An Extreme Value Analysis of Log Returns of Financial Assets

by

Saumitra Mazumder
Supervisor: Dr. Foivos Xanthos

A thesis
presented to Ryerson University

Toronto, Ontario, Canada, 2019

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis presents a study of extreme movements in financial asset returns. According to Extreme Value Theory, the distribution of an extreme financial asset return obeys a Fréchet distribution. Such a distribution of extreme realizations depend on the characterization of the tail of the distribution of the asset return. We investigate this tail behaviour by the Hill method for estimating the heaviness of the tail.

Using data from a random collection of stocks and indexes, we show empirically that financial asset returns have characteristics that preclude the log-normal assumption of distribution of prices in the Black-Scholes model. Furthermore, we show that the distribution of asset returns skew away from the mean and such a distribution has heavy tails. Hence, asset returns will have extreme realizations with greater probability that predicted by Black-Scholes model.

Acknowledgements

I would like to thank my supervisor, Professor Foivos Xanthos. Working on this thesis has been a rewarding experience in no small part thanks to your encouragement and patience.

Dedication

To my puppy, Juniper.

Table of Contents

List of Tables

List of Figures

Introduction

Extreme movements in prices of financial assets are rare, but their effects are profound and long-lasting. The study of these so-called Black Swan events lead greater attention to modelling the distribution of extreme realizations of financial asset returns. Ultimately, risk managers must relate their risk measures (VAR, cVAR) to the probability of an extraordinary event under normal market conditions.

It was first noted by Mandelbrot and then others after that the returns of financial assets were poorly fitted by a Gaussian distribution. Empirical analysis by Mandelbrot showed that the distribution of returns have tails that are heavier than those proposed by the famed Black-Scholes model [22, 23]. Furthermore, asset return distributions display strong asymmetry. Such characteristics imply that a Gaussian model for returns does not correspond to reality and will grossly underestimate the risk of extreme market swings. Following Mandelbrot's pioneering work, numerous studies have documented that financial time series can be well approximated with tails that decay by the power law.

In this thesis, we considered Extreme Value Theory to study rare events and its application to financial assets. Since extreme events occur at the tails of asset return distributions, we use Extreme Value Theory to estimate and make inferences about how fast a tail of a distribution decays. Extreme Value Theory studies the limiting distribution of large realizations (either negative or positive) of a series.

Outline: We begin in Chapter 1 by giving an overview of financial asset returns and an exposition of the Black-Scholes assumption for prices of financial assets. In Chapter 2 we discuss kurtosis and skewness. We show that asset returns have kurtosis and skewness that preclude the log-normal assumption of prices in the Black-Scholes model. We also discuss how such measures are not sufficient for determining which type of distribution an asset return will have. Chapter 3 represents the bulk of this thesis, wherein we review Extreme Value Theory for independent, identically distributed random variables. We then extend these results to random variables with a type of weak dependence. Chapter 4 introduces our methodology for estimating the decay of the tails of asset returns. Chapter 5 presents our data and some inferences that can be made. Finally, in Chapter 6 we present the conclusions that we drew from a study of 10 selected assets. Some motivations for further research are also presented.

Chapter 1

Preliminaries

1.1 Financial Returns

Consider the closing value of a financial asset at time t – perhaps a stock, exchange rate or market index – as S_t (see Fig. 1.1). Denote the logarithm of this price as $X_t = \ln S_t$.

Definition 1.1.1 (Non-Overlapping Logarithmic Return)**.** *Consider the sequence of random variables $\{X_t\}_{t\in\mathbb{N}}$ that represent the daily logarithmic closing price of a financial asset. The* non-overlapping (log-)return *of an asset over some time scale, say $\Delta t \in \mathbb{R}$, is denoted by:*

$$r_t = X_{t+\Delta t} - X_t \quad \textit{where } t = i\Delta t, i \in \mathbb{N}$$

In the proceeding, our empirical analysis focuses on time increments of one day ($\Delta t = 1$, see Fig. 1.2), five days ($\Delta t = 5$) and twenty-five days ($\Delta t = 25$), henceforth referred to as daily, weekly and monthly returns.

For small price movements, the logarithmic return is approximately equal to the relative return $\left(\dfrac{S_{t+\Delta t} - S_t}{S_t}\right)$. However, in contrast to the relative return series, the logarithmic return series $\{r_t\}_{t\in\mathbb{N}}$, does not depend on monetary units, thus facilitating comparisons between assets [14]. Henceforth, we will refer to the logarithmic return as simply the return.

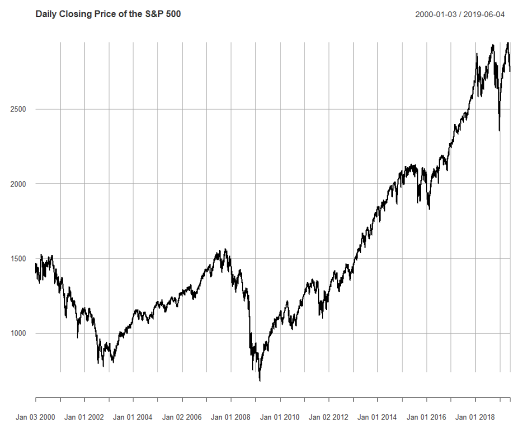

Figure 1.1: The daily closing price for the S&P 500 index

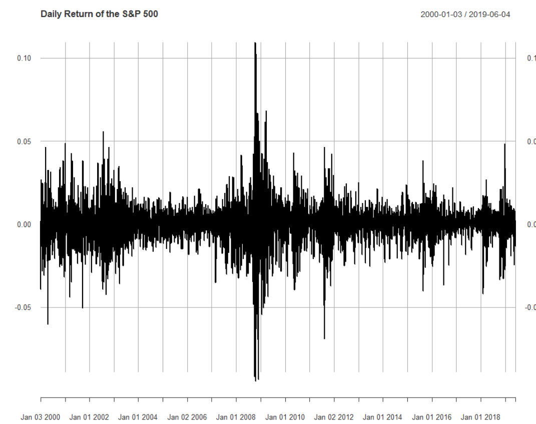

Figure 1.2: The daily log returns for the S&P 500 index

1.2 The Black-Scholes Model

It was not until the celebrated model by Fischer Black, Myron Scholes and Robert Merton, eponymously called the Black-Scholes (-Merton) formula, that there was a method available to analytically value options. In the model, the price of a financial asset (S_t) is modelled by the following stochastic differential equation (SDE)

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \tag{1.1}$$

Where, W_t is a *standard Wiener process* with distribution $\mathcal{N}(0, t)$.

Consider $X_t = \ln S_t$. By applying Itô's Lemma, Eq. (1.1) admits the following classical decomposition

$$dX_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t. \tag{1.2}$$

Thus, the *logarithm* of a stock is *generalized a Wiener process* with distribution $\mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right) t, \sigma^2 t\right)$.

Discretizing Eq. (1.2), for small Δt we have

$$X_{t+\Delta t} - X_t = \left(\mu - \frac{\sigma^2}{2}\right) \Delta t + \sigma \left(W_{t+\Delta t} - Wt\right),$$

i.e. the process' increments in a small interval of length Δt after time t is $\left(\mu - \frac{\sigma^2}{2}\right)\Delta t$ plus a random fluctuation which is $\mathcal{N}\left(0, \sigma^2 \Delta t\right)$ distributed [15, 30].

Since X_t is normally distributed, S_t itself is log-normally distributed. Hence, the Black-Scholes Model implies that *prices* of an underlying financial asset have log-normal distribution. Accordingly, the series of non-overlapping financial returns as defined in Definition 1.1.1 should be also normally distributed.

However, the empirical analysis presented in this thesis and from others have shown that the distribution of financial return assets are often skewed and heavier-tailed than can be described by a normal distribution. This can be visually confirmed via the so-called QQ-plot (see Fig. 1.3).

In general, the QQ-plot for a financial return series shows the upper tail turning upwards away from the norm line. Similarly, the lower tail turns downward away from the norm line. This suggests that return series have distributions with heavier tails than predicted by the Black-Scholes Model. Indeed, the empirical return distributions tend to have larger forth moments than predicted by the Black-Scholes model [22]. Furthermore, it is often observed that the distribution of returns tend to decay with a Pareto like tail [8].

As such, realized volatility in returns tend to exceed that which are expected under the log-normal assumption of prices in the Black-Scholes model. Such discrepancies have obvious implications for both risk management and instrument pricing methodologies. Fig. 1.4 shows the characteristic peak near the mean of heavy-tailed distributions of return series when compared to predicted log-normal price assumption of the Black-Scholes model.

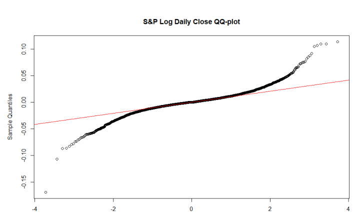

Figure 1.3: The QQ-plot of Daily Log Returns of the S&P index.

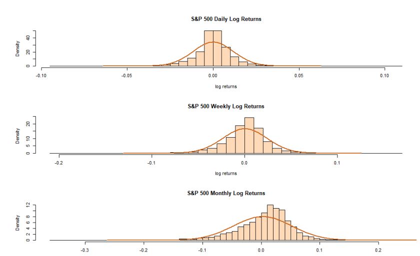

Figure 1.4: The histogram of the log return of the S&P 500 and the predicted distribution of the Black-Scholes Model for daily, weekly and monthly returns

The above characteristics implies that the Black-Scholes Model is insufficient for studying financial returns. A collection of characteristics contrary to the Black-Scholes model were first documented by Mandelbrot [22] and are common across a wide range of financial instruments, markets and times scales. These so-called *Stylized Facts* suggest that the stochastic process driving the volatility of an asset return deviates from the log-normal assumption of the Black-Scholes model[1].

[1] For a list of these Stylized Facts, see Empirical properties of asset returns: stylized facts and statistical issues by Rama Cont [8].

Chapter 2

Kurtosis, Skewness and The Return Distribution

It is often difficult to obtain descriptive statistics which require the probability distributions of asset returns, since in general, such distributions are unknown. However, it is possible to approximate the true distribution by a simpler distribution obtained by a limiting argument using Definition 2.0.1 and Theorem 2.0.2.

Definition 2.0.1 (Convergence in Distribution of Random Variables)**.** *A sequence of random variables $\{X_i\}_{i\in\mathbb{N}}$ having distribution functions $\{F_i\}_{i\in\mathbb{N}}$ is said to* Converge in Distribution (or weakly converge) *to a random variables X, having distribution function F, denoted $X_n \xrightarrow{d} X$, if*

$$F_n(x) \to F(x) \quad \text{as } n \to \infty$$

for all continuity points x to F

In probability, we can prove convergence in distribution by looking at corresponding characteristic functions. However, for our purposes, we will use the following theorem,

Theorem 2.0.2 (Helly-Bray Theorem)**.** *For a sequence of random variables $\{X_i\}_{i\in\mathbb{N}}$ and random variable X, the following are equivalent*

(i) $X_n \xrightarrow{d} X$,

(ii) for all real bounded and continuous functions g, $\mathbb{E}\left[g(X_n)\right] \to \mathbb{E}\left[g(X)\right]$ as $n \to \infty$.

The idea of Theorem 2.0.2 is that convergence in distribution can be translated into the convergence of expectations for all real bounded and continuous functions. These two concepts will also be useful in Chapter 3.

