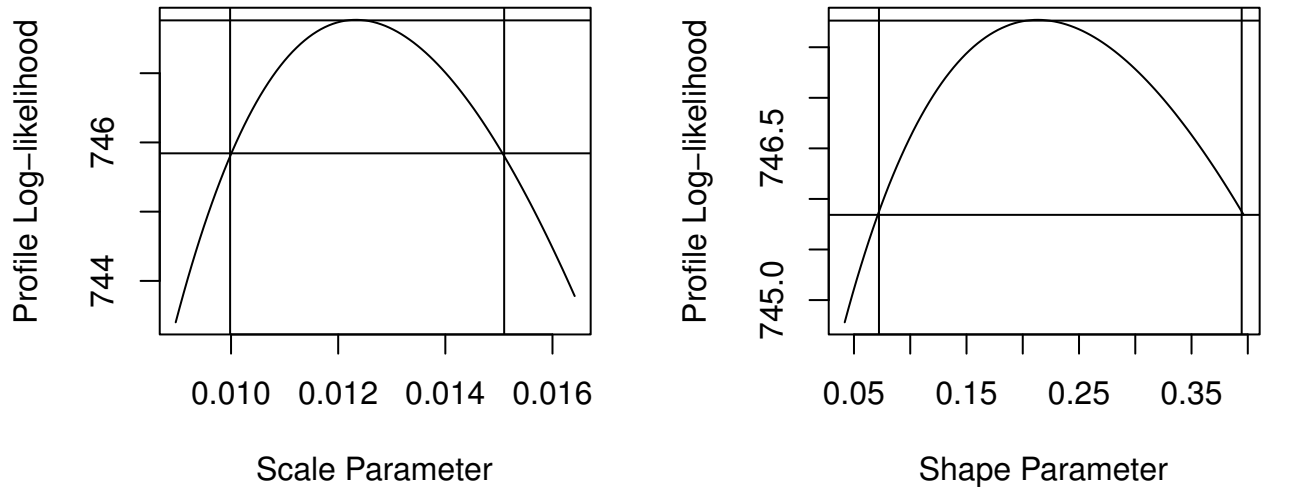


Figure 3.25: Plots of the profile likelihood functions and corresponding confidence intervals for parameters γ (right) and β (left) obtained using the MLE method (for positive returns).

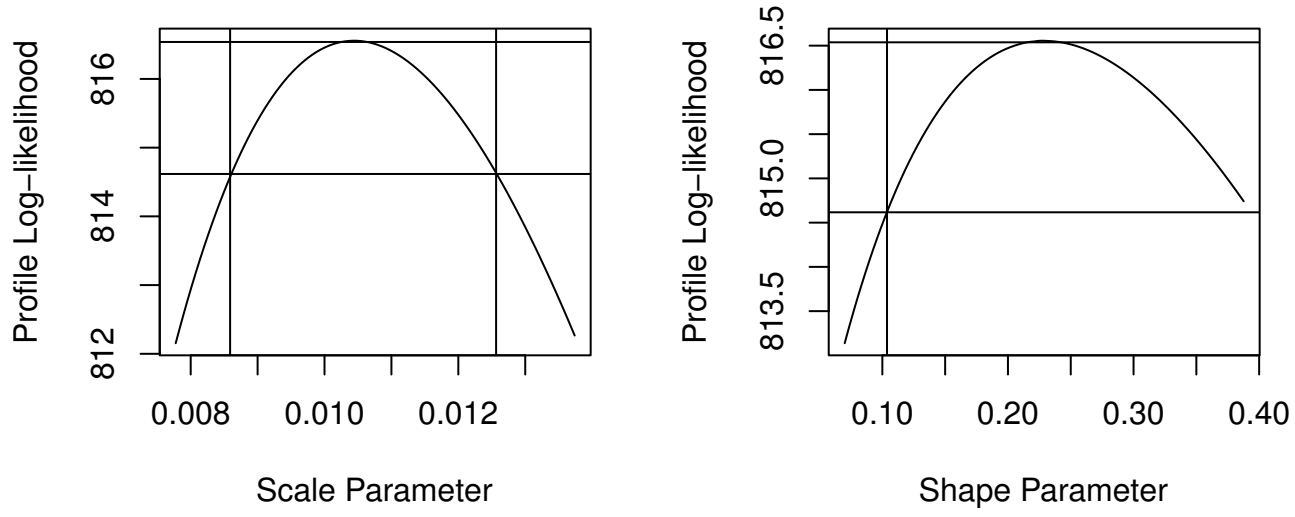


Figure 3.26: Plots of the profile likelihood functions and corresponding confidence intervals for parameters γ (right) and β (left) obtained using the MLE method (for negative returns).

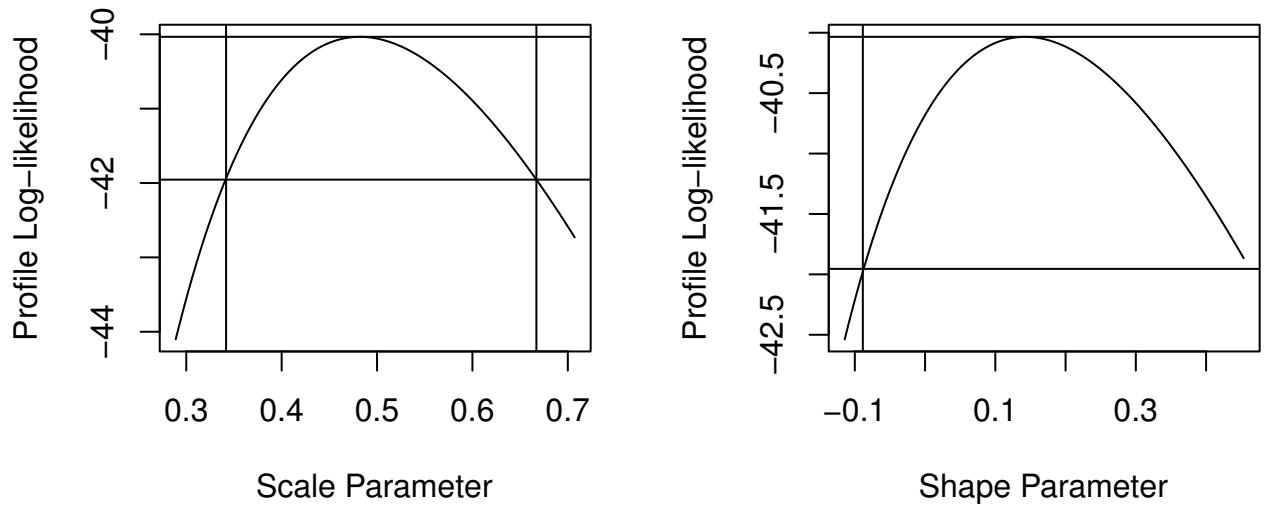


Figure 3.27: Plots of the profile likelihood functions and corresponding confidence intervals for parameters γ (right) and β (left) obtained using the MLE method (for trading volumes).

3.4 Estimating Tail Risk: VaR and ES

Table 3.9 records the estimates of VaR and ES of the negative returns at 99%, 98% and 95% quantile levels, according to their respective threshold values fitted in their GPD models. Recall, in this case, VaR measures the worst possible scenario on the market value of Bitcoin prices, given a certain level of confidence. For instance, for the left tail of the returns with a threshold $u = -0.05$, at the 99% level of confidence, VaR is estimated to be -0.0696819. This means that, given normal conditions in the Bitcoin market, one can expect that an hourly decrease in the price of this cryptocurrency would not exceed 6.968%. And in the extreme case that it does exceed this quantile, the actual loss made will be 5.668%.

In Table 3.10, for the GPD model with $u = -0.03$, the VaR = -0.04491356 at the 95% level of confidence suggests that for the lower 5% hourly returns, the hourly loss in the trading value of Bitcoin can be expected to be 3.499%. Therefore, for a given investment in Bitcoin, one can be 95% confident that, in the worst-case scenario, the loss in returns will not be greater than 3.499% during a one hour period. Similarly, at a 98% level of confidence, it is expected that the hourly loss will not exceed 2.493% in one trading day, given that it exceeds the estimated maximum possible loss (VaR).

Table 3.9: VaR and ES estimates for the first threshold selection method applied to hourly BTC/USD.

Quantile	Threshold	VaR estimate	ES estimate
99%	$u = -0.05$	0.0696819	0.05667889
99%	$u = -0.04$	0.04694447	0.03493279
99%	$u = -0.03$	0.02992652	0.01589365
99%	$u = -0.02$	0.00952886	0.0048738
99%	$u = -0.01$	0.20278227	0.34537825
99%	$u = -0.0$	0.00424861	0.00175139
99%	$u = -0.01975$	0.00997842	0.00305873
98%	$u = -0.05$	0.07538828	0.06481923
98%	$u = -0.04$	0.05133343	0.04221133
98%	$u = -0.03$	0.03703038	0.02493042
98%	$u = -0.02$	0.01726099	0.00452125
98%	$u = -0.01$	0.12274605	0.25143903
98%	$u = -0.0$	0.0041877	0.0018123
98%	$u = -0.01975$	0.01697747	0.00544559
95%	$u = -0.05$	0.08132645	0.07329023
95%	$u = -0.04$	0.05555862	0.04921828
95%	$u = -0.03$	0.04491355	0.03498541
95%	$u = -0.02$	0.02612953	0.01529711
95%	$u = -0.01$	0.02879256	0.14116498
95%	$u = -0.0$	0.00400115	0.00199885
95%	$u = -0.01975$	0.02500517	0.01519978

The results from the backtesting procedures essentially compare the losses predicted by the VaR models to those that actually occured over the testing or sample period. Ideally, the actual losses should not signifcantly exceed the expected losses, otherwise the models perform poorly in risk estimation.

Table 3.10 tabulates the tail risk estimates (VaR and ES) for the models fitted to the hourly BTC/USD data. The VaR estimates are backtested using the Kupiec LR test and Christoffersen test and the number of violations, test statistics and p-values are recorded. The ES estimates are backtested using the McNeil and Frey (2000) test and p-values are recorded. The packages in R do not have these tests built-in to cater for VaR and ES estimates that are calculated using GPD/GEVD models that are not integrated with GARCH models. See Appendix for detailed code on the backtesting procedures for VaR and ES estimates.

The Kupiec LR test results are given in Table 3.10. They generally show high p-values for all three GPD models at all levels of significance 1%, 2% and 5%. However, the model with threshold $u = -0.07$ displays the lowest p-values for the Kupiec and Christoffersen tests, meaning it shows poor performance and it is rejected at 2% and 5% levels of signifance. Based on the Kupiec test, the validity of all the other VaR models is confirmed since they exhibit high p-values, thus they produce accurate VaR estimates. The Christoffersen test, on the other hand, unfortunately rejects majority of the models.

