ΕΛΛΗΝΙΚΟ ΑΝΟΙΚΤΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΜΕΤΑΠΤΥΧΙΑΚΕΣ ΣΠΟΥΔΕΣ ΣΤΑ ΜΑΘΗΜΑΤΙΚΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Θεωρία Ακραίων Τιμών και Κατανομές με Βαριές Ουρές: Εφαρμογή στον Υπολογισμό Ακραίων Αποδόσεων Περιουσιακών Στοιχείων»

«Extreme Value Theory and Heavy Tailed Distributions: Application to calculate extreme asset returns»

ΟΝΟΜΑ ΦΟΙΤΗΤΡΙΑΣ: ΠΑΠΑΛΕΞΑΝΔΡΟΠΟΥΛΟΥ ΕΛΕΝΗ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ Α: ΜΙΧΑΛΗΣ ΑΝΟΥΣΗΣ
ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ Β: ΚΩΝΣΤΑΝΤΙΝΟΣ ΠΟΛΙΤΗΣ

ΑΘΗΝΑ, ΜΑΪΟΣ 2023

Abstract

Extreme value theory (EVT) aims to predict occurrence of rare events and is concerned with the study of the asymptotical distribution of these extreme events. To model extreme observations, we use two different distribution families the Generalized Extreme Value (GEV) and the Generalized Pareto Distribution (GPD).

The first is based on the extreme value distribution of the Gumbel (thin-tailed), Fréchet (fat-tailed) or Weibull (finite upper bound) distributions which are generalized as the generalized extreme value distribution (GEV). The second is based on the generalized Pareto distribution (GPD) and is known as the peak-over-threshold (POT) approach.

In this thesis, EVT methods were used to investigate and fit the empirical distribution of the weekly and monthly maximum and minimum return series of the ASE, and BEGCGA indices to the theoretical GEV and GPD distributions, to determine whether returns follow a heavy-tailed stable distribution.

We have applied two approaches of extreme value theory, the Block Maxima and the Peaks Over Threshold (POT) approach, as well as the parametric approach of the Maximum Likelihood Estimate Method (MLE) for the distribution parameter estimation.

Keywords

Extreme value theory, block maxima, generalize extreme value distribution, generalized Pareto distribution, peaks over threshold, financial time series.

Περίληψη

Η Θεωρία των Ακραίων Τιμών στοχεύει να προβλέψει την εμφάνιση σπάνιων γεγονότων και ασχολείται με τη μελέτη της ασυμπτωτικής κατανομής αυτών των ακραίων γεγονότων. Για να μοντελοποιήσουμε ακραίες παρατηρήσεις, χρησιμοποιούμε δύο διαφορετικές οικογένειες κατανομών, την κατανομή της Γενικευμένης Ακραίας Τιμής (GEV) και τη Γενικευμένη κατά Pareto Κατανομή (GPD).

Η πρώτη βασίζεται στην κατανομή ακραίων τιμών όπως η κατανομή Gumbel (λεπτή ουρά), η κατανομή Fréchet (παχιά ουρά) ή η κατανομή Weibull (πεπερασμένο άνω όριο) που γενικεύονται ως η γενικευμένη κατανομή ακραίων τιμών (GEV). Η δεύτερη βασίζεται στη γενικευμένη κατανομή Pareto (GPD) και είναι γνωστή ως «προσέγγιση κορυφής» πάνω από το όριο (POT).

Σε αυτή τη θέση, χρησιμοποιήθηκαν μέθοδοι από τη Θεωρία Ακραίων Τιμών για τη διερεύνηση και την προσαρμογή της εμπειρικής κατανομής της εβδομαδιαίας και μηνιαίας μέγιστης και ελάχιστης απόδοσης τόσο του Μετοχικού Δείκτη του ΧΑΑ, όσο και του ομολογιακού BEGCGA στις θεωρητικές κατανομές GEV και GPD.

Εφαρμόσαμε δύο προσεγγίσεις της θεωρίας ακραίων τιμών, την προσέγγιση Block Maxima και την προσέγγιση Peaks Over Threshold (POT), καθώς και την παραμετρική προσέγγιση της μεθόδου εκτίμησης μέγιστης πιθανότητας (MLE) για την εκτίμηση των παραμέτρων της κατανομής.