2.1 Kurtosis and Skewness

Let X be a random variable, then $\mu = \mathbb{E}[X]$ is defined as the mean. Denote $\mu_q = \mathbb{E}[(X-\mu)^q]$ as the q^{th} *central moment* of random variable X. We can define a measure for the "heaviness" of the tail of probability distribution function f of random variable X and how much the f "skews" away from μ for well-behaved continuous distributions as follows

Definition 2.1.1 (Skewness and Kurtosis)**.** *The* Skewness *coefficient of the distribution is defined to be*

$$\tau = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}.$$

The kurtosis excess *coefficient is*

$$\kappa = \frac{\mu_4}{\mu_2^2} - 3.$$

Let $\{X_i\}_{i=1}^n$ be realizations of X. We formulate the sample skewness and kurtosis to estimate the true statistic. Denote $\hat{\mu_q} = \dfrac{\sum_{i=1}^n (X_i - \hat{\mu})^q}{n}$ as the q^{th} *sample moment*, where $\hat{\mu} = \dfrac{\sum_{i=1}^n X_i}{n}$ is the *sample mean.*

Definition 2.1.2 (Sample Skewness and Sample Kurtosis)**.** *The* sample Skewness *is*

$$\hat{\tau} = \frac{\hat{\mu_3}}{\hat{\mu_2}^{\frac{3}{2}}}$$

And sample kurtosis excess *is*

$$\hat{\kappa} = \frac{\hat{\mu_4}}{\hat{\mu_2}^2} - 3$$

When $\{X_i\}$ are iid and normal, the sample estimators converge asymptotically as follows, $\sqrt{n}\hat{\tau} \sim \mathcal{N}(0,6)$ and $\sqrt{n}\hat{\kappa} \sim \mathcal{N}(0,24)$ [1].

A normal distribution has skewness coefficient (τ) of zero. When a probability distribution distribution f is unimodal, a negative skew coefficient commonly indicates that f has more density to the left of the median. Similarly, a positive skew coefficient indicates that density is concentrated to the right of the median. Hence for well behaved distributions, any value skewness coefficient value other than zero indicates the probability density function of the return series is at least asymmetric and has more mass at one tail over the other. However, if one tail is long while the other is fat, then the previous characterization is voided [29].

A normal distribution has *kurtosis excess* (κ) of zero. Distributions with zero excess kurtosis are called mesokurtic, but are not necessarily normal. Distributions with positive excess kurtosis are called leptokurtic, and tend to have a distinct peak near the mean, decline faster than a normal distribution, and have heavy tails. Distributions with negative excess kurtosis are called platykurtic, they have a characteristic flat top near the mean and shorter, thinner tails.

While the measures of kurtosis and skewness can be used in identifying if the distribution of a return series is not normal, they are not sufficient for identifying what type the distributions of returns are (see Example 2.1.3).

Example 2.1.3. *Consider the* Laplace(μ, β) *distribution*

$$f(x \mid \mu, \beta) = \frac{1}{2\beta} e^{-\frac{|x-\mu|}{\beta}}.$$

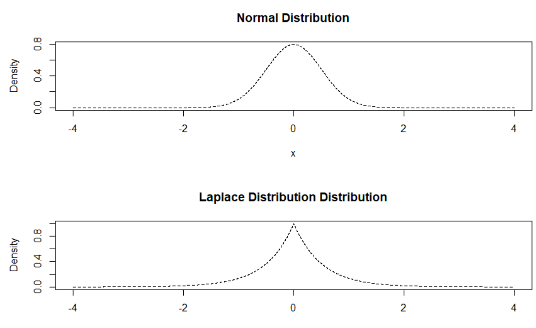

Figure 2.1: Comparison of $Laplace(0, \frac{1}{2})$ and $\mathcal{N}(0, \frac{1}{2})$

It can be shown that the mean, variance, skewness and kurtosis excess are as follows [26]

$$\textit{(mean)}\ \ \mu = \mu$$

$$\textit{(variance)}\ \ \mu_2 = 2\beta^2$$

$$\textit{(skewness)}\ \ \tau = 0$$

$$\textit{(kurtosis excess)}\ \ \kappa = 0$$

Then the Laplace $\left(0, \frac{1}{2}\right)$ *distribution has the same descriptive statistic as a normal distribution* $\mathcal{N}\left(0, \frac{1}{2}\right)$*, however it is clearly not normal (See Fig. 2.1). Hence, while kurtosis excess and skewness may be used to reject the Gaussian hypothesis, it cannot be used to confirm it.*

Chapter 3

Tail Density

Mandelbrot proposed that so-called *heavy-tailed distributions* govern the asset returns of economic and financial return series [22]. He observed that the number of extreme observations found in the data of financial return series preclude a Gaussian density [23]. Such time series are characterized by their high variability (See Chapter 2). Ultimately, his research showed that estimation of extreme events using the Black-Scholes model is insufficient, as such a model would underestimate the probability of having either a large positive or negative realizations [5].

Accordingly, Mandelbrot stated that financial asset returns were best modelled as independent or at least stationary processes, whereby the behaviour of the tails of the underlying distribution, $1 - F(x)$, follow a Pareto distribution for sufficiently large x. The Pareto distribution is a heavy-tailed and skewed distribution. Since Pareto distributions are heavy-tailed, they are often used to model rare events.

Definition 3.0.1 (Heavy Tailed Distribution)**.** *The distribution a random variable X with distribution function F has a Pareto right tail with* tail index $\alpha > 0$ *if the exceedance probability*

$$\Pr[X > x] \equiv 1 - F(x) = \mathcal{L}(x)x^{-\alpha} \quad \textit{for } x \geq x_0,$$ [1]

where $\mathcal{L}(x)$ is some slowing varying function[2] *and x_0 is some threshold above which the*

[1] See Karamata's Theory to observe that the tail with Pareto distribution function can be written as $\Pr[X > x] = \mathcal{L}(x)x^{-\alpha}$.

[2] A measurable function $\mathcal{L} : (0, +\infty) \to (0, +\infty)$ is called slowly varying if for all $a > 0$,

$$\lim_{x \to \infty} \frac{\mathcal{L}(ax)}{\mathcal{L}(x)} = 1.$$

power law holds.

Here, the *tail index* α, summarizes the "weight" of the tail of the distribution of the random variable X. When x is large enough, $\mathcal{L}(x)$ converges to a constant, although it may converge very slowly. Distributions that have tails with the above characteristic include the Pareto and the Student-t distributions [24]. Heavy-tailed model are said to occur naturally in financial asset returns [8].

Clearly, an estimation of the tail index α is crucial, however this estimation is a non-trivial problem.

3.1 Extreme Value Theory

Initially, suppose that the return series of financial assets, $\{r_t\}_{t=1}^{n}$ are *independent and identically distributed* (this restriction will be lifted in Section 3.1.1) with distribution function F. In order to solve the general limit problem of extremes, we rely on the concepts from probability theory as defined in Definition 2.0.1 and Theorem 2.0.2.

Definition 3.1.1 (Maximum Order Statistic)**.** *For a sample of* n *observations* $\{r_t\}_{t=1}^{n}$*, the ordered samples,* $r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(n)}$ *are called the order statistics where* $r_{(j)}$ *is the* j^{th} *largest realization and is called the* j^{th} order statistic*. A special case of the order statistic is as follows,*

$$r_{(n)} = \max_{0 \leq t \leq n} \{r_t\}$$

is defined as the **Maximum order statistic** *of the time series* $\{r_t\}$*.* [3]

Recall from Definition 1.1.1, $\{r_t\}$ represents values of a process measured on a regular time scale. Hence, $r_{(n)}$ in Definition 3.1.1 represents the largest observation of the stochastic process over n time units.

When the distribution F of such a process is known the limiting distribution function of the largest order statistic, denoted $F_{(n)}(\cdot)$ can be derived for any value of n as follows

[3] We focus on the properties of maximum return, however any analysis can be applied also to the minimum return, since $r_{(1)} = -\max_{1 \leq t \leq n}\{-r_t\}$.

$$
\begin{aligned}
F_{(n)}(x) = \Pr[r_{(n)} \le x] &\overset{max.}{=} \Pr[r_1 \le x, r_2 \le x, \ldots, r_n \le x] \\
&\overset{ind.}{=} \prod_{j=1}^{n} \Pr[r_j \le x] \\
&= \prod_{j=1}^{n} F(x) = [F(x)]^n
\end{aligned} \tag{3.1}
$$

Intuitively, this states that since extreme realizations occur near the upper end of F, the asymptotic behaviour of $r_{(n)}$ must somehow be related to the right tail of the probability density f, near its right endpoint [6].

Denote $x_F := \sup\{x : F(x) < 1\}$ as the rightmost endpoint of F (for now, assume that $x_F < \infty$, x_F may indeed be infinite). It can be shown that $r_{(n)}$ converges in probability to x_F as $n \to \infty$.

Proposition 3.1.2. *$r_{(n)}$ converges in probability to x_F*

Proof: Denote $x_F := \sup\{x : F(x) < 1\}$ and fix $\epsilon > 0$. Then,

$$
\begin{aligned}
\lim_{n\to\infty} \Pr[|r_{(n)} - x_F| > \epsilon] &= \lim_{n\to\infty} \Pr[x_F < r_{(n)} - \epsilon \cup x_F > r_{(n)} + \epsilon] \\
&\le \lim_{n\to\infty} \left[\Pr[r_{(n)} > x_F + \epsilon] + \Pr[r_{(n)} < x_F - \epsilon]\right] \\
&= \lim_{n\to\infty} \left[1 - [F(x_F + \epsilon)]^n + [F(x_F - \epsilon)]^n\right] \\
&= 1 - 1 - 0 = 0
\end{aligned}
$$

Hence, $r_{(n)} \xrightarrow{p} x_F$. ∎

However, F is generally unknown (hence, $F_{(n)}$ is also not known). Furthermore, when $x_F = \infty$ it is not clear as to how to find a limiting distribution of the maximum order statistic. Regardless, $[F(x)]^n$ is said to be a degenerate distribution function[4], since for all $x < x_F$, $\Pr\left[r_{(n)} \le x\right] = 0$, and for all $x > x_F$ $\Pr\left[r_{(n)} \le x\right] = 1$ as $n \to \infty$. Such a distribution is not of practical use to us.

[4] A distribution function where for some constant x_0,

$$
F(x) = \begin{cases} 0 & \text{for } x < x_0 \\ 1 & \text{if } x \ge x_0 \end{cases}
$$

.

In order to analyze the extreme behaviour of a financial return series, we need conditions on F that ensure that there exists a pair of sequences of numbers $(\alpha_n > 0, \beta_n)$ which reduce and scale $r_{(n)}$, such that the distribution of the standardized extremes $\dfrac{r_{(n)} - \beta_n}{\alpha_n}$ is non-degenerate. In other words, we seek a limiting distribution which is not concentrated on a single point [6].

Hence, we concern ourselves with sequences $\{(\alpha_n > 0, \beta_n)\}_{n\in\mathbb{N}}$ that appropriately standardize the distribution of $r_{(n)}$ such that

$$\lim_{n\to\infty} \Pr\left[\frac{r_{(n)} - \beta_n}{\alpha_n} < x\right] = \lim_{n\to\infty} F_{(n)}(\alpha_n x + \beta_n) \to G(x)$$

Where G is non-degenerate. More explicitly, we seek conditions on the the tail of F such that

$$\lim_{n\to\infty} [F(\alpha_n x + \beta_n)]^n = \lim_{n\to\infty} \left(1 - n\frac{1 - F(\alpha_n x + \beta_n)}{n}\right)^n,$$

converges to some non-trivial limit. Recalling an elementary result in calculus, $\lim_{n\to\infty} \left(1 - \frac{c_n}{n}\right)^n = e^{-c} \iff \lim_{n\to\infty} c_n = c$, we obtain Proposition 3.1.3.

Proposition 3.1.3. [5] *Denote* $\overline{F}(x) = \Pr(X > x)$ *and* $u_n := \alpha_n x + \beta_n$. *For* $0 \le \tau \le \infty$ *and a sequence of real numbers* $\{u_n\}_{n\ge1}$, *it holds for* $n \to \infty$ *that the following are equivalent*

(i) $n\overline{F}(u_n) \to \tau$,

(ii) $\Pr[r_{(n)} \le u_n] = F^n(u_n) \to e^{-\tau}$.

Proposition 3.1.3 states that depending on the tail of F, it is expected that for pairs of real number sequences $\{\alpha_n > 0, \beta_n\}$ that an asymptotic distribution *can* emerge. The proposition *does not* state that such an asymptotic distribution will occur.