The Kupiec test only requires a model to have its proportion of violations close to the tail probability in order to be accepted as a good estimator of risk. On the other hand, the independence test included in the Christoffersen test tests for clustering among the VaR violations and rejects models that exhibit clustered violations. This means, even if a model's proportion of violations is close to the level of significance, if the violations are clustered (i.e. they are not independently distributed), the model can be rejected. Thus, the Christoffersen test has a high rate of rejecting models that could in fact be accurate, making it less efficient than the Kupiec test. This explains why the Kupiec LR test is used more extensively than the Christoffersen test.

The ES backtesting results from the McNeil and Frey (2000) t- test indicate suitable ES estimates for all the models at all levels of significance. All the GPD models produce high p-values, with the model with $u = -0.03$ producing the highest p-value. The means that all the risk models are valid as they generate accurate ES estimates.

Table 3.10: VaR and ES estimates and backtesting results for the second threshold selection method applied to hourly BTC/USD.

Quantile	**Threshold**	**VaR**	**ES**	**Kupiec LR test**		**Christoffersen**	**McNeil&Frey**
				violations	**p-value**	**p-value**	**p-value**
99%	$u = -0.03$	0.02992651	0.01589365	249	0.817	0.002	0.8413
99%	$u = -0.04$	0.04694447	0.03493279	57	0.229	0.018	0.3918
99%	$u = -0.07$	0.1302061	0.11054469	11	0.007	0.049	0.1587
98%	$u = -0.03$	0.03703038	0.02493042	134	0.571	0.012	0.2826
98%	$u = -0.04$	0.05133343	0.04221133	43	0.137	0.005	0.1587
98%	$u = -0.07$	0.14069586	0.1233527	6	0.002	< 0.0001	0.1582
95%	$u = -0.03$	0.04491356	0.03498541	66	0.174	0.017	0.1587
95%	$u = -0.04$	0.05555862	0.04921828	35	0.037	0.005	0.1587
95%	$u = -0.07$	0.15268891	0.13799622	4	< 0.0001	< 0.0001	0.1587

Table 3.11 records VaR and ES estimates calculated at the 95% level using methods Historical simulation (a parametric approach) and RiskMetrics (an non-parametric approach) discussed in Sections 2.6.2.1 and 2.6.2.2 respectively. Under the Historical simulation method, VaR is computed by taking the 95%, 98% or 99% quantiles of the returns from the historical data, with no assumptions of a particular distribution.

While, the RiskMetrics method computes the VaR using the formula (upon estimating parameters mean μ and standard deviation σ),

$$VaR_p = \gamma + \sigma \cdot Q_{norm}(p).$$

ES is then computed using

$$ES_p = \gamma + \sigma \cdot ES_{norm}(p).$$

From Table 3.11 there is no clear pattern of which method produces bigger risk estimates. However, the existing literature does say that the normal distribution model underestimates VaR. Comparing the EVT VaR estimates from Table 3.10 and the Riskmetrics (normal) estimates Table 3.11, it is quite clear that the VaR estimates from the GPD models are larger than those of the Riskmetrics - especially for higher quantile levels (i.e. as we move further into the tail). This implies that the normal model underestimates the size of cryptocurrency losses and further confirms the inefficiency and lack of accuracy of these two methods compared to the EVT approach (Section 2.6.2.3) when calculating and managing risks in the crypto market.

Table 3.11: VaR and ES estimates using methods Historical simulation and RiskMetrics.

	Quantile	Historical simulation	Riskmetrics (Normal)
VaR	99%	0.01336546	0.01619055
	98%	0.02176592	0.02020487
	95%	0.02801351	0.02288111
ES	99%	0.0237911	0.02029287
	98%	0.03441449	0.02380944
	95%	0.04400851	0.02620793

3.5 Evaluating the Better Investment: Bitcoin or Gold

For this section, historical data for daily adjusted closing Bitcoin prices and Gold spot prices (GC=F COMEX Delayed Price in USD) obtained from Yahoo Finance for the period 04 January 2017 - 19 Oct 2023 are utilised. Figures 3.29 and 3.28 show the time series of daily Bitcoin and Gold prices for the 6 year period. The period 2017-2021 shows very moderate fluctuations in Bitcoin prices compared to the period 2021-2023 where very large increases in BTC prices can be seen. This is likely due to, among other things, the extreme effects of COVID-19. Towards the end of 2023, however, they appear to decrease considerably.

The Gold prices also increase substantially from 2021 onwards, but they remain at considerably high levels. Overall, the time series indicate the presence of trends and seasonality which are eliminated by adjusting the data before modelling the detrended log returns. The negative returns of these assets are modelled below.

Table 3.12 provides the summary statistics for both returns and it can be noted that the kurtosis for both variables is higher than 3, indicating a heavier tail than the normal distribution. However, the

Gold returns have a lighter tail as their skewness is positive and the kurtosis is only 4.76 . Gold returns have a smaller standard deviation than BTC, which could mean they exhibit less variation or volatility.

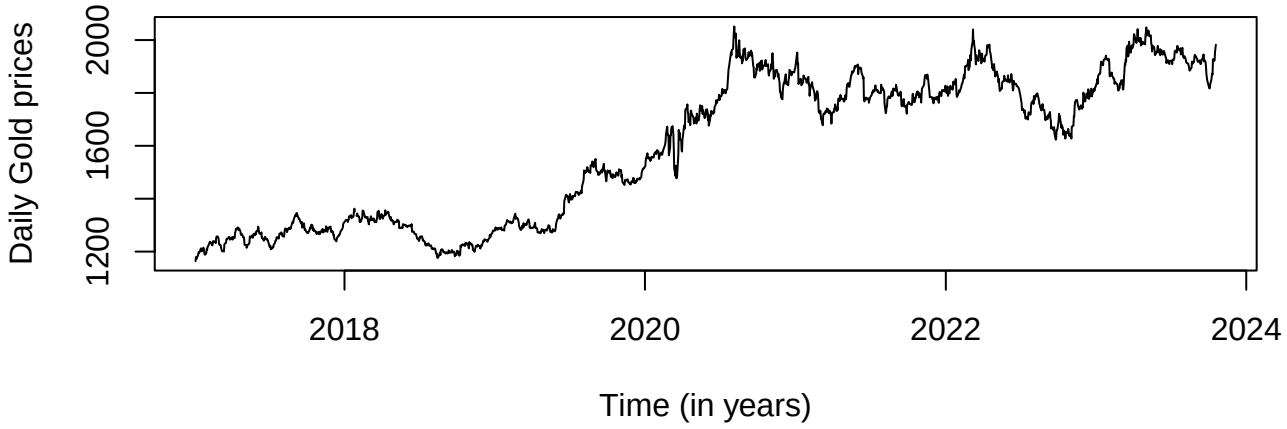


Figure 3.28: Time series of Gold prices (USD).

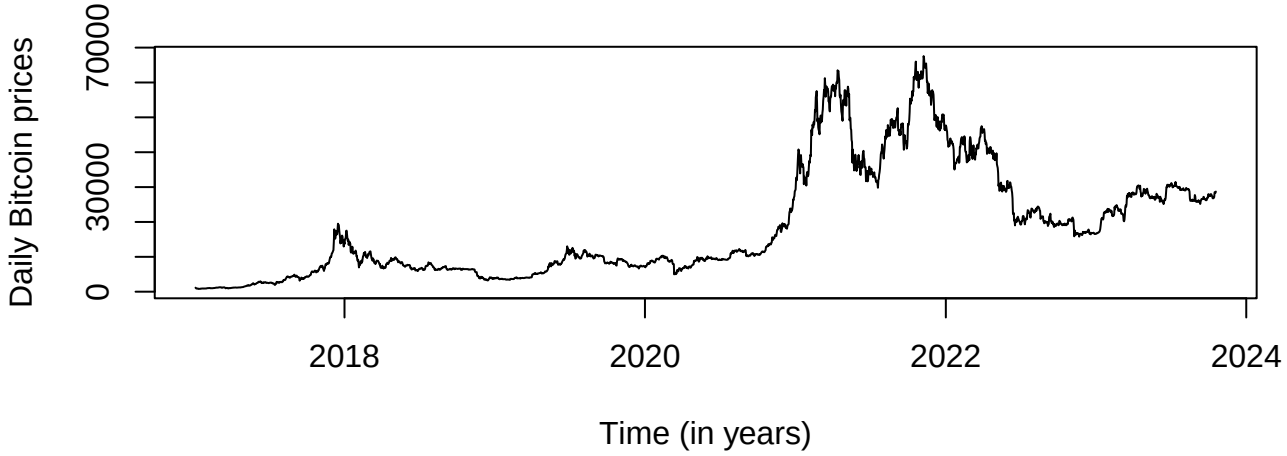
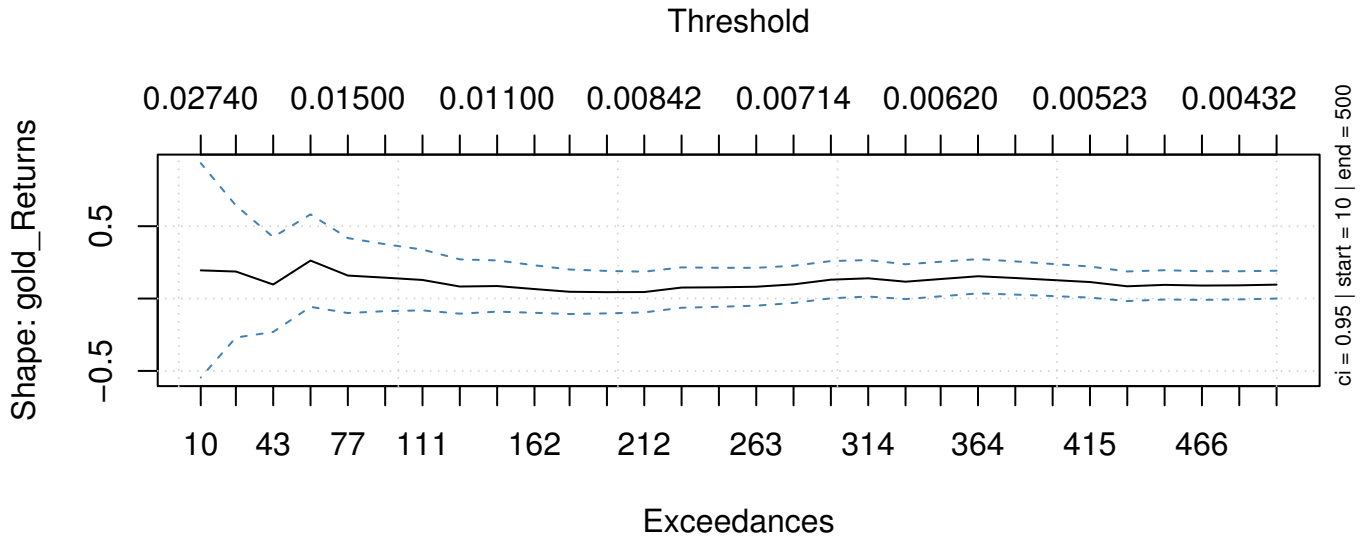


Figure 3.29: Time series of Bitcoin prices.