Λέξεις - Κλειδιά

Θεωρία ακραίων τιμών, block maxima, γενικευμένη κατανομή ακραίων τιμών, generalized Pareto distribution γενικευμένη κατά Pareto κατανομή, peaks over threshold, χρηματοοικονομικές χρονοσειρές.

Contents

Commonly used notations

μ – Location parameter

σ – Scale parameter

ξ – Shape parameter

α - Tail index

u - Threshold

i.i.d. – independent and identically distributed

$\Phi_\alpha(x)$ - Fréchet distribution

$\Psi_\alpha(x)$ - Weibull distribution

$\Lambda_\alpha(x)$ - Gumbel distribution

$H_\xi(x)$ - Generalized Extreme Value distribution

$G_\xi(x)$ - Generalized Pareto distribution

Abbreviations

BM – Block Maxima

EDF – Empirical Distribution Function

EVT – Extreme Value Theory

POT - Peaks Over Threshold

GEV – Generalized Extreme Value

GPD – Generalized Pareto Distribution

MLE – Maximum Likelihood Estimate

ASE – Athens Stock Exchange General Index

BEGCGA - Bloomberg Series-E Greece Govt All > 1Yr Bond Index

List of Tables

List of Figures

Introduction

Extreme Value Theory (EVT) is a well-developed theory in the field of probability, that studies the distribution of extreme realizations of a given distribution function, or of a stochastic process, satisfying suitable assumptions. The foundations of the theory were set by Fisher and Tippett (1928) and Gnedenko (1943), who proved that the distribution of the extreme values of an independent and identically distributed (i.i.d.) sample from a cumulative distribution function F, when adequately rescaled, can converge (and indeed does converge, for the majority of known cumulative distribution functions F) to one out of only three possible distributions.

Extreme Value Theory (EVT) originated from astronomy and the need to keep or reject outliers in data (Kotz & Nadarajah, 2003). Outliers are stragglers — extremely high or extremely low values — in a data set that can throw off statistics. EVT has developed into a theory that is applicable to almost every area of science and business. For example, the theory can model and predict a diverse range of phenomena such as the maximum heights of ocean waves or the strength of financial markets (Coles, 2013). The theory, which uses extreme value distributions, is widely used in economics, finance, materials science, reliability engineering and many other fields.

Kinnison (1985) credits Fourier in 1824 for the oldest remarks in statistical literature about extreme values. Fourier stated that, for the Gaussian distribution, the probability of a deviation being more than three times the square root of two standard deviations from the mean is about 1 in 50,000, and consequently the observation associated with this deviation could be neglected. This seems to be the origin of the common but erroneous statistical "rule" that plus or minus three standard deviations from the mean should be considered the maximum range of valid samples from a Gaussian distribution irrespective of the number of samples taken.

The identification of limiting distributions for maxima is attributable to Fisher & Tippett (1928), whose work was further developed by von Mises (1936), Gnedenko (1943), and many others. Slightly earlier, Fréchet (1927) had identified a functional equation, which he called the *stability postulate*, that provides a mathematical basis for extrapolation and thus lies at the heart of the classical theory of extremes. His stability postulate is now referred to as *max stability*. An early advocate for the application of extreme value theory to engineering problems was E.J. Gumbel, whose 1958 book remained essentially the only text on the topic until the appearance of one by Galambos (1978). Since then, the number of books on extreme value theory has multiplied: Resnick (1987) presents an important early treatment of probability models for extremes; Embrechts et al. (1997) provide a careful account oriented toward applications in finance and insurance; Kotz & Nadarajah (2000) survey the literature up to the start of the twenty-first century; Coles (2001) offers an accessible account of modeling statistical extremes that has done much to bring these ideas to a wide readership; Beirlant et al. (2004) provide a broader description of both probabilistic and statistical aspects; and de Haan & Ferreira (2006) and Resnick (2006) provide detailed mathematical treatments from different perspectives.

The application of extreme value theory began in the middle 1930s with the work of E.J. Gumbel, first in Germany and then in the U.S. when World War II engulfed Europe. Gumbel's first application was the consideration of the longest duration of life, or old age (Gumbel, 1935). Then, he showed that the statistical distribution of floods, long studied by engineers, could be understood using extreme value theory (Gumbel, 1941). These procedures have been extensively applied to other areas of science. For example, Tiago de Oliveira (1983) presents the proceedings from the 1983 NATO Advanced Study Institute on Statistical Extremes and Applications to obtain a complete perspective of the field, along with a series of applications.