[5] For proofs of this proposition, see Extreme Value Analysis: an Introduction by Charras-Garrido and Lezaud [6] or Extremes and Related Properties of Random Sequences and Processes by Leadbetter, Lindgren and Rootzén [20].

3.1.1 Limiting Distribution of The Largest Order Statistic of An iid Time Series

Definition 3.1.4 (Maximum Domain of Attraction (MDA))**.** *If there exists a sequence of pairs of real numbers* $(\alpha_n > 0, \beta_n)$ *such that*

$$\Pr\left[\frac{r_{(n)} - \beta_n}{\alpha_n} \leq x\right] = F^n\left(\alpha_n x + \beta_n\right) \to G(x) \quad \text{as } n \to \infty$$

where G *is some non-degenerate distribution function, then* F *is said to be in the* maximum domain of attraction *of* G*, denoted* $F \in MDA(G)$*.*

Here after, we assume that the underlying marginal distribution function F is continuous and strictly increasing. It follows from Definitions 2.0.1 and 3.1.4, that if $F^n\left(\alpha_n x + \beta_n\right) \to G(x)$ as $n \to \infty$ then the random variables $\{X_n := \frac{r_{(n)}-\beta_n}{\alpha_n}\}_{n\in\mathbb{N}} \xrightarrow{d} X$, where X has distribution G as defined in Definition 3.1.4.

Two natural questions arise from Definition 3.1.4:

(i) Can more than one possible non-degenerate function appear as a limit in Definition 3.1.4, i.e., can $F \in MDA(G)$ and $F \in MDA(H)$?

(ii) What are the characteristics of F for which there exist sequences α_n and β_n such that $F \in MDA(G)$?

Theorem 3.1.5 (Khintchine's Theorem)**.** [6] *Let* $\{F_n\}_{n\in\mathbb{N}}$ *be distribution functions, and let* G *be a non-degenerate distribution function. Suppose there exists a pair of sequences of real numbers* $\{\alpha_n > 0, \beta_n\}$ *such that*

$$F_n\left(\alpha_n x + \beta_n\right) \to G(x)$$

then it holds that there are is another pair of sequences of real numbers $\{a_n > 0, b_n\}$ *such that*

$$F_n\left(a_n x + b_n\right) \to H(x)$$

if and only if

$$\frac{a_n}{\alpha_n} \to \alpha, \quad \frac{b_n - \beta_n}{\alpha_n} \to \beta, \quad \text{as } n \to \infty$$

then $H(x) = G(\alpha x + \beta)$

[6] For a proof Khintchine's Theorem, see Extremes and Related Properties of Random Sequences and Processes by Leadbetter, Lindgren and Rootzén [20]

Theorem 3.1.5 makes precise that different choices of the scaling sequences α_n and β_n lead to distributions that are related by a transformation. Hence, the norming sequences such that a limiting distribution exists are not necessarily unique, but the asymptotic distribution function is. Hence, we require a characterization on the tail of F such that norming sequences α_n and β_n exist.

Theorem 3.1.6 (Fisher-Tippet, Gnedenko)**.** *Suppose $F \in MDA(G)$ for some non-degenerate limiting distribution G as defined in Definition 3.1.4, then G must have one of following distribution functions known as* Generalized Extreme Value (GEV) distribution.

$$G_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-\frac{1}{\xi}}\right) & \text{if } \xi \neq 0 \\ \exp\left(-e^{-x}\right) & \text{if } \xi = 0 \end{cases} \quad \text{where } x \text{ is such that } 1+\xi x > 0.$$

The case $\xi = 0$ can be thought of as the limit of the distribution function as $\xi \to 0$.

The Gumbel distribution corresponds with $\xi = 0$, the Fréchet distribution with $\xi > 0$ and the Weibull distribution with $\xi < 0$ (See Fig. 3.1). Here the parameter ξ is referred to as the *shape parameter* and it governs the behaviour of the limiting distribution. $\frac{1}{\xi}$ is called the *tail index* of the distribution [25].

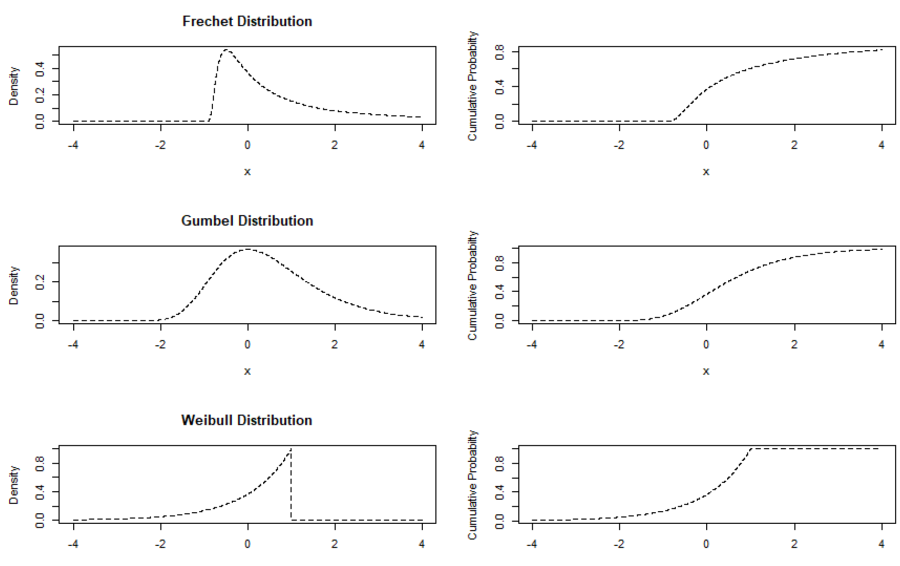

Figure 3.1: The probability density and marginal density respectively of a standard GEV distribution in three cases: $\xi = 1$ (Fréchet); $\xi = 0$ (Gumbel); and $\xi = -1$ (Weibull). For all cases $(\alpha_n, \beta_n) = (1, 0)$

Example 3.1.7 (McNeil, Frey and Embrechts). *If the underlying distribution is a Pareto distribution* $Pa\left(\frac{1}{\xi},\kappa\right)$ *with df* $F(x) = 1 - \left(\frac{\kappa}{\kappa+x}\right)^{\frac{1}{\xi}}$ *for* $\xi > 0, \kappa > 0, x \geq 0$*, we can take normalizing sequences* $c_n = \xi\kappa n^{\xi}$ *and* $d_n = \kappa n^{\xi} - \kappa$*. Using Definition 3.1.4 we get,*

$$\begin{aligned} F^n(c_n x + d_n) &= \left(1 - \left[\frac{1}{\xi x n^{\xi} + n^{\xi}}\right]^{\frac{1}{\xi}}\right)^n, \\ &= \left(1 - \frac{(1+\xi x)^{-\frac{1}{\xi}}}{n}\right)^n, \qquad 1+\xi x \geq n^{-\xi}, \\ \lim_{n\to\infty} F^n(c_n x + d_n) &= \exp\left(-(1+\xi x)^{-\frac{1}{\xi}}\right) \qquad 1+\xi x > 0 \end{aligned}$$

from which we conclude that $F \in MDA\left(G_{\xi>0}\right)$.

It may be noted that Theorem 3.1.6 is a special case of Proposition 3.1.3 using a linear parameterization where $\tau = -\ln G(x)$, $u_n = \alpha_n x + \beta_n$, thus a necessary and sufficient condition for the limit G is

$$n(1 - F(\alpha_n x + \beta_n)) \to -\ln G(x), \quad \text{as } n \to \infty$$

for any x and some pair of real sequences $\{\alpha_n > 0, \beta_n\}$. This explains the relevance of the tail $1 - F(x)$ for the Maximum Domain of Attraction criterion from Definition 3.1.4.

The aforementioned Fisher-Tippet-Gnedenko Theorem 3.1.6 has three important implications:

(i) The tail behaviour of the random variable, denoted $\overline{F}(x)$, determines the limiting distribution of the normalized maximum $r_{(n)}$ and not the specific distribution, $F(x)$ (See Theorem 3.1.8) [28].

(ii) The maximum of a sample of iid random variables after proper normalization can converge to a non-degenerate distribution and that distribution will be one of 3 possible distributions, the Gumbel distribution ($\xi = 0$), the Fréchet distribution ($\xi > 0$) , or the Weibull distribution ($\xi < 0$).

(iii) Although the above result states that when the limiting distribution of maxima has a limit, it will be in the *GEV distribution* family, it does not ensure the existence of such a limit. (E.g., Poisson Distribution).

Of course, α_n and β_n can be determined when F is known (See Example 3.1.7). However, we wish to determine the conditions on $\overline{F}(x)$ whereby the asymptotic distribution of the maximum order statistic exists when F is not known. We consider only distributions $\overline{F}(x)$ that are considered "heavy-tailed" (See Definition 3.0.1) since they correspond to Mandelbrot's observations of financial assets returns.

The following characterization of the tail behaviour was given by Gnedenko wherein the tail of distribution F decays by the power law.

Theorem 3.1.8 (Fréchet MDA, Gnedenko)**.** *For $\xi > 0$*

$$\overline{F}(x) = \mathcal{L}(x)x^{-\frac{1}{\xi}} \iff F \in MDA(G_\xi)$$

Where, $\mathcal{L}(x)$ is some slowing varying function as $x \to \infty$.

Accordingly, Theorem 3.1.8 states that any distribution that falls into the *maximum domain of attraction* of a Fréchet class ($\xi > 0$) GEV distribution can be modelled by a Pareto distribution. Equivalently, the tail index α of a heavy-tailed distribution as defined in Definition 3.0.1 is related to the parameter ξ for a Fréchet class GEV distribution.

Consequently, the entire tail of F can be utilized for fitting a distribution (see Chapter 4). Theorem 3.1.8 states that the distribution of interest has a *negative tail index*, $(-\frac{1}{\xi})$ and the tail decays with rate $\alpha = \frac{1}{\xi}$. Recall that ξ is called the *shape parameter* which is the main parameter of interest.

Extending Extreme Value Theory to Strictly Stationary Processes

Until now, our assumptions underlying the exposition of the asymptotic distribution of the maximum value presented in Section 3.1 required that the return series be iid. However, the iid assumption is too simple to be a description of real-life phenomena. In particular, the iid assumption is violated for financial asset returns, since often the realizations of a process are dependent on its recent history (see Fig. 3.2). Furthermore, successive observations for returns exhibit nonlinear correlation [8]. This dependence can have a significant impact on the analysis and interpretation of such data.

3.1.2 Volatility and Nonlinear Dependence

Definition 3.1.9. *The auto-correlation function (ACF) of $\{X_i\}$ at lag h is*

$$\rho(i,h) = \frac{Cov(X_i, X_{i+h})}{\sqrt{Var(X_i)Var(r_{X+h})}}$$

Typically, a return series for financial assets do not exhibit any significant auto-correlation. That is, $\rho(h)$ rapidly decays to zero as h increases. However, this *does not* imply that such time series are. Independence however, implies that any nonlinear function of returns will have *no* auto-correlation [8]. In general, this property does not hold (see Fig. 3.2) for financial asset returns.