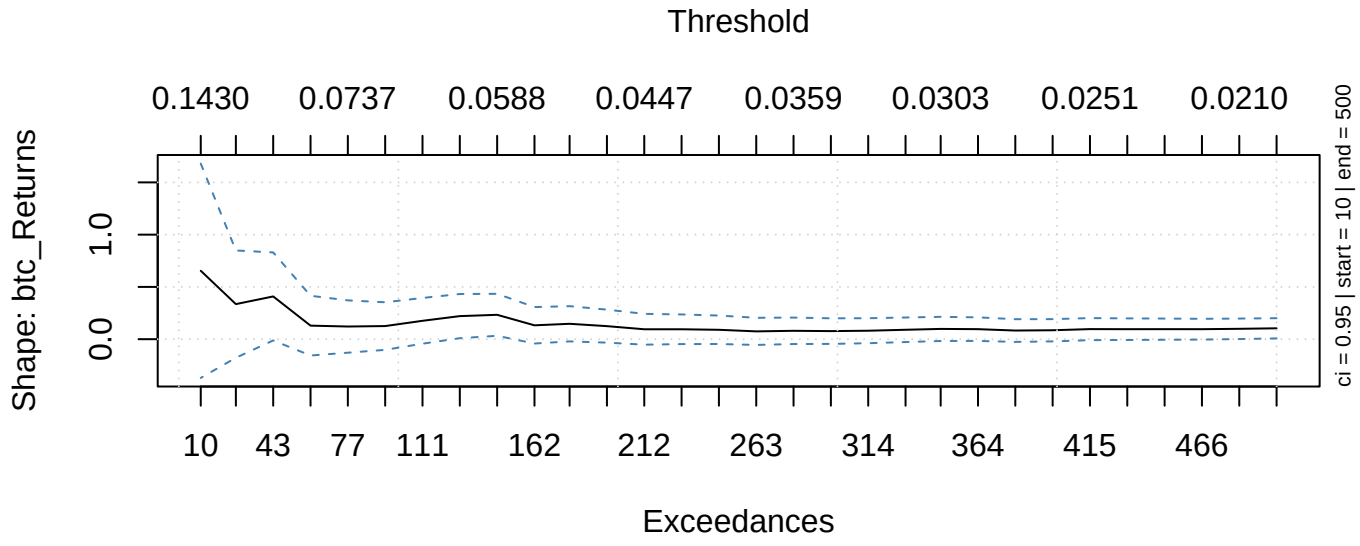
Table 3.12: Descriptive statistics for the daily returns of Bitcoin and Gold.

Statistics	Bitcoin returns	Gold returns
Observations	2479	1710
Minimum	-0.465	-0.058
Q1	-0.015	-0.005
Median	0.001	0.000
Mean	0.001	-0.000
Q3	0.018	0.004
Maximum	0.255	0.051
Skewness	-0.71	0.17
Kurtosis	11.15	4.76
Standard deviation	0.04	0.01
Variance	0.002	0.000
Range	0.69	0.11
IQR	0.033	0.009

Figure 3.30 (a) plots the EVI parameter for varying threshold values for gold returns and Figure 3.30 (b) plots those for daily Bitcoin returns. In both shape plots it can be seen that the shape parameter only stabilises for high threshold values. In Figure 3.30 (a), the threshold points $u = 0.017$, and $u = 0.02$ display a stability point for the shape parameter, while multiple threshold values demonstrate this for Bitcoin in Figure 3.30 (b) (such as the set of threshold points $u = \{0.055;\ 0.059;\ 0.087;\ 0.1;\ 0.12\}$). Mean excess plots are subsequently plotted to aid in threshold selection.



(a) Gold returns.



(b) Bitcoin returns.

Figure 3.30: Plots of the shape parameter against various thresholds for the negative returns of Gold (a) and Bitcoin (b).

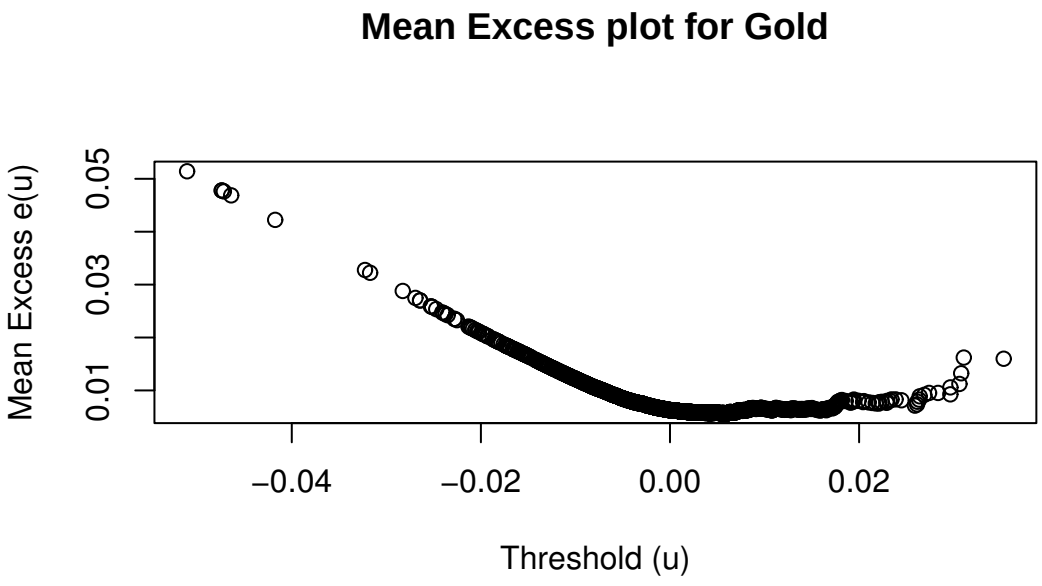


Figure 3.31: Mean excess plot for the negative returns of Gold.

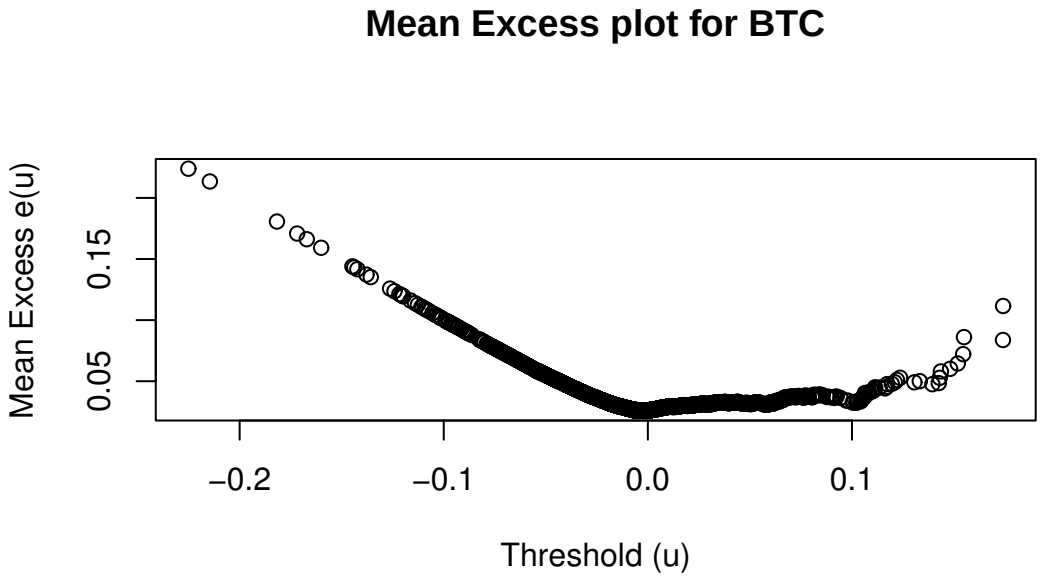


Figure 3.32: Mean excess plots for the negative returns Bitcoin.

The mean excess functions are plotted in Figure 3.32. The mean excess plot for BTC shows an upward linear pattern at the intervals with thresholds (u) approximately equal to 0 to 0.035, or somewhere between 0.06 and 0.08, or a threshold greater than 0.1. So, to select sufficiently high thresholds, the following thresholds will be used for Bitcoin returns 0.055, 0.075 , and 0.12. For the gold returns, thresholds 0.008, 0.017, and 0.027 are selected. This choice of thresholds is formulated using a combination of the observations made from both the mean excess plots and the shape plots. The GPD parameter estimates for these threshold levels are recorded in Tables 3.13 and 3.14.

Table 3.13: Parameter estimates for the negative returns of Gold.

u	N_u	p	$\hat{\gamma}$	$se(\hat{\gamma})$	$\hat{\beta}$	$se\left(\hat{\beta}\right)$
-0.008	237	0.139	< 0.00001	0.048	0.006	0.001
-0.017	55	0.032	< 0.00001	0.097	0.007	0.001
-0.027	11	0.006	< 0.00001	0.414	0.009	0.005

From Table 3.14, the shape parameters are much higher than those of the gold returns, with the model with threshold $u = -0.12$ showing the highest EVI value of 0.329. From Table 3.13 the shape parameters of the gold returns are estimated as a value very close to zero which confirms that the tails exhibit lighter tail behaviour.

Table 3.14: Parameter estimates for the negative returns of Bitcoin.

u	N_u	p	$\hat{\gamma}$	$se(\hat{\gamma})$	$\hat{\beta}$	$se\left(\hat{\beta}\right)$
-0.055	156	0.063	0.147	0.082	0.027	0.003
-0.075	76	0.031	0.138	0.104	0.032	0.005
-0.12	18	0.007	0.329	0.267	0.031	0.011

Figures 3.33 - 3.37 plot the model diagnostics to evaluate how well the fitted models describe gold returns and Bitcoin returns. These diagnostic plots correspond to the models with thresholds $u = -0.017$ and $u = -0.075$ for gold and Bitcoin respectively. From Figure 3.33, the GPD model shows a fairly good fit to the extreme gold returns, noting some departures from the 45-degree line in the pobability plot. The QQ plot shows that the underlying distribution of the returns can indeed be approximated by GPD as majority of the quantile points lie close to straight line. The fitted density plot also confirms a good fit. The same observation is made for the fitness of the GPD model to extreme BTC returns from Figure 3.34.