There has been a growing awareness that some economic data display distributional characteristics that are flatly in- consistent with the hypothesis of normality. Frequency functions have been observed with too much mass near the mean and in the extreme tails, and not enough mass in the intervals between one and two standard deviations (roughly) from the mean. This pattern of behavior holds true of time series and cross-section data alike and has been observed for such widely disparate variables as stock and commodity price changes; sales, employment, or asset size measures of business firms; and personal incomes.

Tail fatness is a well-documented feature of most financial time series, since the pioneering studies of Mandelbrot (1963, 1967). Therefore, econometric, and financial models based on an assumption of normality may lead to serious underestimation of the probability of extreme events to occur, i.e., those events that, because of their magnitude (in absolute value) pertain to the tails of the distribution underlying the given time series.

Extremal events in finance have the advantage that they are mostly quantifiable in units of money (as opposed to other extremal occurrences, such as floods and earthquakes, which might cost the loss of lives). Such events include the stock market crashes of 1929 and 1987. Incidents in the derivatives markets include the collapse of Barings Bank and the losses of Metallgesellschaft, Procter and Gamble, Kashima Oil, Orange County (California), and Sumitomo Bank. However, most such events also have a non-quantifiable component. These events can cause a larger shock to the underlying financial system. This is sometimes called *contagion risk*, or the risk that a securities market decline in one country will spread to another, causing serious market losses, or that the failure of one large financial institution will lead to the failure of others. Thus, it is important to study these circumstances.

What is of importance here are the magnitudes of the changes in prices, rather than the average variations. A single, extremely negative return, or a combination of several smaller negative returns, can possibly wipe out so much capital that the firm or portfolio becomes illiquid or insolvent. Trading limits need to be set. These limits are a function of the probability of a single negative return so large that capital is endangered.

If in fact the normal law does not describe economic data, are there compact and simple alternative laws which are descriptive? Research in probability theory has uncovered a class of simple probability distributions, called *stable distributions*, which contain the normal distribution as a special case. The density function of these distributions is all more *leptokurtic* (peaked and thicker tailed) than the normal distribution.

An extreme value approach to the modeling of rare and damaging events quite frequently involves *heavy tailed distributions* associated with *power decaying tails*.

Limit theory for sums, maxima or point processes is closely related to the power law behavior of tails, or of normalizing constants even of characteristic functions in the neighborhood of the origin. Exact power laws mainly occur in the very limit, but in extreme value distributions power laws do not appear in pure form but slightly disturbed by *slowly varying functions*. A power law times a slowly varying function is called *regularly varying*.

Daily or weekly returns, measured by the relative differences of consecutive prices or differences of log-prices, are the appropriate quantities which must be investigated. The special feature of this type of return series is the alternation between periods of tranquility and volatility. Most importantly, there is empirical evidence that distributions of returns can possess heavy (fat) tails so that a careful analysis of returns is required. Fat-tailed distributions exhibit more probability mass in the tails than distributions such as the standard normal distribution. This means that extremely high and low realizations will occur more frequently than under the hypothesis of normality.

In this first part of the paper, we review the foundations of extreme value theory (EVT) as a part of *probability theory* and the main *statistical methods* developed to deal with it.

We focus on extremal events of the form $\{X > x\}$ for some random variable X and large x. We want to estimate tails in their far regions and high quantiles. Asymptotic properties of the tails can be derived, by an analogy to the Central Limit Theorem.

The main result of the EVT is that the distribution of the tails is limited to only three types:

- No tails (df is cut off),
- Exponential function
- Power function

To model extreme observations using Extreme Value Theory (EVT), we can use two different statistical distribution families: Generalized Extreme Value (GEV) and Generalized Pareto Distribution (GPD).

We apply two approaches of extreme value theory, the Block Maxima and the Peaks Over Threshold (POT) approach, as well as the parametric approach of the Maximum Likelihood Estimate Method (MLE) for the distribution parameter estimation and the non-parametric approach of the Hill estimator.

Our aim is to illustrate the tail distribution estimation of two financial series of daily returns. We can say that the goal is to try understanding which theoretical GEV and GPD distributions better fit our data, weekly and monthly maximum, and minimum return series of some financial series in the space of equity and bond local markets.