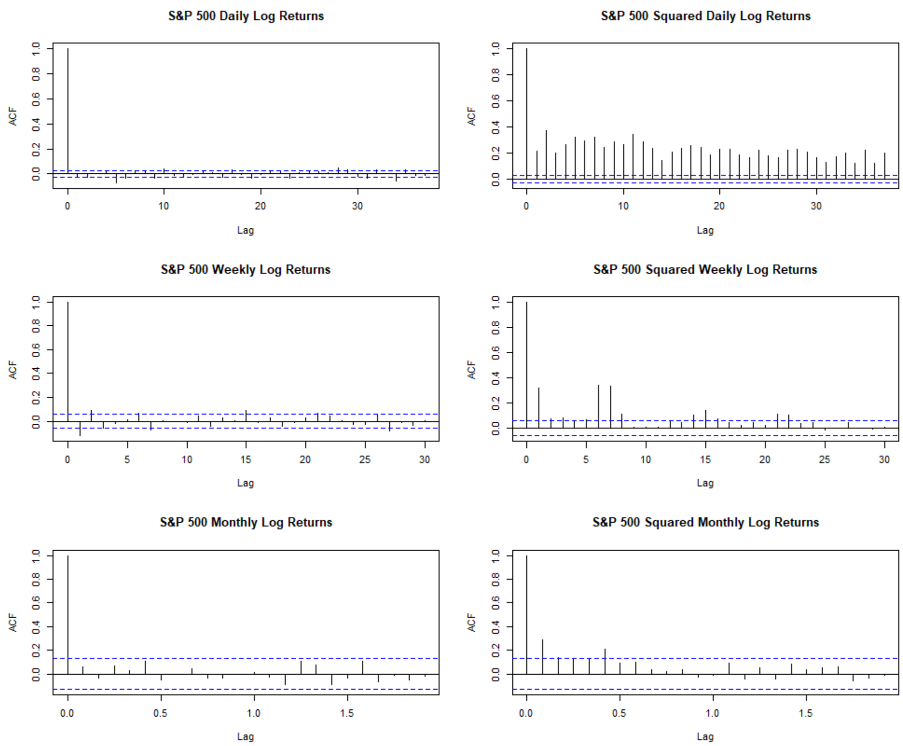

Figure 3.2: ACF for Squared and Squared Returns For the S&P 500

Extreme events in financial returns returns tend to occur in clusters caused by local dependence [6]. Hence, we require a modification of the standard methods for analyzing extreme events. We will try to generalize the results of extreme value theory for iid sequences by permitting a type of weak dependence. In our following analysis, we will assume that the time series $\{r_t\}$ is stationary as defined in Definition 3.1.10 and has a marginal distribution function that is continuous and strictly increasing.

3.1.3 Stationary Returns

As previously stated, in general returns of financial assets are typically not iid [6]. Nonetheless, in general, the statistical properties of financial asset return processes remain stable over time. This time invariance in statistical properties corresponds with the *stationary hypothesis* [8].

Definition 3.1.10 (Stationary). *We will say that the time series* $\{X_i\}$ *is* **stationary** *(sometimes referred to as weakly stationary) if*

(i) $\mathbb{E}(X_i) = \mu$ *is independent of* i*,*

(ii) $Cov(X_{i+h}, X_i)$ *is independent of* i *for each* $h \in \mathbb{Z}$*,*

We will say that $\{X_i\}$ *is* **strictly stationary** *(also referred to as strongly stationary) if* $(X_{i_1}, \cdots, X_{i_j}) \sim (X_{i_1+h}, \cdots, X_{i_j+h})$ *for each* $j \in \mathbb{N}$ *and all* $i_1, ..., i_j, h \in \mathbb{Z}$

Hence, stationarity corresponds to a series whose variables may be mutually dependent, but whose stochastic properties are homogeneous through time [30].

It follows from the above definition that an iid process is strictly stationary. Furthermore, if the process is strictly stationary *and* has finite second moment, then such a process is also stationary [30].

Proof: Suppose we have a strictly stationary process $\{X_i\}_{i\in\mathbb{Z}}$ as defined in Definition 3.1.10 and such a process has *finite second moment*. Then,

(i) $\ldots, X_{-1}, X_0, X_1, \ldots$ have the same distribution function. For some fixed x,

$$f_{X_j}(x) = f_{X_k}(x) \text{ for any } j, k \in \mathbb{Z}.$$

Hence, $\mathbb{E}(X_i) = \mu \quad \forall i \in \mathbb{Z}$.

(ii) (X_{i_1}, X_{i_2}) and (X_{i_1+h}, X_{i_2+h}) have the same *joint distribution function* for any i_1, i_2, h. For some fixed x_1, x_2,

$$f_{X_{i_1}, X_{i_2}}(x_1, x_2) = f_{X_{i_1+h}, X_{i_2+h}}(x_1, x_2) \text{ for any } i_1, i_2, h \in \mathbb{Z}.$$

Hence, $Cov(X_{i_1+h}, X_{i_2+h}) = Cov(X_{i_1}, X_{i_2}) \quad \forall i_1, i_2, h \in \mathbb{Z}$

$\blacksquare$

If we can extend the preceding exposition of extreme value theory to strictly stationary sequences, then we can extend it to stationary sequences given that a stationary sequence has finite variance. If a strictly stationary sequence is close to independent for large enough time separation, this extension follows naturally.

3.1.4 Maxima of Strictly Stationary Returns

In this section, we assume that $\{r_t\}_{t\in\mathbb{N}}$ is *strictly stationary*. We shall further assume that the dependence between r_j and r_k falls off in some specified way as $\left|j-k\right|$ increases.

The following condition was given by Leadbetter and Rootzén [19],

Condition 3.1.11 ($D(u_n)$)**.** *A strictly stationary series $\{r_t\}$ is said to satisfy the $D(u_n)$ condition if, for all $i_1 < \cdots i_p < j_1 \cdots < j_q$ with $j_1 - i_p > l$,*

$$\begin{aligned}\Big|&\Pr\left[r_{i_1} \leq u_n, \cdots, r_{i_p} \leq u_n, r_{j_1} \leq u_n, \cdots, r_{j_q} \leq u_n\right] \\ &- \Pr\left[r_{i_1} \leq u_n, \cdots, r_{i_p} \leq u_n\right] \Pr\left[r_{j_1} \leq u_n, \cdots, r_{j_q} \leq u_n\right]\Big| \leq \alpha(n,l)\end{aligned}$$

Where $\alpha(n,l_n) \to 0$ for some sequence l_n such that $\dfrac{l_n}{n} \to 0$ as $n \to \infty$

Under the $D(u_n)$ condition, for some threshold sequence of u_n that increases with n, the difference of probabilities in Condition 3.1.11 is sufficiently close to zero and have little effect on the limit laws for extremes. Since correlation is also a measure of the independence of two random variables, Condition 3.1.11 states that $\rho(h) \to 0$ as $h \to \infty$ for any non-linear function of r_t and r_{t+h} (see Fig. 3.3). In other words, Condition 3.1.11 states that two events sufficiently far away become asymptotically independent as n increases.

Theorem 3.1.12. *Let $\{r_t\}$ be a* strictly stationary process *for which there exists a sequence of pairs of real numbers $(\alpha_n > 0, \beta_n)$ such that*

$$\Pr\left[\frac{r_{(n)} - \beta_n}{\alpha_n} \leq x\right] \to G(x) \quad \text{as } n \to \infty$$

converges for some non-degenerate distribution function, G and the $D(u_n)$ condition is satisfied for $u_n := \alpha_n x + \beta_n$ for every real x. Then G is a member of the generalized extreme value family of distributions as defined in Theorem 3.1.6.

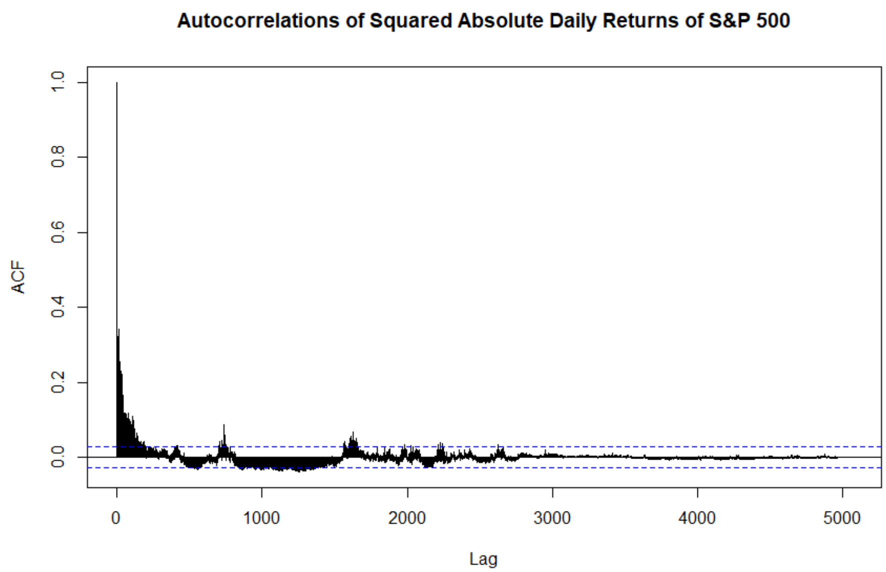

Figure 3.3: Squared Daily Returns For the S&P 500

Hence, by Theorem 3.1.12, a strictly stationary series satisfying Condition 3.1.11 has the same asymptotic distribution as an iid series. In other words, provided that the long-range dependence of a series is very small, the maxima of such a series follows the same distributional limit as described in Theorem 3.1.6 [7]. This *does not* however state that the parameters of the limiting distribution (α_n, β_n) are the same.

Thus far, we concerned ourselves with *possible* forms of the limiting distribution of maxima of strictly stationary series. Now we concern ourselves with whether such a limit exists. It will be seen that the classical criterion for Maximum Domains of Attraction may be applied under Condition 3.1.11 for strictly stationary series.

It is convenient at this point to introduce the concept of an associated sequence of random variables.

Definition 3.1.13 (Associated Sequence of Random Variables)**.** *Let $\{r_t\}$ be a strictly stationary series with marginal distribution F. We defined an* associated sequence *of independent random variables, $\{\epsilon_t\}$ such that each ϵ_t also has distribution function F.*

We will compare the maxima of a strictly stationary series $\{r_t\}$ to its associated iid series ϵ_t. Since the marginal distributions of $\{r_t\}$ and $\{\epsilon_t\}$ are the same, any difference in the limiting distribution of maxima must be attributable to dependence in the series $\{r_t\}$

Definition 3.1.14 (Extremal Index)**.** *Let $\{r_t\}$ be a strictly stationary time series with marginal distribution function F and θ a non-negative number. Suppose for every $\tau > 0$ there exists a sequence $u_n(\tau)$ such that the following hold*

(i) $n[1 - F(u_n(\tau))] \to \tau$,

(ii) $\Pr[r_{(n)} \leq u_n(\tau)] \to e^{-\theta\tau}$, *as* $n \to \infty$

Then θ is called the extremal index *of the time series $\{r_t\}$*

It can be shown that $0 < \theta \leq 1$ is well defined and *not* dependent on the particular choice of sequence $u_n(\tau)$. Furthermore, the extremal index can be seen as a measure of the effect of dependence on the maxima.

Proposition 3.1.15. [7]*Denote $\overline{F}(x) = \Pr(X > x)$. Let $\{r_t\}$ and $\{\epsilon_t\}$ be sequences as defined in Definition 3.1.13. By Definition 3.1.14, we have that for $\tau > 0$ and some general sequence u_n that the following are equivalent*

(i) $n\overline{F}(u_n) \to \tau$,

(ii) $\Pr[\epsilon_{(n)} \leq u_n] = [F(u_n)]^n = \left(1 - \frac{n\overline{F}(u_n)}{n}\right)^n \to e^{-\tau}$,

(iii) $\Pr[r_{(n)} \leq u_n] \to e^{-\theta\tau}$

as $n \to \infty$

Thus, Proposition 3.1.15 implies that for large n, $\Pr\left[r_{(n)} \leq u_n\right] \approx \left(\Pr\left[\epsilon_{(n)} \leq u_n\right]\right)^\theta = F^{n\theta}(u_n)$, provided Condition 3.1.11 holds.

Proposition 3.1.15 for strictly stationary series is comparable to Proposition 3.1.3 for iid series. Whereas the introduction of dependence into a sequence can significantly affect various extremal properties, it does not affect the type of asymptotic distribution of the maxima (See Theorem 3.1.17). It is clear that any iid sequence for which $u_n(\tau)$ may be chosen satisfying Proposition 3.1.3 has extremal index $\theta = 1$. This is obvious as $\{r_i\}$ and $\{\epsilon_i\}$ will comprise the same series. However, the converse is not true. That is one can have strictly stationary series with $\theta = 1$ which are *not* iid [7]. In other words, a series for which $\theta = 1$ means that dependence is negligible at *high* levels, but *not* at the extreme levels that are relevant to a particular application [7].