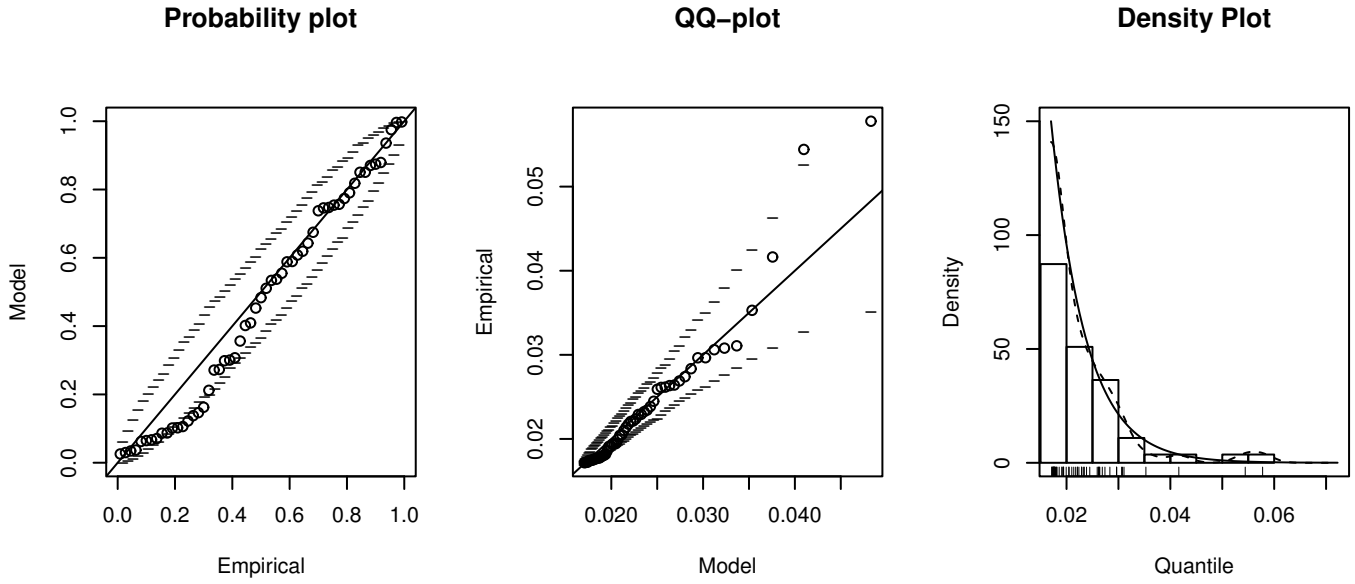


Figure 3.33: Diagnostic plots for Gold returns modelled with threshold $u = -0.017$.

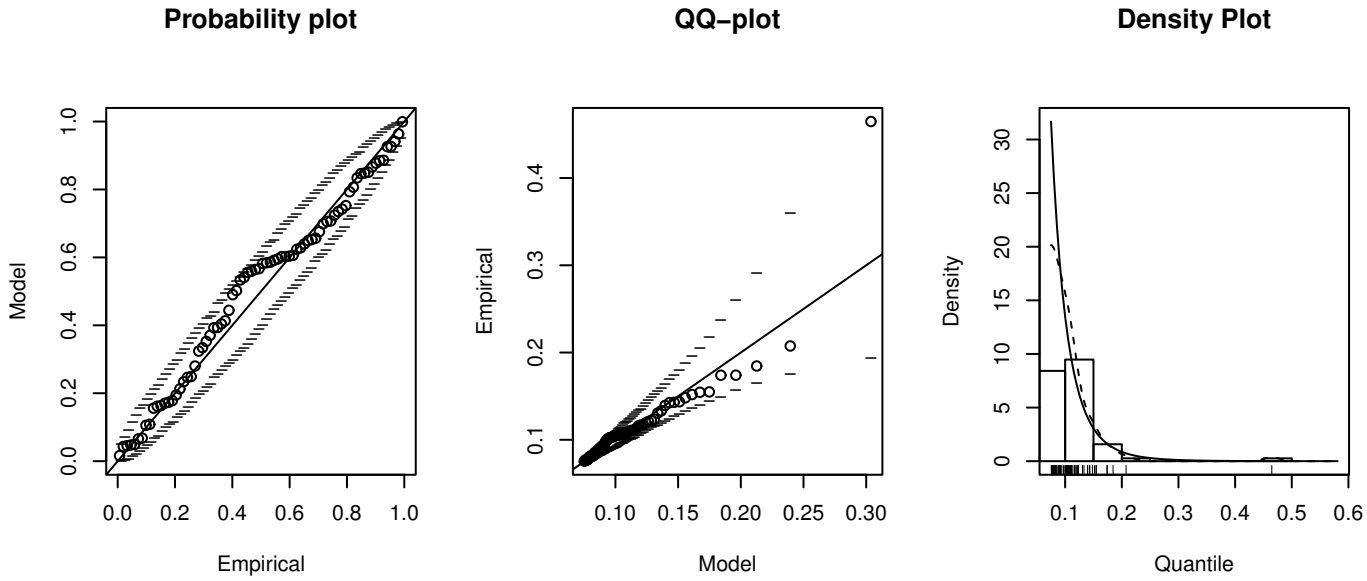


Figure 3.34: Diagnostic plots for Bitcoin returns modelled with threshold $u = -0.075$.

The empirical excess distribution plots for both Gold and BTC in Figure 3.35 indicate that fitted distributions model the negative returns very well since the pattern displayed by the points matches that of the GPDs very closely. The complimentary diagnostic plots in Figures 3.36 and 3.37 also suggest a good fit of the models. The graphed distribution tails indicate that the extreme returns in the tail

of the returns distribution are well described by the fitted GPD models (despite a few insignificant deviations).

Most of the points on the residual plots are surrounding the red line for the BTC returns, however quite a lot of the points deviate from the red line for the Gold returns. The QQ plot against the exponential displays a few quantiles above the exponential quantile - implying that the gold returns exhibit a tail that is heavier than the exponential distibution, and an even heavier tail for the Boitcoin returns in Figure 3.37.

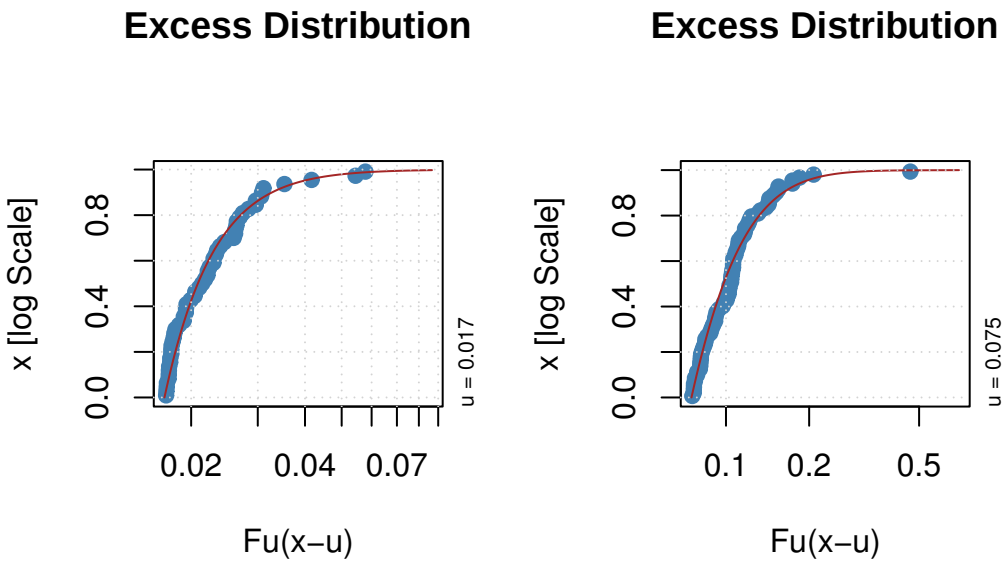


Figure 3.35: Excess distribution plots of Gold returns (left) and Bitcoin returns (right).

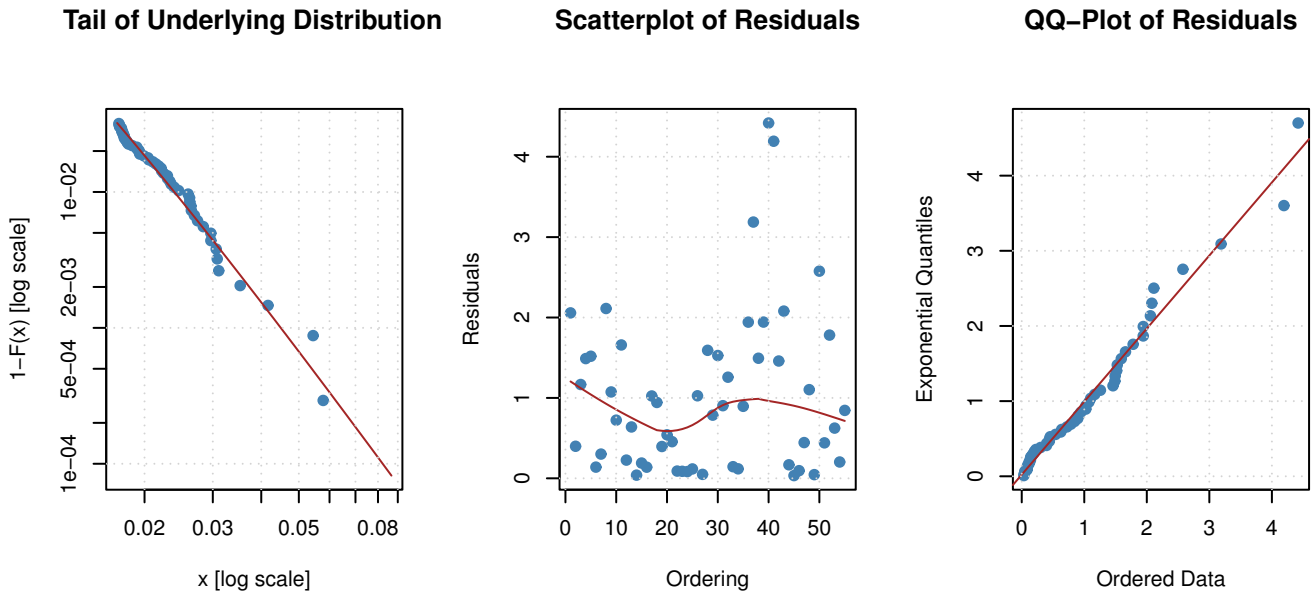


Figure 3.36: Additional diagnostic plots for Gold returns modelled with threshold $u = -0.017$.