Theoretical Background

Fisher and Tippett (1928) published the paper that is now considered the foundation of the asymptotic theory of extreme value distributions. The Fisher-Tippett theorem specifies the form of the limit distribution for centered and normalized maxima. Fisher and Tippett argue that if there exists a normalizing constant and a centering constant, then the distribution follows one of the three extreme value distributions.

The three families of possible limit laws are known as *extreme value distributions*. These three distributions have been found adequate to describe the extreme value distributions of all statistical distributions. Fisher and Tippett first derived the *limit laws for maxima*. They were the first to describe the extremely slow convergence of the distribution of the largest value in samples from a Gaussian distribution towards its asymptote. Thus, the use of a Gaussian distribution was inappropriate for extreme value analysis. It has since been found that the exponential distribution is more useful because it leads to simple development and expression of the fundamental theorems. The results can then be generalized to other distributions, particularly those of the exponential family.

The main theorem of EVT is usually compared to the *central limit theorem of probability*. As the central limit theorem states the *convergence of the standardized sum of any sequence of iid random variables* (satisfying some technical assumptions) to a given distribution (the normal), so the results from Fisher, Tippett and Gnedenko guarantee that the *standardized maximum of a sequence of iid random variables* (again, under technical assumptions) converges to some given distribution family. However, while the central limit theorem deals with *sums of random variables*, EVT focuses on *maxima*. Moreover, the central limit theorem provides a unique limit distribution, while EVT includes three different families of asymptotic distributions.

More precisely, given *n* i.i.d. random variables $X_1, \dots, X_n$, let M_n denote the maximum of this collection, i.e.

$$M_n = max\{X_1, \dots, X_n\}.$$

It is known that, if *F* is the cumulative distribution function of X_j, *j* = 1, . . ., *n*, the distribution function of M_n is given by F^n. Since the actual distribution *F* is not known, we are interested in studying the asymptotic distribution of M_n. Anyway, directly taking the limit would yield,

$$\lim_{n \to +\infty} F^n(x) = \begin{cases} 1, & if \quad F(x) = 1 \\ 0, & if \quad F(x) < 1 \end{cases}$$

i.e., a degenerate distribution. A distribution function *G* is non-degenerate if there exists *x* such that $G(x) \in]0, 1[$.

To obtain a non-degenerate limit distribution for the maxima, Fisher and Tippett (1928) *standardized the random variable* M_n by means of an *affine transformation* with a location parameter d_n and a positive scale parameter c_n, thus reducing themselves to study the asymptotic distribution of $\frac{M_n - d_n}{c_n}$.

Assuming the existence of a whole sequence (d_n, c_n) of such parameters (norming constants, for short) indexed by n, Gnedenko (1943) proved that the distribution of the sequence of standardized maxima converges to one of the following three non-degenerate types of distribution:

- *Fréchet*: $H_\xi(x) = \begin{cases} 0, & if\ x \leq 0 \\ exp\{-x^{-1/\xi}\}, & if\ x > 0 \end{cases}$ ξ > 0,
- *Gumbel*: $H_0(x) = exp\{-exp(-x)\},\ x \in \mathbb{R}$,
- *Weibull*: $H_\xi(x) = \begin{cases} exp\{-(-x)^{1/\xi}), & if\ \ x \leq 0 \\ 1, & if\ \ x > 0, \end{cases}$ ξ > 0.

The previous three types of distributions are called *extreme value (EV) distributions*, while *ξ* is called the *shape parameter* and accounts for the behavior of the tail of the distribution. It is now well known that:

- *Fréchet type* is the asymptotic distribution of the extremes of *fat-tailed distributions*, such as stable laws (with *characteristic index* $0 < \alpha < 2$), Student t, log-gamma and Pareto distributions (roughly speaking, distributions with *power-like decaying tails*). Notice that the index *α* is one of the main parameters entering the characteristic function of stable laws.
- *Gumbel family* is the reference class for the extremes of distributions with (essentially) *exponentially decaying tails* with finite moments of any order, like the normal; other distributions in this family are the log-normal, gamma, chi-squared, Gumbel and standard Weibull.
- *Weibull type* is the asymptotic distribution for *short-tailed distributions*, i.e., those with finite right endpoint $x_F = \sup\{x \in \mathbb{R}: F(x) < 1\}$, like the beta.