[7]For a proof, see Extremes and Related Properties of Random Sequences and Processes by Leadbetter, Lindgren and Rootzén [20]

Example 3.1.16 (Coles, Bawa, Trenner and Dorazio)**.** *Let* $\{Y_i\}_{i\in\mathbb{N}}$ *be a independent sequence of random variables with distribution function*

$$F_Y(y) = e^{-\frac{1}{(a+1)y}}, \quad y > 0,$$

where $0 \leq a \leq 1$ *is a parameter. Define the* $\{X_i\}_{i\in\mathbb{N}}$ *as follows*

$$X_0 = Y_0, \quad X_i = max\{aY_{i-1}, Y_i\}, \quad i = 1, \ldots, n$$

For each $i = 1, \ldots, n$,

$$\Pr[X_i \leq x] = \Pr[aY_{i-1} \leq x, Y_i \leq x] = e^{-\frac{1}{x}},$$

provided $x > 0$. *Hence, the marginal distribution of* $\{X_i\}$ *for* $i = 1, 2, \ldots$ *is a standard Fréchet distribution. It can be further shown that* $\{X_i\}_{i=1,2,\ldots}$ *is strictly stationary.*

Suppose there is another sequence of random variables say, $\{X_i^\star\}_{i=1,2,\ldots}$ *that is iid and has the same marginal standard Fréchet distribution as* $\{X_i\}$ *, then*

$$\Pr\left[X^\star_{(n)} \leq nx\right] = \left[e^{-\frac{1}{nx}}\right]^n = e^{-\frac{1}{x}}$$

by applying Eq. (3.1)*. On the other hand*

$$\begin{aligned}
\Pr\left[X_{(n)} \leq nx\right] &= \Pr[X_1 \leq nx, \ldots, X_n \leq nx] \\
&= \Pr[Y_1 \leq nx, aY_1 \leq nx \ldots, aY_{n-1} \leq nx, aY_n \leq nx] \\
&= \Pr[Y_1 \leq nx, \ldots, Y_n \leq nx] \quad \textit{when } 0 \leq a \leq 1 \\
&= \left[e^{\left(-\frac{1}{(a+1)nx}\right)}\right]^n \\
&= \left[e^{\left(-\frac{1}{x}\right)}\right]^{\frac{1}{a+1}}
\end{aligned}$$

In particular, we have $\Pr\left[X_{(n)} \leq nz\right] = \left(\Pr\left[X^\star_{(n)} \leq nz\right]\right)^{\frac{1}{a+1}}$. *Hence the extremal index value* $\theta = \frac{1}{a+1}$.

Theorem 3.1.17 (Distribution of The Maxima of Strictly Stationary Time Series)**.** *Let* $\{r_t\}_{t\in\mathbb{N}}$ *denote a strictly stationary time series with common distribution* F *under Condition 3.1.11 and let* ϵ_t *represent the associated iid process as definition in Definition 3.1.13. Then for some extremal index* $\theta \in \mathbb{R}, 0 < \theta \leq 1$

$$\Pr\left[\frac{\epsilon_{(n)} - \beta_n}{\alpha_n} \leq x\right] \to G(x) \quad \textit{for a non-degenerate limit } G(x)$$

$$\Longleftrightarrow$$

$$\Pr\left[\frac{r_{(n)} - \beta_n}{\alpha_n} \leq x\right] \to G^\theta(x) \quad \textit{for a non-degenerate limit } G^\theta(x) \textit{ is also non-degenerate}$$

as $n \to \infty$

Theorems 3.1.12 and 3.1.17 imply that under Condition 3.1.11the distribution of the maxima of a stationary series converges *and* that limiting distribution is related to the limiting distribution of iid series with the same marginal distribution as the strictly stationary series by the extremal index. Furthermore, from Theorem 3.1.6, the limiting distribution G is said to be a Generalized Extreme Value distribution G_ξ. It can be easily verified that for any GEV distribution G_ξ its power G_ξ^θ is also a GEV distribution with the *same shape parameter* ξ. That is, Theorem 3.1.17 states that the asymptotic distribution of the maxima of the associated iid process is a GEV distribution if and only if the maxima of the strictly stationary time series is asymptotically distributed according to a GEV with the *same shape parameter* ξ.

Thus, for large enough n , the above implies that

$$\Pr\left[r_{(n)} \leq \alpha_n x + \beta_n\right] \approx \Pr^\theta\left[\epsilon_{(n)} \leq \alpha_n x + \beta_n\right] = F^{n\theta}(\alpha_n x + \beta_n) \tag{3.2}$$

provided that Proposition 3.1.15 holds.

By the way of comment, it is trivially true that $G_\xi^\theta = G_\xi$ if $\theta = 1$. For cases where $\theta = 1$ there is no tendency to cluster at high levels and large sample maxima from the time series behave exactly like maxima from similarly sized iid samples. In other words, the limiting distribution of $r_{(n)}$ and $\epsilon_{(n)}$ are the same. For cases where $\theta < 1$ extreme values tend to cluster. Indeed for $0 < \theta < 1$, the limiting distribution may also be taken to be the same by a simple change of normalizing constants [19]. We note that ARCH and GARCH processes have $\theta < 1$ [24].

The above Eq. (3.2) *does not* say that every strictly stationary process has an extremal index. However, for time series under Condition 3.1.11, such as the financial asset returns that interest us, an extremal index generally exists [24].

The following derivation for $\xi \neq 0$ was given by Tsay [28],

Proof:[Tsay] Assume that for an strictly stationary sequence $\{r_i\}$ under Condition 3.1.11 with marginal distribution function F has associated iid sequence $\{\epsilon_i\}$ such that $\epsilon_{(n)}$ has limiting distribution with parameters ξ, α, β. Then by Theorem 3.1.17, we have

$$\Pr\left[r_{(n)} \leq \alpha_n x + \beta_n\right] = \Pr^{\theta}\left[\epsilon_{(n)} \leq \alpha_n x + \beta_n\right]$$

by Theorem 3.1.6, we have

$$\begin{aligned}
&= \exp\left[-\theta\left(1+\xi\frac{x-\beta}{\alpha}\right)^{-\frac{1}{\xi}}\right] \\
&= \exp\left[-\left(\frac{1}{\theta^{\xi}}+\xi\frac{x-\beta}{\alpha\theta^{\xi}}\right)^{-\frac{1}{\xi}}\right] \\
&= \exp\left[-\left(\xi\frac{\frac{\alpha}{\xi}-x-\beta}{\alpha\theta^{\xi}}\right)^{-\frac{1}{\xi}}\right] \\
&= \exp\left[-\left(1+\xi\frac{x-\beta+\frac{\alpha}{\xi}-\frac{\alpha\theta^{\xi}}{\xi}}{\alpha\theta^{\xi}}\right)^{-\frac{1}{\xi}}\right] \\
&= \exp\left[-\left(1+\xi\frac{x-\left\{\beta-\frac{\alpha}{\xi}\left(1-\theta^{\xi}\right)\right\}}{\alpha\theta^{\xi}}\right)^{-\frac{1}{\xi}}\right] \\
&= \exp\left[-\left(1+\xi\frac{x-b}{a}\right)^{-\frac{1}{\xi}}\right] \qquad (3.3)
\end{aligned}$$

■

Hence, a strictly stationary series $\{r_t\}$ under Condition 3.1.11, if it has limiting distribution with the *shape parameter* ξ, has the same shape parameter ξ as its associated iid random variables $\{\epsilon_t\}$. Furthermore, such limiting distributions are related by $a = \alpha\theta^{\xi}$ and $b = \beta - \alpha\frac{\left(1-\theta^{\xi}\right)}{\xi}$ (See Theorem 3.1.5). A similar result can be shown for $\xi = 0$, where $a = \alpha$ and $b = \beta + \alpha\ln(\theta)$.

The practical implication of Theorem 3.1.17 is that under certain conditions, a series with dependence does not invalidate iid Extreme Value Theory. Furthermore, since θ only alters the pair real sequences $\{\alpha_n > 0, \beta_n\}$ and not the shape parameter ξ, the precise value of θ may be unimportant in finding how fast $\overline{F}(x)$ decays [19].

Chapter 4

Fitting the Shape Parameter ξ

In this chapter, we consider the estimation of the *shape parameter* ξ in the case of a "heavy-tailed" distribution. Equivalently, we place distribution F in the Fréchet domain of attraction (see Theorem 3.1.8). That is we assume that the returns of a financial asset $\{r_t\}$ at any time t, has underlying distribution F that has a tail that decays like a power function, i.e., $\overline{F}(x) = x^{-\frac{1}{\xi}}\mathcal{L}(x)$ for large enough x. Furthermore, we assume that the return series is independent or at the very least is stationary under Condition 3.1.11. For such random variables, the asymptotic distribution of the maxima given a pair of normalizing constants is

$$F^{\theta n}(\alpha x + \beta) \to G_\xi^\theta(x) = \exp\left(-\left(1 + \xi\frac{x-b}{a}\right)^{-\frac{1}{\xi}}\right) \quad \text{as } n \to \infty \text{ and } \xi > 0$$

(See Eq. (3.3)).

Estimation of ξ is a non-trivial problem since the marginal distribution function F of a return series is difficult to estimate beyond observed data.

We discuss a useful definition below.

Definition 4.0.1 (Quantile Function)**.** *Suppose there is a series of realizations $\{X_i\}$ from a random variable X. Then the inverse of the distribution function $F(x) = \Pr[X < x]$ is*

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

and is called the Quantile Function*.*

Although we do not know the underlying distribution of $\{X_i\}$ we can exploit Theorems 3.1.8 and 3.1.17 to impose a Fréchet class distribution ($\xi > 0$) on the asymptotic distribution of the maxima of $\{X_i\}$. Then for sufficiently large x, $\overline{F}(x) = x^{-\frac{1}{\xi}}\mathcal{L}(x)$. Equivalently, we have

$$\ln Q(1-p) = \ln \mathcal{L}^{*}\left(\frac{1}{p}\right) - \xi \ln(p), \tag{4.1}$$

where $\mathcal{L}^{*}(x)$ is some other slowly varying function[1].

The slope of Eq. (4.1) is somehow related to ξ. When considering *only the k largest observations*, a natural estimator for $Q(1-\frac{k}{n+1})$ is the order statistic $X_{(n-k+1)}$.

Hence, for the largest k order statistics, the plot $\left\{\left(\ln\left(\dfrac{n+1}{k}\right), \ln\left(X_{(n-k+1)}\right)\right)\right\}_{k=1,\ldots,n}$ is nearly linear if $\mathcal{L}(x)$ is constant. When $\mathcal{L}(x)$ is not constant, this behaviour is only seen in small enough values of k [2]. I.e., for sufficiently large x, $\mathcal{L}(x)$ converges to a constant. Such a plot is called a Pareto Quantile (or Zipf) Plot (See Fig. 4.1).

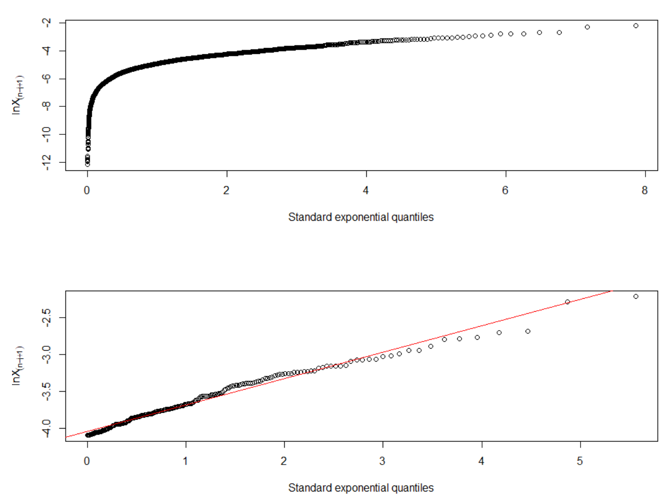

Figure 4.1: Top: The Pareto Quantile Plot for Positive Log Returns of S&P 500 Index. Bottom: The Pareto Quantile Plot for Positive Log Returns of S&P 500 Index for 0.1n largest order statistics.