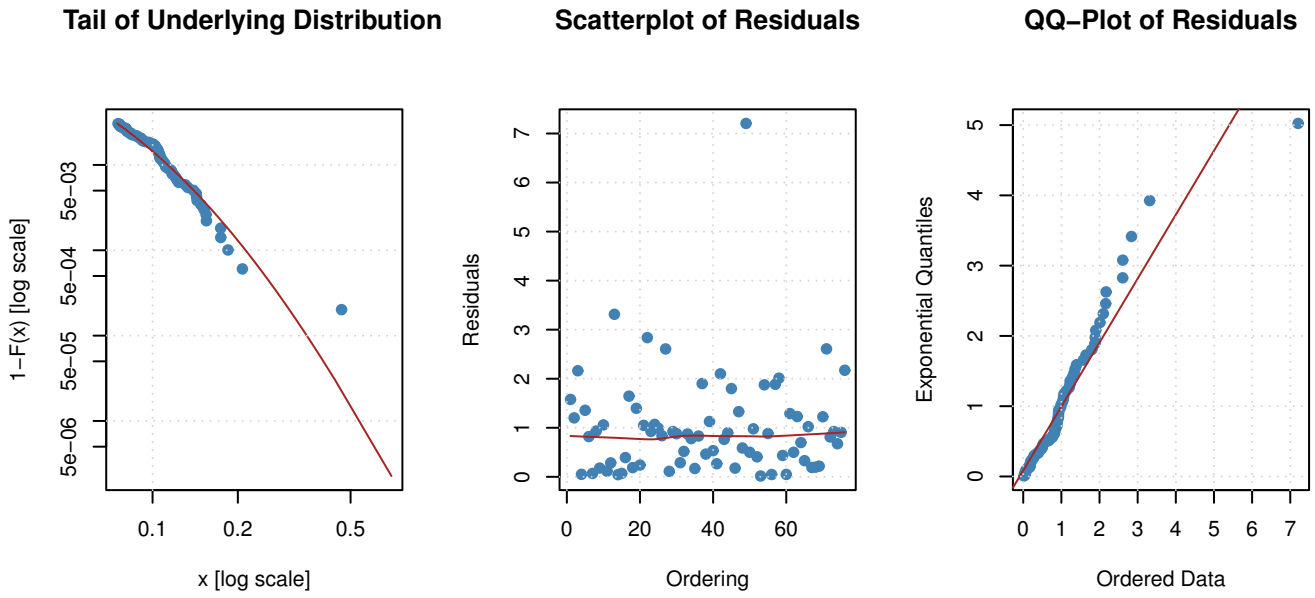


Figure 3.37: Additional diagnostic plots for Bitcoin returns modelled with threshold $u = -0.075$.

In Tables 3.15 and 3.16 the risk estimates as well as backtesting results are recorded for gold returns and Bitcoin returns respectively, at various quantile levels. In Table 3.15, in the worst case scenario, under the model with threshold $u = -0.008$, an investor in gold can expect (with 98% confidence) that

the market value of gold will not decrease by more than 1.974%. For the same model, they can expect very similar percentage in loss at 95% and 98% confidence. For the model with the largest $u = -0.027$, an investor can expect the market value to lose less than 7.233% in a day, with 99% confidence.

In Table 3.16, in extreme conditions in the Bitcoin market, traders can expect tomorrow's loss not to exceed 14.17675%, at the 99% confidence level (modelled at the highest threshold $u = -0.12$). But in the rare event that this VaR is exceeded, they can expect a total loss of 10.494% in returns. The smallest VaR estimated for Bitcoin is 3.013%, with the largest being 15.428%.

Notably, the Bitcoin returns produce higher VaR estimates than the gold returns, at all quantile levels 95%, 98% and 99%, making Bitcoin the riskier asset since it produces higher Value-at-risk. This means that for an individual that invests in both Gold and BTC, the prospects of extreme risk (i.e. losses) are greater in BTC compared to Gold. The VaR estimates for gold returns are very close to one another for each respective confidence level. The backtesting results with p-values that exceed the significance levels imply model adequacy in risk estimation. Those with significantly low p-values indicate poor performance in risk estimation.

Majority of the models for Gold generate satisfactory performance in estimating VaR, with the model with $u = -0.027$ showing very poor performance as its VaR estimations are rejected at all levels of significance. The risk models for BTC generally show efficiency in risk estimation as they exhibit high p-values. For the ES estimates, the McNeil and Frey (2000) test shows that none of the models are rejected at any level of significance as the p-values are significantly high. This observation applies to both Gold and BTC returns. Thus, the validity of all the models in the estimation of ES is confirmed.

Table 3.15: Tail risk estimates (VaR and ES) and backtesting results applied to daily Gold returns at different quantile levels.

Quantile	**Threshold**	**VaR**	**ES**	**Kupiec LR test**		**Christoffersen**	**McNeil&Frey**
				violations	**p-value**	**p-value**	**p-value**
99%	$u = -0.008$	0.01979683	0.01379683	32	0.661	0.042	0.2828
99%	$u = -0.017$	0.04098806	0.03398806	3	0.15	0.004	0.1587
99%	$u = -0.027$	0.07232673	0.06332673	0	< 0.0001	< 0.0001	0.1587
98%	$u = -0.008$	0.01973591	0.01373592	32	0.872	0.02	0.4993
98%	$u = -0.017$	0.04091699	0.03391699	3	0.093	0.007	0.1587
98%	$u = -0.027$	0.07223536	0.06323536	0	< 0.0001	< 0.0001	0.1527
95%	$u = -0.008$	0.01954937	0.01354937	33	0.383	0.02	0.8413
95%	$u = -0.017$	0.04069936	0.03369936	3	0.014	< 0.0001	0.1587
95%	$u = -0.027$	0.07195554	0.06295554	0	< 0.0001	< 0.0001	0.1587

Table 3.16: Tail risk estimates (VaR and ES) and backtesting results applied to daily BTC returns at different quantile levels.

Quantile	Threshold	VaR	ES	Kupiec LR test		Christoffersen	McNeil&Frey
				violations	p-value	p-value	p-value
99%	$u = -0.055$	0.05428533	0.04907219	158	0.004	< 0.001	0.1348
99%	$u = -0.075$	0.07915548	0.05166312	67	0.258	0.019	0.1587
99%	$u = -0.12$	0.14176753	0.10494283	12	0.099	0.005	0.1582
98%	$u = -0.055$	0.0301311	0.00673824	348	< 0.0001	< 0.0001	0.1587
98%	$u = -0.075$	0.06092044	0.02154344	131	0.113	0.004	0.1587
98%	$u = -0.12$	0.1542774	0.1416263	7	0.028	0.009	0.1587
95%	$u = -0.055$	0.06564714	0.0346928	106	< 0.001	< 0.0001	0.1588
95%	$u = -0.075$	0.09013603	0.05543623	51	0.635	0.0198	0.4964
95%	$u = -0.12$	0.08651274	0.07158315	55	0.691	0.02	0.1587

3.6 Assessing Model Performance

Table 3.17 records some results on how well the GPD models fit the extremes in the tails of positive returns, negative returns and volumes. The one-sample K-S test is implemented and AIC values from the fitted models are recorded. Evidently, all the models cannot be rejected at all levels of significance since they all display significantly high p-values. This means that they all fit the extreme observations very well, i.e. the estimated asymptotic distributions of the extremes display a good fit with the empirical distributions of the tails of the returns and volumes.

Notably, the model with the lowest K-S test statistic and highest p-value shows the best fit compared to the others. These can be identified as the hourly BTC/USD GPD models with thresholds $u = 0.029$, $u = -0.04$ and $u = 2.18$ for positive returns, negative returns and volumes respectively. It can be concluded that all the models for BTC/USD considered in the table can indeed be used to forecast potential Bitcoin returns and trading volumes when the market is faced with extreme conditions. For the Gold returns, the model with $u = -0.008$ seems to have the highest p-value and lowest test statistic compared to the others, making it the best fit for negative gold returns.

The model with the lowest AIC value is considered optimal since the smaller the value, the better the fit of the model. From Table 3.17, these are the hourly BTC/USD GPD models with thresholds $u = 0.705$, $u = -0.07$ and $u = 3.3$ for positive returns, negative returns and volumes respectively. These are actually the GPD models with the highest threshold values - which is not suprising since (as discussed in Section 2.4.1) fitting a GPD model to the tail requires the selection of a sufficiently high threshold in order to ensure that not too many data observations are included in the estimation process and biasedness is minimal. This means that the AIC value is influenced by the number of exceedances. The same observation can be concluded for the daily Gold and Bitcoin returns.

Table 3.17: Kolmogorov-Smirnov Test results (test statistics and p-values) and AIC values for assessing the goodness of fit of the GPD models on the tails of positive returns, negative returns and volumes.

	K-S test		**AIC**
Hourly BTC/USD GPD Models	**Test statistic**	**p-value**	
$u = 0.029$	0.0304323	0.9815	-1491.527
$u = 0.036$	0.0653	0.5938	-854.269
$u = 0.075$	0.12074	0.9276	-98.407
$u = -0.03$	0.042842	0.7595	-1629.013
$u = -0.04$	0.051751	0.9345	-695.013
$u = -0.07$	0.15535	0.8385	-64.153
$u = 2.18$	0.027103	0.9902	171.7
$u = 2.67$	0.082573	0.4968	84.068
$u = 3.3$	0.10183	0.9053	39.709
Daily Gold GPD models			
$u = -0.008$	0.049188	0.615	-1927.105
$u = -0.017$	0.15388	0.1329	-437.345
$u = -0.027$	0.27257	0.3256	-77.489
Daily BTC GPD models			
$u = -0.055$	0.0666	0.4941	-712.414
$u = -0.075$	0.1079	0.3165	-348.395
$u = -0.12$	0.1558	0.7181	-72.358

Chapter 4

Conclusion

In this thesis, the use of Extreme value theory in modelling cryptocurrency returns is explored for periods before the Covid-19 pandemic as well as after. The focus was on high-frequency (hourly and daily) data for the currency pair Bitcoin against the US Dollar. Data on Gold prices (in USD) was also included in the analysis for a comparison of financial risk.