Notice that the existence of the sequence of norming constants (location and scale parameters) is not always guaranteed.

The asymptotic distribution of extreme values belongs to the family of the *generalized extreme value distributions* which depend on a real parameter, called the *extreme value index* (EVI).

The purpose of the statistical theory of extreme values is to model observed extremes in samples of some specified size, or the number of data points in a related group or set of values. The essential conditions are that:

(1) the phenomenon being measured is a random (stochastic) variable,

(2) the initial distribution from which the samples with extreme values have been drawn remains constant from one set of samples to the next (or that any change that occurs may be measured and a transformation of the data may be found to eliminate the effects of the change), and

(3) the observed extremes are statistically independent.

The distributions of extremes may be characterized by certain statistics, such as means, medians, modes and a statistic called the *expected extreme*. The initial distribution from which the samples containing the extremes are obtained and the sizes of these samples must be known to derive the exact extreme value distribution. However, if the type of the distribution is known and sample sizes are large, asymptotic theory can be used. Asymptotic theory yields simple formulations for statistical tests.

A *general asymptotic theory* for sums of independent and identically distributed variables includes results as the *central limit theorem* and the *law of large numbers*. The α-stable and Brownian motion processes are *continuous-time limits* of appropriate *partial sum processes* and considered random walks in continuous time. Abstract weak convergence techniques are needed in order to prove that partial sum processes converge toward α-stable process or Brownian motion process. Fluctuations of maxima are considered to be classical extreme theory. The results for sums are like the results for partial maxima and order statistics.

Definition and basic properties of Stable Distribution

The word “stable” is due to Levy (1924) and relates to the fact that the convolution of two stable laws is the same law (up to changes of scale and location). *Stable distributions* are the only possible limit laws for sums of independent identically distributed random variables when properly normalized and centered. Thus, by *Central Limit Theorem*, the Gaussian Distributions are stable. It is well known that the convolution of two Gaussian variables is also Gaussian. Moreover, we know that the standard central limit theorem requires a second moment. So, the other stable distributions must appear as limits when the summands do not have a finite variance.

The normal characteristic, and reason for the term stable, is that they retain their shape (up to scale and shift) under addition.

A random variable (a distribution function or “df”) is called *stable* if it satisfies the identity (or symbolically, ($\overset{d}{=}$)) in the law:

$$c_1X_1 + c_2X_2 \overset{d}{=} b(c_1, c_2)X + a(c_1, c_2),$$

for i.i.d. X, X_1, and X_2; for all non-negative numbers c_1, c_2; and appropriate real numbers $b(c_1,c_2) > 0$ and $a(c_1,c_2)$.

This identity shows which distributions are closed (up to changes of location and scale) under convolution and multiplication with real numbers. A general definition of a convolution between two functions may be defined by $f \star g(x) = \int_{-\infty}^{+\infty} f(x-y)g(y)m\{\mathrm{dy}\}$ where m stands for an arbitrary measure.

Definition: A random variable (a distribution, a df) is called *stable* if for each n there exist constants $c_n > 0$, d_n such that

$$S_n = X_1 + X_2 + \ldots + X_n \overset{d}{=} c_nX + d_n, \;\; n \geq 1 \,,$$

and is not concentrated at one point. The distribution is stable in the *strict sense* if the above equation holds with $d_n = 0$. Also, is called symmetric stable if it is stable and symmetrically distributed around 0, i.e., $X \stackrel{d}{=} -X$.

The above equation can be written as

$$c_n^{-1}(S_n - d_n) \stackrel{d}{=} X .$$

We can say that if a distribution is stable then it is a limit distribution for sums of iid random variables.

The symbol $\stackrel{d}{=}$ means equality in distribution, i.e., the right- and left-hand sides have the same distribution. Some authors use the term sum stable to emphasize the stability under addition and to distinguish it from other concepts, e.g., max-stable, min-stable, etc. The normal distributions satisfy this property: the sum of normal terms is normal.