When we consider the k largest observations, a naive estimator of the slope of Eq. (4.1) is

[1] The two slowly varying $\mathcal{L}(x)$ and $\mathcal{L}^{*}(x)$ functions are linked together via the deBruyn conjugation.

$$\hat{\xi}(k) = \frac{\frac{1}{k}\left(\sum_{i=1}^{k} \ln X_{(n-i+1)}\right) - \ln X_{(n-k+1)}}{\frac{1}{k}\left(\sum_{i=1}^{k} \ln \frac{i}{n+1}\right) - \ln \frac{k}{n+1}} \tag{4.2}$$

A subsample of $k < n$ largest observations in Eq. (4.2) [18] is selected such that the choice of k increases with the overall sample size, n.

This estimator is then refined further to the Hill Estimator by assuming that the denominator converges to 1 for large enough n.

4.1 Using the Hill Estimator to Determine ξ

It is clear that an estimator of ξ will use the largest order statistic $X_{(n)}$. However, it is unclear as to how large k is. If k is too large, then it is unlikely that $\hat{\xi}$ will be accurate, since this includes realizations outside the tail region. However, if k is not large enough, then $\hat{\xi}$ will be distorted by the fact that there are too few observations available to make an estimation. In other words, the right choice of the k largest observations is crucial for a proper estimation of the shape parameter ξ. The right choice of k depends on where the tail of a distribution begins.

Hill proposed the following non-parametric estimator for ξ for heavy-tailed distributions based on a known high threshold, $x > x_0 > 0$, where the denominator of Eq. (4.2) converges to 1 for sufficiently large n.

Definition 4.1.1 (Hill Estimator of ξ)**.** *Let $k = 1, 2, \ldots, n$, then the following is the* Hill estimator *of shape parameter ξ*

$$\hat{\xi}_h(k) = \frac{1}{k}\left(\sum_{i=1}^{k} \ln X_{(n-i+1)}\right) - \ln X_{(n-k+1)} \quad \textit{Hill Estimator} \tag{4.3}$$

If the a time series is iid, the Hill estimator ($\hat{\xi}_h(k)$) is asymptotically normal as follows [18]

$$\sqrt{k}\left(\hat{\xi}_h(k) - \xi\right) \xrightarrow{d} \mathcal{N}(0, \xi).$$

In general, there is no consensus as to the best choice of k. In practice, one may plot the estimator $\hat{\xi}_h(k)$ against k and find a k such that the estimate appears to be stable [28].

This so-called Hill plot proposes a graphical method for choosing an appropriate threshold by identifying the relevant number of upper order statistics. That is, we find a region in the Hill plot where the variance seems to subside. This is referred to as the "Eye-Ball technique" [9]. However, the Hill plot may not be practical for finding the appropriate threshold since the region of stability is not always obvious from the graph (See Fig. 4.2).

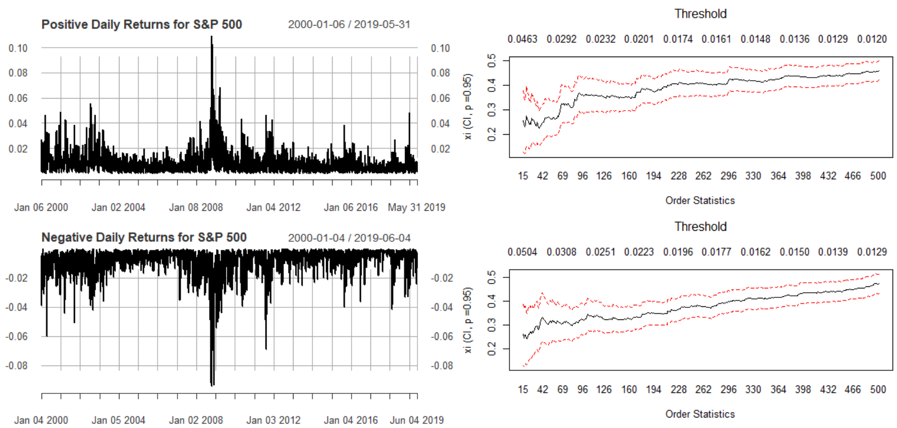

Figure 4.2: The hill plot of the estimators of ξ for the upper(or right) tail and lower (or left) tail for daily returns of S&P 500 for the 500 largest order statistics.

In addition, a severe bias for $\hat{\xi}_h(k)$ can appear in many instances. This is due to the slowly varying function $\mathcal{L}(x)$ converging too slowly. I.e., the assumption that the tails of F follow a strict Pareto distribution, $(\lambda x^{-\alpha})$ is sometimes too optimistic.

Example 4.1.2. *Consider 200000 realizations of a strict Pareto random variable X with parameter 3. Then*

$$\overline{F}(x) = \Pr(X > x) = \lambda x^{-3}.$$

Figs. 4.3 and 4.4 show the typical behaviour of a heavy-tailed for the entire data set as well as the largest 10% of the realizations. The red line is a fitted line using linear regression.

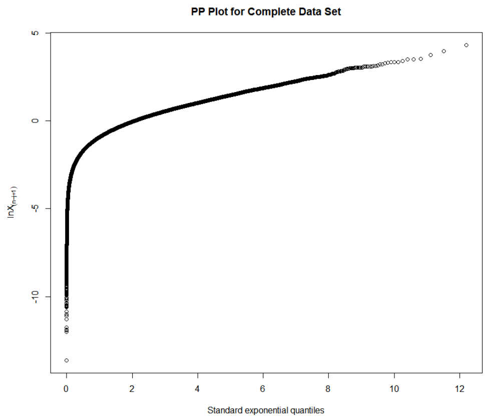

Figure 4.3: PP Plot for Entire Pareto Set

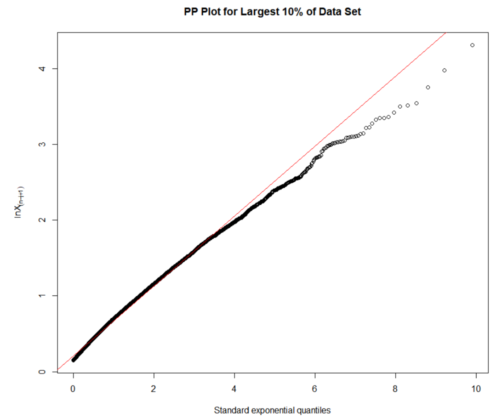

Figure 4.4: PP Plot for Largest 10% of Pareto Set

The Pareto plot of the largest order statistics is nearly linear. We can plot the Hill Estimator $\hat{\xi}_h(k)$ *against* k *to find a region of stability.*

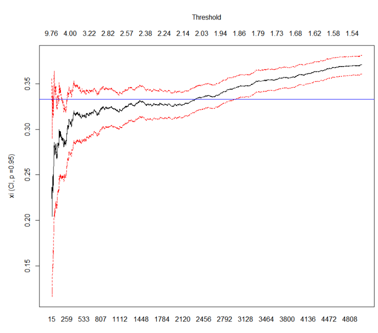

Figure 4.5: Hill Plot of Largest 5000 Order statistics. The blue line is the true value of the shape parameter.

Table 4.1: The Hill estimates for Pareto with parameter 3. Standard errors are in parenthesis.

	Right Tail		
k	2000	3000	4000
X_i	0.3275(0.0073)	0.3469(0.0063)	0.3604(0.0056)

Hence, we can say that $\xi \approx 0.34 \pm 0.014$*. Accordingly, an estimate for the right tail is* $\hat{\overline{F}}(x) = cx^{-2.94}$ *when* x *is sufficiently large and* c *is some unknown constant.*

Example 4.1.2 shows that even in a best case scenario, estimation of shape parameter ξ is difficult. In general, the bias for $\hat{\xi}_h(k)$ grows with the number of the largest observations k [17]. Estimates are much better when k is small [28].

The Hill estimator is only applicable when the underlying distribution F converges to the Fréchet class of GEV distributions ($\xi > 0$) [28]. We use the Hill estimator instead of other estimators since we assume that the distributions of interest (E.g., financial asset returns) are "heavy-tailed". The Hill estimator is more efficient than other estimators [21]. In our analysis, we choose exceedance k between the largest 10% to 20% of the order statistics [18, 3, 9].

4.2 Simulated ARCH Process

Introduced by Engle in 1982, the Autoregressive Conditional Heteroscedasticity(ARCH) models are popular for modelling volatility [30]. It reproduces the same type of volatility changes as are observed in financial asset returns.

Consider a process $\{Y_n, n \geq 1\}$ which satisfies the following stochastic difference equation (SDE), $\{Y_n = A_n Y_{n-1} + B_n\}_{n\geq 1}$, for $Y_n \geq 0$, where $\{(A_n > 0, B_n > 0)\}_{n\geq 1}$ are iid real-valued random pairs.

We simulated 1000 realizations of the following ARCH(1) process,

Definition 4.2.1 (ARCH(1)). *A process $\{\epsilon_i\}$ is called a ARCH(1) process if*

$$\left\{\epsilon_i = X_i \left(\beta + \lambda \epsilon_{i-1}^2\right)^{\frac{1}{2}}\right\}_{i\geq 1}$$

Where $\{X_i\}$ are iid standard normal random variables, $\beta > 0$, $0 < \lambda < 1$.

Thus, Definition 4.2.1 satisfies the previously stated SDE, where $B_i = \lambda X_i^2$ and $A_i = \beta X_i^2$. The tails of such a processes are further discussed in de Haan et al [12]. The tail index parameter, α, is greater than two and is obtained from the equation $\Gamma(\alpha+0.5) = \sqrt{\pi}\,(2\lambda)^{-\alpha}$. Such a process will have at least a finite second moment [21]. For instance, when $\lambda = \dfrac{1}{2}$, the tail index $\alpha = 2.365$ [12].

We simulate the following ARCH(1) process (See Fig. 4.6):

$$\begin{aligned} \epsilon_i &= X_i \sigma_i. \\ \sigma_i^2 &= 1 + 0.5\epsilon_{i-1} \end{aligned} \tag{4.4}$$

Fig. 4.6 plots a 1000 realizations of the ARCH(1) process with $\lambda = 0.5$ and $\beta = 1$. An ARCH(1) process exhibits clustering of extreme values similar to those of financial asset returns. The density of such a process seems to show the characteristic peak near the mean of a leptokurtic (or heavy-tailed) distribution.