An extreme value analysis was conducted on the extreme (tail) returns of Bitcoin at high-frequency. As was anticipated, they display significant deviations from the normality assumption, owing to their leptokurtic (heavy-tail) behaviour. Upon studying and comparing the BM and POT methods, the Peaks-over-Threshold approach was employed to identify and describe extremes and model them using the GPD. Several methods of selecting appropriately high thresholds were implemented on the data to analyse both positive and negative returns. The GPD models provide a good fit to the tails of the distribution of returns of Bitcoin, as was confirmed by the diagnostic plots and tests for 'goodness of fit'.

The underexplored relationship between returns and trading volume at the extreme tails was examined using the extreme dependence and correlation modelled with bivariate EVT. The results indicate that the correlation between positive returns and trading volume decreases from a strong positive correlation for small thresholds, to a weak positive correlation in the extreme right tail where extreme events in the Bitcoin occur or extreme market condition are observed. The same observation was made for negative returns in the left tail. This implies that positive (or negative) returns and trading volumes might not be strongly correlated at times when the market experiences booms or troughs. This finding is in line with the notion of cryptocurrencies behaving like assets despite inheriting financial traits far different from traditional financial assets.

The application was extended to include daily BTC-USD data that span the same time period but for prices from a different cryptocurrency exchange platform. The significance of this particular analysis is to check the robustness of the EVT models. Similar outcomes were observed – confirming, to a certain degree, the robustness of the models. The Bitcoin market, like most financial markets, exhibits

high volatility and general instability, thus extreme price movements are prone to occur.

The results from this dynamic tail relationship analysis can assist traders (which include ordinary citizens) during extreme market conditions. Valuable information about the market is conveyed by the dependence structure and can aid in understanding how the volumes of transactions contribute towards a change in Bitcoin prices and vice versa, and how losses can then be minimised and returns (i.e. profits) increased. The weak positive correlation between extreme returns and volumes implies that trading strategies reliant on speculation are neither optimal nor practical during extreme events, as they can result in trade misinterpretation among the market participants. As substantiated by Chan et al. (2022), this can result in the market becoming relatively illiquid, which then induces extreme price fluctuations.

EVT models provide a methodology that aids in understanding the extreme gains in the best case scenario and the losses in the worst case scenario, as well as the frequency thereof. The implications of extreme price changes include extreme losses. The prediction of extreme losses plays a significant role in financial risk management. The risk models estimate VaR and ES using the POT method. The accuracy of these risk estimates is evaluated using various backtesting procedures. Majority of the models perform well in risk assessing financial risk.

The inclusion of an application to South African data would have been an interesting study for this research, however due to data limitations the study was restricted to the USD. The USD currency dominates as the most universal and popular in the crypto industry and for many crypto exchanges this is the sole currency used for Bitcoin trading activities. The South African Rand (ZAR), on the other hand, is not as popular since it only showed significant presence in the Bitcoin market years after its launch in 2009. For this reason, majority of the trading platforms either have very limited data on the BTC-ZAR pair or no data at all.

An investigation was conducted to determine the more risky asset to invest in. The negative return observations of Gold and Bitcoin exhibit heavy-tail properties. The GPD models fitted to the distribution of the losses of these assets perform well. Bitcoin returns produced higher VaR estimates compared to the Gold returns. This suggests that Bitcoin has higher prospects of extreme risk (losses) compared to Gold, especially during turbulent periods. The conclusion drawn is that Bitcoin is the riskier asset. Although this finding seems to contradict the idea of Bitcoin being perceived as the new Gold, it does not necessarily conclude that Bitcoin is not the new 'digital Gold' as the results are only based on the analysis of the markets during extreme or turbulent conditions. Overall, the backtesting results indicate that the risk models perform well in risk estimation, thus confirming the validity of the models. Therefore, the usefulness of these EVT models cannot be underestimated in predicting potential extreme losses as large unexpected losses can bear heavy consequences for investors and financial institutions.

References

Abad, P., Benito, S., and López, C. (2014). A comprehensive review of Value at Risk methodologies. *The Spanish Review of Financial Economics*, 12(1):15–32.

Ali, N., Zaimi, N., and Ali, N. M. (2020). Statistical modelling of Malaysia trading gold price using extreme value theory approach. *Advances in Mathematics: Scientific Journal*, 10(1):1857–8365.

Andreeva, V. O., Tinykovb, S. E., Ovchinnikovaa, O. P., and Parahinc, G. P. (2012). Extreme value theory and peaks over threshold model in the Russian Stock Market. *Journal of Siberian Federal University. Engineering Technology*, 1(5):111–121.

Balcilar, M., Bouri, E., Gupta, R., and Roubaud, D. (2017). Can volume predict Bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, 64:74–81.

Balduzzi, P., Kallal, H., and Longin, F. (1996). Minimal returns and the breakdown of the price-volume relation. *Economics Letters*, 50(2):265–269.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.

Bee, M., Dupuis, D. J., and Trapin, L. (2016). Realising the extremes: Estimation of tail-risk measures from a high-frequency perspective. *Journal of Empirical Finance*, 36:86–99.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley and Sons, Chichester.

Chan, S., Chu, J., Zhang, Y., and Nadarajah, S. (2022). An extreme value analysis of the tail relationships between returns and volumes for high frequency cryptocurrencies. *Research in International Business and Finance*, 59:101541.

Cheah, E.-T. and Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics letters*, 130:32–36.

Chen, G. M., Firth, M., and Rui, O. M. (2001). The dynamic relation between stock returns, trading volume and volatility. *Financial Review*, 36(3):153–174.

Chikobvu, D. and Jakata, O. (2020). Analysing extreme risk in the South African financial index (J580) using the generalised extreme value distribution. *Statistics, Optimization & Information Computing*, 8(4):915–933.

Chinhamu, K., Huang, C. K., and Chikobvu, D. (2017). Evaluating risk in precious metal prices with generalised lambda, generalised pareto and generalised extreme value distributions. *South African Statistical Journal*, 51(2):159 – 182.

Chinhamu, K., Huang, C.-K., Huang, C.-S., Chikobvu, D., et al. (2015a). Extreme risk, value-at-risk and expected shortfall in the gold market. *International Business & Economics Research Journal (IBER)*, 14(1):107–122.

Chinhamu, K., Huang, C.-K., Huang, C.-S., and Hammujuddy, J. (2015b). Empirical analyses of extreme value models for the South African mining index. *South African Journal of Economics*, 83(1):41–55.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

CoinMarketCap (2022). Cryptocurrency market capitalisations.

CoinMarketCap (2023). Cryptocurrency market capitalisations.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer, London.

Corlu, C. G. and Corlu, A. (2015). Modelling exchange rate returns: which flexible distribution to use? *Quantitative Finance*, 15(11):1851–1864.

Costinot, A., Roncalli, T., and Teiletche, J. (2000). Revisiting the dependence between financial markets with copulas. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1032535`, accessed February 2023.

Cryptocompare (2023). Bitcoin historical data.

De Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*, volume 21. Springer, London.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431.

Dicks, A., Conradie, W. J., and De Wet, T. (2014). Value at Risk using GARCH volatility models augmented with Extreme Value Theory. *Studies in Economics and Econometrics*, 38(3):1–18.

Do, H. X., Brooks, R., Treepongkaruna, S., and Wu, E. (2014). How does trading volume affect financial return distributions? *International Review of Financial Analysis*, 35:190–206.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, New York.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1992). Stock prices and volume. *The Review of Financial Studies*, 5(2):199–242.

Gilli, M. and Këllezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2):207–228.

Gkillas, K. and Katsiampa, P. (2018). An application of extreme value theory to cryptocurrencies. *Economics Letters*, 164:109–111.

Gnedenko, B. V. (1948). On a local limit theorem of the theory of probability. *Uspheki Mat. Nauk*, 3(25):187–194.

Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University press, New York.

Gunduz, L. and Hatemi-J, A. (2005). Stock price and volume relation in emerging markets. *Emerging Markets Finance and Trade*, 41(1):29–44.

Haugh, M. (2016). An introduction to copulas. *IEOR E4602: Quantitative risk management. Lecture notes. Columbia University*.

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3(5):1163 – 1174.

Hussain, S. I., Masseran, N., Ruza, N., and Safari, M. A. M. (2021). Predicting Extreme Returns of Bitcoin: Extreme Value Theory Approach. *Journal of Physics: Conference Series*, 1988(1):012091.

Islam, M. T. and Das, K. P. (2021). Predicting Bitcoin return using extreme value theory. *American Journal of Mathematical and Management Sciences*, 40(2):177–187.

Jansen, D. W. and De Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The review of economics and statistics*, pages 18–24.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics letters*, 158:3–6.

Katsiampa, P., Gkillas, K., and Longin, F. (2018). Cryptocurrency market activity during extremely volatile periods. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3220781`, accessed August 2022.

Khan, G. R., Abdulrahman, A. T., Alamri, O., Iqbal, Z., and Ahmad, M. (2021). Risk analysis of Gold prices in Pakistan using Extreme Value Theory. *Mathematical Problems in Engineering*, 2021:1–18.

Kotz, S. and Nadarajah, S. (2000). *Extreme value distributions: theory and applications*. World Scientific, London.

Kupiec, P. H. et al. (1995). *Techniques for verifying the accuracy of risk measurement models*, volume 95. Division of Research and Statistics, Division of Monetary Affairs, Federal Reserve Board.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Lin, H. Y. (2013). Dynamic stock return-volume relation: Evidence from emerging Asian markets. *Bulletin of Economic Research*, 65(2):178–193.

Liu, W., Semeyutin, A., Lau, C. K. M., and Gozgor, G. (2020). Forecasting value-at-risk of cryptocurrencies with RiskMetrics type models. *Research in International Business and Finance*, 54:101259.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676.

Longin, P. and Pagliardi, G. (2016). Tail relation between return and volume in the US stock market: an analysis based on extreme value theory. *Economics Letters*, 145:252–254.

Marsh, T. and Wagner, N. (2004). Return-volume dependence and extremes in international equity markets. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=424926`, accessed August 2022.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4):271–300.

Naeem, M., Saleem, K., Ahmed, S., Muhammad, N., and Mustafa, F. (2020). Extreme return-volume relationship in cryptocurrencies: tail-dependence analysis. *Cogent Economcis Finance*, 8(1):1834175.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York.

Osterrieder, J. and Lorenz, J. (2017). A statistical risk assessment of Bitcoin and its extreme tail behavior. *Annals of Financial Economics*, 12(1).

Osterrieder, J., Lorenz, J., and Strika, M. (2016). Bitcoin and cryptocurrencies – not for the faint-hearted. *International Finance and Banking*. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2867671`, accessed August 2022.