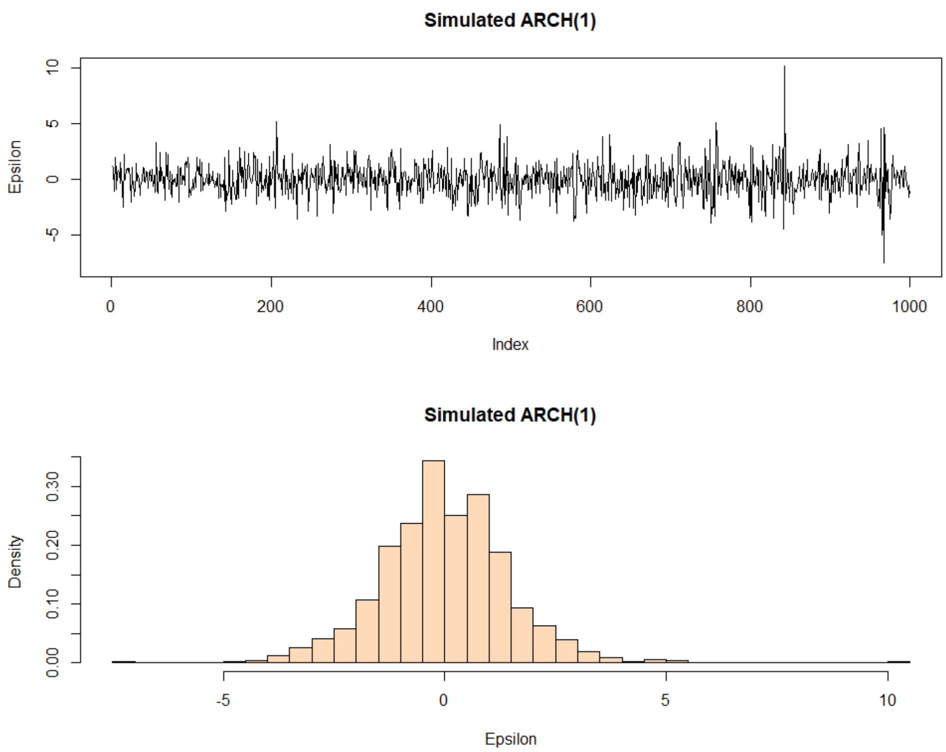

Figure 4.6: A simulated ARCH(1) process with $\lambda = 0.5$ and $\beta = 1$.

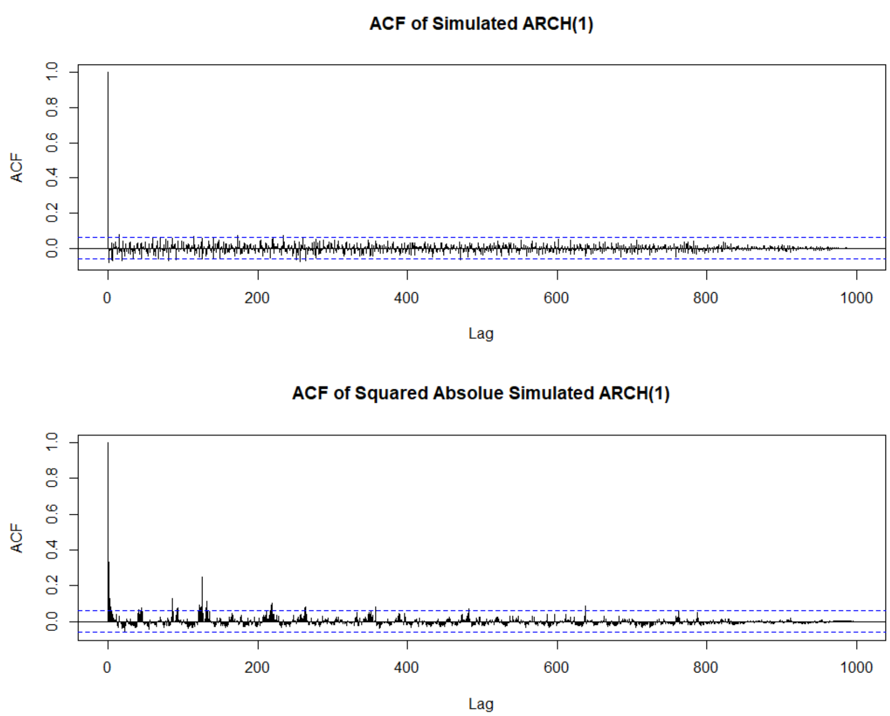

Figure 4.7: ACFs of Simulated ARCH(1) Process

It is known that the above specified ARCH(1) process has a right tail that is modelled by $\overline{F}(x) = \lambda x^{-2.365}$ for large x [12]. Hence, the shape parameter, $\xi = \frac{1}{\alpha} = 0.422$. Such a process is also stationary [12].

Thus by Theorem 3.1.8, we estimate the distribution of the maxima of Eq. (4.4) by assuming such a process has distribution such that $F \in MDA(G_{\xi>0})$. This is a reasonable assumption since the ARCH(1) process is heavy-tailed [12]. For the right tail, we find a stable region around the 150 largest order statistics. Similarly, for the left tail, we find a stable region around the 100 largest order statistics. Table 4.2 summarizes the estimated tail index parameters, ξ using Hill's method. The tail index parameters are different from zero at a significance level of 5%.

Using these estimates, we have the following estimated decays of the tails of F

$$1 - F(\epsilon) \to c_r \epsilon^{-2.27}$$

and

$$1 - F(-\epsilon) \to c_l (-\epsilon)^{-2.79}$$

for large enough ϵ, and c_r and c_l are unknown constants.

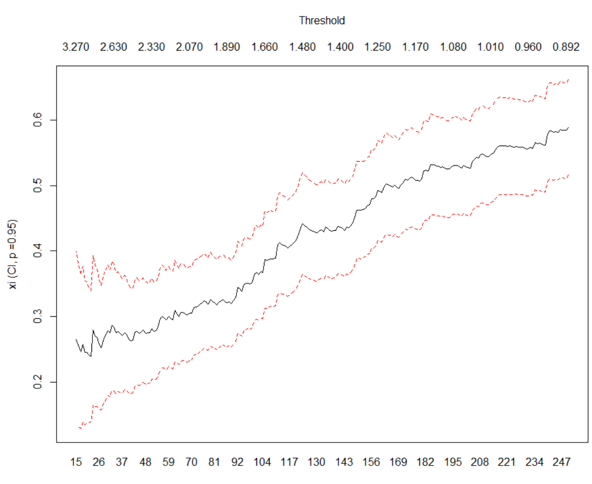

Figure 4.8: Hill Plot for the upper (or right) tail.

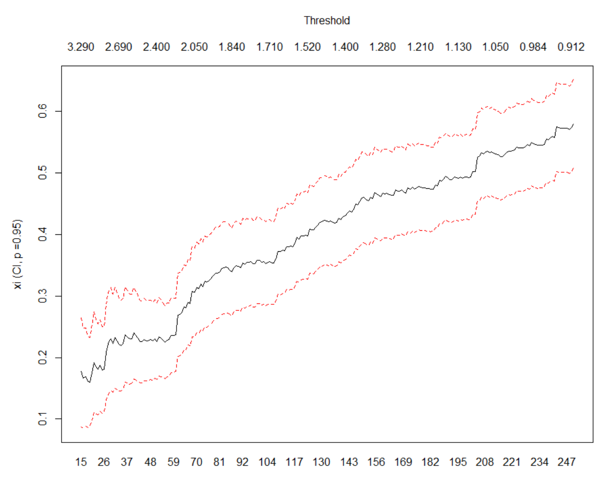

Figure 4.9: Hill Plot for the lower (or left) tail.

According to de Haan et al., the exact constant c for an ARCH(1) process is of little practical importance, since such a constant does not play a role in the rate of decay of the tails [12].

Table 4.2: Results of the Hill estimator for ARCH(1) simulation. Standard errors are in parenthesis.

		Upper (or Right) Tail	
k	135	140	145
ϵ_i	0.434(0.038)	0.440(0.037)	0.465(0.037)
		Lower (or Right) Tail	
k	95	100	105
$-\epsilon_i$	0.349(0.036)	0.358(0.035)	0.357(0.034)

Figs. 4.8 and 4.9 shows the Hill Plots for the largest 250 order statistics for both upper and lower tails. The plots indicate that the tail index parameter, α, appears to be larger for the negative extremes, indicating the ARCH simulation may have a heavier right tail when compared to the left tail. Overall, these results indicate that the ARCH(1) simulations from Eq. (4.4) has distribution belonging to the Fréchet family. Thus confirming the supposition by De Haan, et al. that the ARCH(1) process has heavy tails [12].

Chapter 5

Data Analysis

We collected daily closing prices of a random assortment of stocks and indexes from Bloomberg (See Table 5.1). Using Definition 1.1.1, we obtained three resultant return series $\{r_t\}_{t\in\mathbb{N}}$ for fixed time scales $\Delta t = 1, 5, 25$ for each financial asset. Fig. 5.1 shows the plots of the daily returns from the data set. We can see the typical volatility clustering behaviour in the selected financial asset returns.

The overall aim of the empirical part of this thesis is to examine the tail characteristics of a particular set of indexes and stocks. It includes typical data analysis of examining the descriptive statistics as well as testing the iid hypothesis. A tail index is then computed for all the series chosen for this study. Since it is well established that financial returns show "heavy-tails" [22, 22, 8], the type of general extreme value distribution is known a priori.

Table 5.1: Data Analysed

Data	Daily Symbol	Length
Apple	AAPL	2001-01-03/2019-06-04
Google	GOOG	2014-03-28/2019-06-04
Microsoft	MSFT	2001-01-03/2019-06-04
IBM	IBM	2001-01-03/2019-06-04
GE	GE	2001-01-03/2019-06-04
FTSE	UKX	2001-01-03/2019-06-04
S&P 500	SPX	2001-01-03/2019-06-04
TSX	TSX	2001-01-03/2019-06-04
Dow Jones	DJI	2001-01-03/2019-06-04
DAX	DAX	2001-01-03/2019-06-04

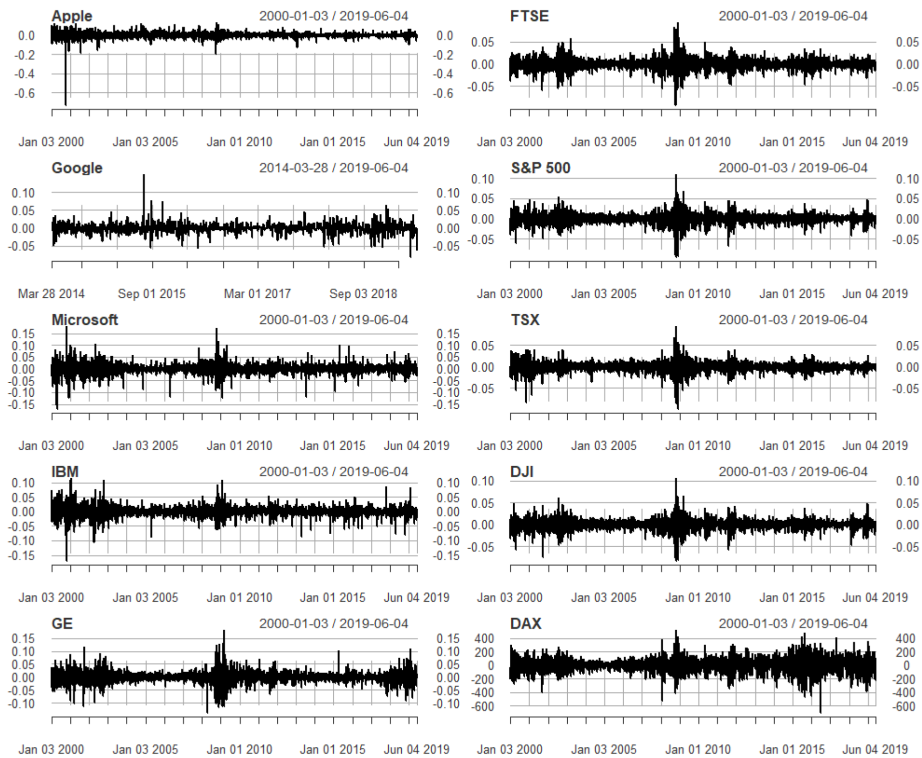

Figure 5.1: The Daily Log Returns of Data Set

5.1 Descriptive Statistics of Financial Return Series

The empirical results from Table 5.2 suggest the underlying distributions a the selected financial time series depart from a Gaussian distribution. Each time series shows the characteristic clustering of extremes with the sample means near zero. In general, a price decrease precedes a price increase.

The descriptive statistics for each financial time series suggest heavy tails and a generally leftward skew. This suggests that the selected time series have longer left tails. We also note that kurtosis excess is typically larger in daily returns and gets smaller as Δt increases. Furthermore, as the time scale increases, the skewness increases. Such behaviour implies that the shape of the respective distributions are not the same at different time scales.