Panjer, H. H. (2006). *Operational Risk: Modeling Analytics*. John Wiley Sons, New Jersey.

Park, M. H. and Kim, J. H. T. (2016). Estimating extreme tail risk measures with generalized Pareto distribution. *Computational Statistics & Data Analysis*, 98:91–104.

Pickands, J. (1975). Statistcal inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Pratiwi, N., Iswahyudi, C., and Safitri, R. I. (2019). Generalized extreme value distribution for value at risk analysis on gold price. *Journal of Physics: Conference Series*, 1217(1):012090.

R Development Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ren, F. and Giles, D. E. (2010). Extreme value analysis of daily Canadian crude oil prices. *Applied Financial Economics*, 20(12):941–954.

Righi, M. B. and Ceretta, P. S. (2015). A comparison of expected shortfall estimation models. *Journal of Economics and Business*, 78:14–47.

Saatcioglu, K. and Starks, L. T. (1998). The stock price-volume relationship in emerging stock markets: the case of Latin America. *International Journal of Forecasting*, 14(2):215–225.

Shamiri, A., Hamzah, N. A., and Pirmoradian, A. (2011). Tail dependence estimate in financial market risk management: Clayton-Gumbel copula approach. *Sains Malaysiana*, 40(8):927–935.

Sigauke, C., Verster, A., and Chikobvu, D. (2013). Extreme daily increases in peak electricity demand: Tail-quantile estimation. *Energy Policy*, 53:90–96.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, volume 8, pages 229–231.

Som, A. and Kayal, P. (2022). A multicountry comparison of cryptocurrency vs gold: Portfolio optimization through generalized simulated annealing. *Blockchain: Research and Applications*, 3(3):100075.

Wentzel, D. C. and Mare, E. (2007). Extreme value theory - an application to the South African equity market. *Investment Analysts Journal*, 36(66):73–77.

Wong, T. S. T. and Li, W. K. (2010). A threshold approach for peaks-over-threshold modeling using maximum product of spacings. *Statistica Sinica*, 20(3):1257–1272.

Zhang, W., Wang, P., Li, X., and Shen, D. (2018). Multi-fractal detrended cross-correlation analysis of the return-volume relationship of Bitcoin market. *Complexity*, 2018(1):8691420.

Zhang, Y., Chan, S., and Nadarajah, S. (2019). Extreme value analysis of high-frequency cryptocurrencies. *High Frequency*, 2(1):61–69.

Appendix A: Code

All the code for this thesis is written in **R** and some of it is provided below:

```
library(extRemes)
library(fExtremes)
library(tseries)
library(extRemes)
library(evir)
library(ismev)
library(psych)
library(nortest)
library(evd)
library(evmix)
library(QRM)
library(quantmod)
library(PerformanceAnalytics)
```

For Figures 2.1 to 2.4:

```
library(extRemes)
library(locfit)
library(logspline)
Plots of Gumbel and Frechet:
x <- revd(10000, loc=0, scale=1, shape=0, type="GEV")
x2 <- revd(10000, loc=0, scale=1, shape=0.1, type="GEV")
x3 <- revd(10000, loc=0, scale=1, shape=0.6, type="GEV")
plot(logspline(x), ylim=c(0, 0.5) , xlim=c(-4, 8), main="GEVD distributions",
xlab="x", ylab="f(x)", type="l", col="purple", lwd=2)
plot(logspline(x2), add=T, col="green", type="l", lwd=2)
plot(logspline(x3), add=T, col="red", type="l",  lwd=2)
colour <- c("purple", "green", "red")
```

```
label <- c("Gumbel", "Frechet (EVI=0.1)", "Frechet (EVI=0.6)")
legend("topright", col=colour, label, lwd=2)
Plots of Gumbel and Weibull:
plot(logspline(x), ylim=c(0, 0.5) , xlim=c(-4, 8), main="GEVD distributions",
xlab="x", ylab="f(x)", type="l", col="purple",  lwd=2)
x4 <- revd(10000, loc=0, scale=1, shape=-0.2, type="GEV")
x5 <- revd(10000, loc=0, scale=1, shape=-0.6, type="GEV")
plot(logspline(x4), add=T, col="blue", lwd=2)
plot(logspline(x5), add=T, col="orange", lwd=2)
colour2 <- c("purple", "blue", "orange")
labell <- c("Gumbel", "Weibull (EVI=-0.2)", "Weibull (EVI=-0.6)")
legend("topright", col=colour2, labell, lwd=2)
Code for GPD with zero and positive tail index:
y1 <- revd(10000, loc=0, scale=1, shape=0, type="GP")
y2 <- revd(10000, loc=0, scale=1, shape=0.1, type="GP")
y3 <- revd(10000, loc=0, scale=1, shape=0.5, type="GP")
plot(logspline(y1), ylim=c(0, 1) , xlim=c(0, 6), main="GPD distributions", xlab="x",
ylab="f(x)", type="l", col="purple", lwd=2)
plot(logspline(y2), add=T, col="red", lwd=2)
plot(logspline(y3), add=T, col="green", lwd=2)
colour3 <- c("purple", "green", "red")
label <- c("EVI=0", "EVI=0.1", "EVI=0.5")
legend("topright", col=colour3, label, lwd=2)
Code for GPD with zero and negative tail index:
y1 <- revd(10000, loc=0, scale=1, shape=0, type="GP")
y4 <- revd(10000, loc=0, scale=1, shape=-0.2, type="GP")
y5 <- revd(10000, loc=0, scale=1, shape=-0.5, type="GP")
plot(logspline(y1), ylim=c(0, 1) , xlim=c(0, 6), main="GPD distributions", xlab="x",
ylab="f(x)", type="l", col="purple", lwd=2)
plot(logspline(y4), add=T, col="blue", lwd=2)
plot(logspline(y5), add=T, col="orange", lwd=2)
colour4 <- c("purple", "blue", "orange")
label2 <- c("EVI=0", "EVI=-0.2", "EVI=-0.5")
legend("topright", col=colour4, label2, lwd=2)
```

For time series plots in Figures 3.1 and 3.2:

```
library(lubridate)
```

```
thedata <- read.csv(file="CRYPTO DATA - BTC-USD KRAKEN.csv")
thedata$newtime <- parse_date_time(thedata$time, orders='mdYHM')
plot(thedata$newtime, thedata$closingprice, xlab= "Time (in years)" ,
ylab= "BTC hourly prices", type = 'l')
plot(thedata$newtime, thedata$volumeto..in.USD., xlab= "Time (in years)" ,
ylab= "Bitcoin trading volume (in USD)", type = 'l')
```

The data:

```
To model log returns, detrendeded volume and negative returns respectively:
logReturns = diff(log(thedata$closingprice))
diffvol = diff(log(thedata$volumeto..in.USD.))
negReturns = -logReturns
```

MLE method of fitting POT models to negative returns:

```
library(POT)
fit1 <- fitgpd(negReturns, threshold = 0.0, est = "mle")
fit2 <- fitgpd(negReturns, threshold = 0.01, est = "mle")
fit3 <- fitgpd(negReturns, threshold = 0.02, est = "mle")
fit4 <- fitgpd(negReturns, threshold = 0.03, est = "mle")
fit5 <- fitgpd(negReturns, threshold = 0.04, est = "mle")
fit6 <- fitgpd(negReturns, threshold = 0.05, est = "mle")
```

Tail dependence parameter and extreme correlation plot:

```
Extract MLE parameter estimates:
fitbvgpd(na.exclude(cbind( negReturns, diffvol)), threshold = c(0.05, 3.138),
model = "log", std.err.type = "observed", corr = TRUE, cshape = FALSE,
cscale = FALSE)$param
fitbvgpd(na.exclude(cbind( negReturns, diffvol)), threshold = c(0.04, 2.657),
model = "log", std.err.type = "observed", corr = TRUE, cshape = FALSE,
cscale = FALSE)$param
fitbvgpd(na.exclude(cbind( negReturns, diffvol)), threshold = c(0.03, 2.227),
model = "log", std.err.type = "observed", corr = TRUE, cshape = FALSE,
cscale = FALSE)$param
fitbvgpd(na.exclude(cbind( negReturns, diffvol)), threshold = c(0.02, 1.744),
```

```
model = "log", std.err.type = "observed", corr = TRUE, cshape = FALSE,
cscale = FALSE)$param
fitbvgpd(na.exclude(cbind( negReturns, diffvol)), threshold = c(0.01, 1.168),
model = "log", std.err.type = "observed", corr = TRUE, cshape = FALSE,
cscale = FALSE)$param
fitbvgpd(na.exclude(cbind( negReturns, diffvol)), threshold = c(0, -0.035),
model = "log", std.err.type = "observed", corr = TRUE, cshape = FALSE,
cscale = FALSE)$param
fitbvgpd(na.exclude(cbind( negReturns, diffvol)), threshold = c(0.0197, 1.727),
model = "log", std.err.type = "observed", corr = TRUE, cshape = FALSE,
cscale = FALSE)$param
thresh <- c(-5, -4, -3, -2, -1, -0.0001, 0, 1, 2, 3, 4, 5)
ExtrCorr <- c(0.004, 0.008, 0.02, 0.325, 0.155, 0.728, 0.763, 0.156, 0.29, 0.132,
0.085, 0.086)
plot(thresh, ExtrCorr, lwd = 2, ylim = c(0,1),
col = c(ifelse(thresh < 0, "black", "red")), pch = 21, cex = 1.2, bg = "red",
type = "o", xlab = "Threshold (%)", ylab = "Extreme correlation",
main = "Extreme correlation: returns and volumes")
```

Mean excess plots:

```
MEplot(logReturns, main = "Mean Excess plot", xlab = "Threshold (u)",
ylab = "Mean Excess e(u)")
MEplot(negReturns, main = "Mean Excess plot", xlab = "Threshold (u)",
ylab = "Mean Excess e(u)")
```

Histograms:

```
library(fExtremes)
par(mfrow=c(1,2)) hist(logReturns, xlab = "Bitcoin log returns",
main = "Histogram of returns", ylim = c(0,100), xlim = c(-0.16, 0.1), prob = TRUE)
curve(dnorm(x, mean = mean(logReturns), sd = sd(logReturns)), col="red",
lwd = 2, add = TRUE)
lines(density(logReturns), col = "blue", lwd = 2)
legend("topleft", legend=c("log returns", "normal"), col=c("blue", "red"),
lty=1, lwd=1.5)
```

Normal QQ plots:

```
par(mfrow=c(1,2))
QQplot(logReturns, reference = "normal")
qqline(logReturns, lwd = 2, col = "blue")
QQplot(diffvol, reference = "normal")
qqline(diffvol, lwd = 2, col = "blue")
```

Empirical distribution plots:

```
par(mfrow=c(1,3))
emdPlot(logReturns, doplot = TRUE, labels = TRUE)
emdPlot(negReturns, doplot = TRUE, labels = TRUE)
emdPlot(diffvol, doplot = TRUE, labels = TRUE)
```

Shape plots:

```
library(evmix)
library(fExtremes)
gpdShapePlot(logReturns, ci = 0.95, models = 30,  start = 10, end = 500,
doplot = TRUE, labels = TRUE, plottype = "normal")
gpdShapePlot(negReturns, ci = 0.95, models = 30,  start = 10, end = 500,
doplot = TRUE, labels = TRUE, plottype = "normal")
gpdShapePlot(diffvol, ci = 0.95, models = 30,  start = 10, end = 500,
doplot = TRUE, labels = TRUE, plottype = "normal")
```

Diagnostic plots:

```
library(POT)
fit8 <- fitgpd(logReturns, threshold = 0.029, est = "mle")
fit11 <- fitgpd(negReturns, threshold = 0.03, est = "mle")
fit15 <- fitgpd(diffvol, threshold = 2.67, est = "mle")
par(mfrow=c(1,3))
plot(fit8, which = 1:3, ci = TRUE)
par(mfrow=c(1,3))
plot(fit11, which = 1:3, ci = TRUE)
par(mfrow=c(1,3))
plot(fit15, which = 1:3, ci = TRUE)
library(fExtremes)
fit17 <- gpdFit(logReturns, u = 0.029, type = "mle")
```

```
fit20 <- gpdFit(negReturns, u = 0.03, type = "mle")
fit24 <- gpdFit(diffvol, u = 2.67, type = "mle")
par(mfrow=c(1,3)) plot(fit17, which = 2:4)
par(mfrow=c(1,3)) plot(fit20, which = 2:4)
par(mfrow=c(1,3)) plot(fit24, which = 2:4)
```

Return level plots:

```
par(mfrow=c(1,2))
plot(fit8, which = 4, ci = TRUE, opy = 365*24, main =
"Return level plot (positive returns)")
plot(fit11, which = 4, ci = TRUE, opy = 365*24, main =
"Return level plot (negative returns)")
```

Mean residual life plots:

```
par(mfrow=c(1,3))
mrlplot(logReturns, u.range = c(0, 0.1), xlab = "Threshold",
ylab = "Mean excess" , main = "Mean residual life plot", conf = 0.95,
col = c('red', 'black', 'red'), lwd = c(1, 1.5, 1))
mrlplot(negReturns, u.range = c(0, 0.1), xlab = "Threshold",
ylab = "Mean excess" , main = "Mean residual life plot", conf = 0.95,
col = c('red', 'black', 'red'), lwd = c(1, 1.5, 1))
mrlplot(diffvol, u.range = c(0, 4), xlab = "Threshold", ylab = "Mean excess",
main = "Mean residual life plot", conf = 0.95, col = c('red', 'black', 'red'),
lwd = c(1, 1.5, 1))
```

Test for 'goodness of fit':

```
library(dgof)
ks.test(logReturns[logReturns >  0.029], "pgpd", shape = 0.203, loc = 0.029,
scale = 0.012)
ks.test(negReturns[negReturns >  0.07], "pgpd", shape = 0.181, loc = 0.07,
scale = 0.027)
ks.test(diffvol[diffvol >  2.18], "pgpd", shape = 0.119, loc = 2.18, scale = 0.449)
```

Profile likelihood procedure:

```
par(mfrow=c(1,2))
confint(fit8, parm = "scale", level = 0.95)
confint(fit8, parm = "shape", level = 0.95)
par(mfrow=c(1,2))
confint(fit11, parm = "scale", level = 0.95)
confint(fit11, parm = "shape", level = 0.95)
par(mfrow=c(1,2))
confint(fit15, parm = "scale", level = 0.95)
confint(fit15, parm = "shape", level = 0.95)
```

For VaR and ES estimation using POT method at different quantile levels, first fit the GPD to obtain parameter estimates then extract the parameter estimates to calculate risk:

```
Risk_est = function(thereturns, u, p, n){
fit1 = fitgpd(negReturns, threshold = u, est = "mle")
V_ar = u + (fit1$param[1]/fit1$param[2])*(((n/fit1$nat)*p)^(-fit1$param[2]) - 1)
ES_est = V_ar/(1 - fit1$param[2]) + (fit1$param[1] -
u*(fit1$param[2]))/(1 - fit1$param[2])
return (list(Value_at_risk = V_ar, Expected_shortfall = ES_est))
}
E.g. to calculate risk estimates for the hourly Bitcoin GPD model (u = 0.0) at 1%
level of significance:
print(Risk_est(thereturns = negReturns, u = 0.0, p = 0.01, n = 24336))
For GPD models with shape parameter (EVI) = 0:
Risk_est2 = function(thereturns, u, p , n){
fit2 = fitgpd(negReturns, threshold = 0.01, est = "mle")
V_ar = u - fit1$param[2]*(log((n/fit1$nat)*(1 - p)))
ES_est = V_ar + fit1$param[2]
return (list(Value_at_risk = V_ar, Expected_shortfall = ES_est))
}
For example, to calculate risk estimates for the hourly Bitcoin GPD model (u = 0.01)
at 5% level of significance:
print(Risk_est(thereturns = negReturns, u = 0.0, p = 0.05, n = 24336))
```

Kupiec LR test and Christoffersen backtests:

The following function generates Kupiec Unconditional Coverage LR Test and Christoffersen backtests, where alpha $= p$ used in the VaR estimation:

```
LRuc = function(thesereturns, p = 0.01, VaR = 0.07, N = 2500){
sp <- length(which(thesereturns > VaR))
LRuc = 2*log((1 - sp/N)^(N - sp)*(sp/N) ^ sp) - 2*log((1 - p)^(N - sp)*(p ^ sp))
return (LRuc) }
E.g. LRuc(negReturns, p=0.01, VaR=0.0696819, N=24336)
The following function calculates Christoffersen Serial Independence Test statistic.
Let n00 = a, n01 = b, n10 = c, n11 = d; The violations vector has 0's on the days in
which violations do not occur and 1's on the ones where they do occur.
LRind <- function(thesereturns, p = 0.01, VaR = 0.07, N = 2500){
violation_vector <- rep(0,length(thesereturns))
violation_vector[which(thesereturns > VaR)] <- 1
nij_vector <- rep(NA,length(violation_vector))
for (i in seq(length(violation_vector) - 1)){
n00 = a; n01 = b, n10 = c; n11 = d
if ((violation_vector[i] == 0) & (violation_vector[i+1] == 0)) {nij_vector[i] <- 0}
if ((violation_vector[i] == 0) & (violation_vector[i+1] == 1)) {nij_vector[i] <- 1}
if ((violation_vector[i] == 1) & (violation_vector[i+1] == 0)) {nij_vector[i] <- 2}
if ((violation_vector[i] == 1) & (violation_vector[i+1] == 1)) {nij_vector[i] <- 3}
}
nij_vector <- factor(nij_vector, levels = 0:3, labels = c('a','b','c','d'))
n00 <- table(nij_vector)['a']
n01 <- table(nij_vector)['b']
n10 <- table(nij_vector)['c']
n11 <- table(nij_vector)['d']
pi0 = n01 / (n00 +n01)
pi1 = n11 / (n10 +n11)
pi = (n01 + n11) / (n00 + n01 + n10 + n11)
sp <- length(which(thesereturns>VaR))
LRuc = 2*log((1-sp/N)^(N-sp)*(sp/N)^sp) -2*log((1-p)^(N-sp)*(p^sp))
LRind = 2 * log  ((1 - pi0)^(n00) * (pi0^n01) * (1 - pi1)^(n10) * pi1^n11)
- 2 * log((1 - pi)^(n00 + n10) * pi^(n01 + n11))
LRcc = LRuc + LRind
p_value_LRuc = 1 - pchisq(LRuc, 1)
p_value_LRcc = 1 - pchisq(LRcc, 2)
return (list(Violations = violation_vector, Nij = nij_vector, sp = sp, LRuc = LRuc,
LRind = LRind, LRcc = LRcc, p_value_LRuc = p_value_LRuc, p_value_LRcc = p_value_LRcc)
}
```

```
For example:
a <- LRind(thesereturns=negReturns, p = 0.01, VaR = 0.0696819, N = 24336)
a
```

McNeil and Frey Backtest, where m is the number of VaR violations and r is the vector of standardised residual exceedances:

```
ES_backtest = function(thereturns, VaR_est, ES_est, std_err, p){
mu_0 = 0
m <- length(which(thereturns < VaR_est))
r <- rep(0, m)
violations <- rep(0, m)
violations <- thereturns[which(thereturns < VaR_est)]
for (i in (length(violations))){
r[i] = (violations[i] - ES_est)/std_err
}
r_mean = mean(r)
r_sd = sd(r)
T_stat = (r_mean - mu_0)/((r_sd)/sqrt(m))
p_val = pt(T_stat, m - 1)
return (P_value = p_val)
}
e.g. to backtest risk estimates from  hourly BTC at various levels of significance:
k1 <- ES_backtest(thereturns = negReturns, VaR_est = 0.02992651,
ES_est = 0.01589365, p = 0.01, std_err = 0.01)
k1
k2 <- ES_backtest(thereturns = negReturns, VaR_est = 0.04694447,
ES_est = 0.03493279, p = 0.01, std_err = 0.01)
k2
k3 <- ES_backtest(thereturns = negReturns, VaR_est = 0.1302061,
ES_est = 0.11054469, p = 0.01, std_err = 0.01)
k3
k4 <- ES_backtest(thereturns = negReturns, VaR_est = 0.03703038,
ES_est = 0.02493042, p = 0.02, std_err = 0.01)
k4
k5 <- ES_backtest(thereturns = negReturns, VaR_est = 0.05133343,
```

```
ES_est = 0.04221133, p = 0.02, std_err = 0.01)
k5
k6 <- ES_backtest(thereturns = negReturns, VaR_est = 0.14069586,
ES_est = 0.1233527, p = 0.02, std_err = 0.01)
k6
k7 <- ES_backtest(thereturns = negReturns, VaR_est = 0.04491356,
ES_est = 0.03498541, p = 0.05, std_err = 0.01)
k7
k8 <- ES_backtest(thereturns = negReturns, VaR_est = 0.05555862,
ES_est = 0.04921828, p = 0.05, std_err = 0.01)
k8
k9 <- ES_backtest(thereturns = negReturns, VaR_est = 0.15268891,
ES_est = 0.13799622, p = 0.05, std_err = 0.01)
k9
```