In general, the kurtosis excess for each time series was positive. This suggests that the distribution of all of the series are heavier than a Gaussian distribution. Thus each time

Table 5.2: Descriptive Statistics Log Returns

	Daily		Weekly		Monthly	
Data	$\hat{\tau}$	$\hat{\kappa}$	$\hat{\tau}$	$\hat{\kappa}$	$\hat{\tau}$	$\hat{\kappa}$
Apple	-4.41	122.68	-3.76	55.17	-2.55	17.07
Google	0.51	10.62	-0.44	3.07	-0.68	-0.04
Microsoft	-0.13	10.08	-0.46	4.75	-0.55	3.60
IBM	-0.13	8.48	-0.04	5.27	-0.69	3.79
GE	0.04	8.15	-0.16	12.09	-0.45	1.88
FTSE	-0.16	6.71	-0.61	3.50	-0.70	1.06
S&P 500	-0.22	8.98	-1.17	11.99	-0.92	1.67
TSX	-0.67	10.47	-1.16	8.54	-1.15	3.43
Dow Jones	-0.11	8.66	-1.19	11.09	-0.75	1.05
DAX	-0.23	3.02	-0.41	2.01	-0.58	1.63

series realizes more extreme returns than predicted by a Gaussian distribution.

5.2 Auto-correlation

From the auto-correlation of each time series, we can see that the data set is not iid, but may at least be stationary (See Fig. 3.2). In general, the first order auto-correlation is small and little serial correlation is found at higher lags. In addition, a strong second order auto-correlation is found. Each time series shows weak long range dependence corresponding with Condition 3.1.11. As time scale Δt increases, the nonlinear auto-correlation between lags decreases. This non-iid assumption is further support by completion of the Box-Ljung test. The Box-Ljung tests presented in Table 5.3 show that in general, there is strong evidence against an iid hypothesis.

Table 5.3: The p-values for each data series from the Box-Ljung Tests

	Daily		Weekly		Monthly	
Data	Raw Returns	Squared Returns	Raw Returns	Squared Returns	Raw Returns	Squared Returns
Apple	0.00	0.59	0.00	0.99	0.35	0.94
Google	0.00	0.18	0.18	0.49	0.22	0.00
Microsoft	0.00	0.00	0.01	0.00	0.01	0.85
IBM	0.00	0.00	0.01	0.00	0.00	0.00
GE	0.02	0.00	0.00	0.00	0.02	0.00
FTSE	0.00	0.00	0.00	0.00	0.07	0.00
S&P 500	0.00	0.00	0.00	0.00	0.49	0.00
TSX	0.00	0.00	0.00	0.00	0.41	0.00
Dow Jones	0.00	0.00	0.00	0.00	0.24	0.00
DAX	0.00	0.00	0.67	0.00	0.72	0.00

Since there is strong evidence to reject the iid assumption and the descriptive statistics of each return series indicate that the distributions of the return series F are *not* normal, we can thereby assume that that financial returns show evidence contradicting the geometric Brownian motion on which the Black-Scholes Model is based. Furthermore, we can assume that the analyzed financial returns are heavy-tailed and their tails can be modelled by a Pareto type distribution.

5.3 Asymptotic Behaviour

To estimate the heaviness of the tails of each stock return distribution, we plotted the Hill estimator against the largest order statistics. An estimator of ξ for each stock return was found using the "Eye-ball" technique, where we observed the first regions of stability in the Hill Plots for each time series.

Table 5.4 summarizes the Hill estimates for each time series analyzed. All the Hill estimators were less than 0.5 and positive. This suggests that each time series has a finite second moment.

Table 5.4: The Hill estimates for returns. Standard errors are in parenthesis.

	Daily		Weekly		Monthly	
Data	left tail	right tail	left tail	right tail	left tail	right tail
Apple	0.327(0.021)	0.380(0.021)	0.367(0.043)	0.370(0.045)	0.534(0.133)	0.518(0.097)
Google	0.345(0.043)	0.389(0.038)	0.367(0.080)	0.374(0.074)	0.499(0.150)	0.588(0.163)
Microsoft	0.428(0.023)	0.328(0.028)	0.329(0.025)	0.349(0.021)	0.301(0.024)	0.314(0.018)
IBM	0.458(0.025)	0.387(0.025)	0.360(0.048)	0.369(0.045)	0.570(0.124)	0.441(0.107)
GE	0.343(0.028)	0.384(0.027)	0.429(0.055)	0.507(0.065)	0.507(0.065)	0.410(0.099)
FTSE	0.311(0.026)	0.342(0.022)	0.462(0.060)	0.363(0.054)	0.294(0.071)	0.292(0.068)
S&P 500	0.325(0.027)	0.353(0.028)	0.399(0.055)	0.377(0.040)	0.290(0.072)	0.335(0.068)
TSX	0.336(0.032)	0.392(0.024)	0.454(0.057)	0.361(0.039)	0.331(0.078)	0.309(0.064)
Dow Jones	0.337(0.024)	0.369(0.025)	0.428(0.048)	0.331(0.041)	0.348(0.087)	0.336(0.070)
DAX	0.268(0.021)	0.285(0.019)	0.394(0.049)	0.298(0.038)	0.415(0.097)	0.309(0.069)

Our data showed some marked differences in the distributions of stocks and indexes.

In general, the tails of monthly returns for stocks were heavier as time increment Δt increased. This suggests that extreme returns, both positive and negative were likelier between successive months.

The weekly returns for stocks suggests a heavier right tail. One can expect the probability of extreme positive returns to be greater than the probability of extreme negative returns for between successive weeks.

The left tail parameter of was smaller than the right tail parameter for daily returns of indexes. This indicates that one could expect the probability of extreme positive returns to be higher than extreme negative returns for daily return series. This characteristic was reversed for weekly and monthly return series for indexes. Suggesting the opposite is true for weekly and monthly returns.

Furthermore, we observed that as the time increment Δt increases, the shape parameter for the left tail of indexes increased from daily to weekly and then decreased from weekly to monthly. In general, the left tail shape parameter of weekly indexes was the greatest for each time increment. This suggests that for indexes, extreme negative returns are more likely between successive weeks than successive days or months.

The right tail of indexes were in general heavier for monthly series when compared to the tails of weekly returns. Suggesting the extreme positive returns are more likely during successive monthly when compared to successive weeks.

Table 5.5: The average Hill estimates for stocks and indexes.

	Daily		Weekly		Monthly	
Average	left tail	right tail	left tail	right tail	left tail	right tail
Stocks	0.380	0.373	0.370	0.393	0.482	0.454
indexes	0.315	0.348	0.427	0.346	0.335	0.316

When taking the average of each shape parameter ξ for stocks and indexes with each time increment and tail (See Table 5.5), we found that in general both tails are heavier for stocks than for indexes. This suggests that on average, stocks have a greater probability of extreme realizations when compared to indexes.

Chapter 6

Conclusions and Suggestions for Future Research

Extreme Value Theory provides useful techniques for modelling the distribution of extreme events. We found that the asymptotic distribution of extreme movements in financial asset returns are of the Fréchet class of distributions of the general extreme value distribution. Equivalently, the heaviness of the tail of the distribution of asset returns are related to the parameter of Fréchet class of distributions, ξ.

With regard to estimating the shape parameter using Hill's method, we showed that even in the best case scenario, estimating of ξ is a troublesome using the common "Eye-ball" technique. Regardless, for each data set, the left tail (minima) and the right tail (maxima) were modelled. We found that the shape parameter of an asset return changes as time increment Δt increases. Thus, implying that the respective distributions are different. The data showed that stocks generally have heavier tails than indexes, implying that extreme movements in returns are less likely for indexes when compared to stocks. Such behaviour is of crucial importance to risk managers since the results could be useful in testing economic models of booms and crashes which result from speculative bubbles.

We have consider prices to have weak dependence and assumed that prices between financial assets are independent. However, such an assumptions may not be accurate. More accurate risk estimators that consider stronger dependence and co-movement of asset returns are more useful to risk managers.

It is a Stylized Fact that if an asset return series is modelled by a GARCH(p,q) process, then the distribution of the resulting residuals are heavy-tailed. In future research, we plan on modelling the extreme realizations of these residuals using Extreme Value Theory and investigating how modelling such a behaviour can be used to refine risk indicators.

References

[1] Jushan Bai and Serena Ng. Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23(1):49–60, 2005.

[2] J. Beirlant, G. Dierckx, and A. Guillou. Estimation of the extreme-value index and generalized quantile plots. *Bernoulli*, 11(6):949–970, 2005.

[3] Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef L Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[4] Peter J Brockwell, Richard A Davis, and Matthew V Calder. *Introduction to time series and forecasting*, volume 2. Springer, 2002.

[5] Nicolas Champagnat, Madalina Deaconu, Antoine Lejay, Nicolas Navet, and Souhail Boukherouaa. An empirical analysis of heavy-tails behavior of financial data: The case for power laws. 2013.

[6] Myriam Charras-Garrido and Pascal Lezaud. Extreme value analysis: an introduction. *Journal de la Société Française de Statistique*, 154(2):pp–66, 2013.

[7] Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

[8] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. 2001.

[9] Jon Danielsson, Lerby Murat Ergun, Laurens de Haan, and Casper G de Vries. Tail index estimation: Quantile driven threshold selection. *Available at SSRN 2717478*, 2016.

[10] Richard A Davis and Thomas Mikosch. Extreme value theory for garch processes. In *Handbook of financial time series*, pages 187–200. Springer, 2009.

[11] Laurens De Haan and Ana Ferreira. *Extreme value theory: an introduction*. Springer Science & Business Media, 2007.

[12] Laurens De Haan, Sidney Resnick, Holger Rootzén, and Casper de Vries. Extremal behavior of solutions to a stochastic difference equation, with applications to arch processes. *Stochastic Processes and their Applications*, pages 213–224, 1989.

[13] Bruno De Sousa and George Michailidis. A diagnostic plot for estimating the tail index of a distribution. *Journal of Computational and Graphical Statistics*, 13(4):974–995, 2004.

[14] Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.

[15] Jürgen Franke, Wolfgang Karl Härdle, and Christian M Hafner. *Statistics of financial markets*, volume 2. Springer, 2004.

[16] BV Gnedenko. On the limiting distribution of the maximum term in a random series. In *Breakthroughs in statistics*, pages 195–225. Springer, 1992.

[17] Mia Hubert, Goedele Dierckx, and Dina Vanpaemel. Detecting influential data points for the hill estimator in pareto-type distributions. *Computational Statistics & Data Analysis*, 65:13–28, 2013.

[18] Phillip Kearns and Adrian Pagan. Estimating the density tail index for financial time series. *The Review of Economics and Statistics*, 79(2):171–175, 1997.

[19] M. R. Leadbetter and Holger Rootzé. Extremal theory for stochastic processes. *The Annals of Probability*, 16(2):431–478, 1988.

[20] Malcolm R Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Science & Business Media, 2012.

[21] François M Longin. The asymptotic distribution of extreme stock market returns. *Journal of business*, pages 383–408, 1996.

[22] Benoit Mandelbrot. The pareto-levy law and the distribution of income. *International Economic Review*, 1(2):79–106, 1960.

[23] Benoit Mandelbrot. New methods in statistical economics. *Journal of political economy*, 71(5):421–440, 1963.

[24] Alexander J McNeil. Calculating quantile risk measures for financial return series using extreme value theory. Technical report, ETH Zurich, 1998.

[25] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition*. Princeton university press, 2015.

[26] Athanasios Papoulis. *Random Variables and Stochastic Processes*. McGraw Hill, 1976.

[27] Rolf-Dieter Reiss. *Approximate distributions of order statistics: with applications to nonparametric statistics*. Springer science & business media, 2012.

[28] Ruey S Tsay. Extreme values and their applications in finance, 2007.

[29] Paul T Von Hippel. Mean, median, and skew: Correcting a textbook rule. *Journal of Statistics Education*, 13(2), 2005.

[30] Foivos Xanthos. Am8205: Applied statistical methods. 2019.