Limit Property of Stable Law

We can conclude that if a distribution is stable, then it is a limit distribution for sums of i.i.d. random variables. The class of stable (non-degenerate) distributions coincides with the class of all possible (non-degenerate) limit laws for (properly normalized and centered) sums of i.i.d. random variables. This is known as the *central limit theorem*, or *convergence in distribution*, or *weak convergence*. Every sequence $\{S_n\}$ can be normalized and centered in such a way that it converges to a given constant (for instance zero) in probability.

The most common way to describe the class of stable distributions analytically is to determine their characteristic functions. By *the spectral representation of a stable law*, we say:

A *stable distribution has characteristic function*:

$$\varphi_X(t) = Eexp\{iXt\} = exp\{i\gamma t - c|t|^{\alpha}(1 - i\beta sign(t)z(t,\alpha))\},\ t \in \mathbb{R},$$

where γ is a real constant, $c > 0$, $\alpha \in (0,2]$, $\beta \in [-1,1]$, and

$$z(t,\alpha) = \begin{cases} tan(\frac{\pi\alpha}{2}) & if\ \ \alpha \neq 1, \\ -\frac{2}{\pi}ln|t| & if\ \alpha = 1. \end{cases}$$

The quantity γ is just a location parameter, which for simplicity we can assume that $\gamma = 0$.

In the case $c = 0$ we have a degenerate distribution.

The most important parameter in this representation is α. It determines the properties of this class of distributions as moments, tails, asymptotic behavior of sums, normalization.

Definition: The number α in the characteristic function is called the *characteristic exponent*, the corresponding distribution *α-stable*.

- For α = 2 we obtain one of the most important distributions in probability theory the *normal* or *Gaussian* distribution: $\varphi_X(t) = exp\{-ct^2\}$, with zero mean and 2*c* variance. We can see that the normal law is determined only by mean and variance, whereas the other α-stable distributions depend on four parameters. This is due to the fact that the normal distribution is symmetric around its expectation. On the contrary, a stable law for $\alpha < 2$ can be asymmetric and even be concentrated on a half-axis, for $\alpha < 1$.
- For α = 1 we have the *Cauchy laws* with characteristic function: $\varphi_X(t) = exp\{-c|t|(1 + i\beta \frac{2}{\pi} sign(t) ln|t|)\}$
- For α =1/2, and β = 1, we obtain the characteristic function of *Lévy distribution*: $\varphi_X(t) = exp\{- |ct|^{1/2}(1 - sign(t))\}$

For fixed α, the parameters *c* and β determine the nature of the distribution. The parameter *c* is a scaling constant which corresponds to $c = \sigma^2 / 2$ in the Gaussian case. The parameter β describes the skewness of the distribution. If $\beta = 1$ and $\alpha < 1$ the random variable *X* is positive and for $\beta = -1$ and $\alpha < 1$ it is negative. In the case $|\beta| < 1$, $\alpha < 1$, or $\alpha \in [1,2]$, the random variable *X* has the whole real axis as support. If $\beta = 1$, $\alpha \in [1, 2)$ then $\Pr(X \leq -x) = o(\Pr(X > x))$ as $x \to \infty$.

Although the α-stable laws are absolutely continuous, their densities can be expressed only by complicated special functions. Only in a few cases as the *Gaussian* ($\alpha = 2$), the *symmetric Cauchy* ($\alpha =1$, $\beta = 0$), and the *stable inverse Gaussian* ($\alpha = ½$, $\beta = 1$) are these densities expressible explicitly via elementary functions. But there exist asymptotic expansions of the α-stable densities in a neighborhood of the origin or of infinity.

When α is close to 2, the tail becomes thin. When moving down from 2 to 1.5, a stable random variable exhibits a heavier tail. For large values of *x*, the pdf of a stable variable behaves like a power law (sometimes called a Pareto tail): $f(x) \simeq \frac{C_{\pm}^{\alpha,\beta}}{|x|^{1+\alpha}}$.

In empirical applications, the estimated or calibrated value of α is usually between 1.3 and 1.9. An empirical study by Kim et al. (2011) determined this range of α for stable and tempered stable distributions.

Stable distributions are a class of probability laws that have intriguing theoretical and practical properties. Their application to financial modeling comes from the fact that they generalize the normal (Gaussian) distribution and allow heavy tails and skewness, which are frequently seen in financial data.

Classical statistics is basically L^2 theory. Hence, for non-Gaussian stable laws, the classical theory breaks down. Infinite variance models are suitable constructions for phenomena that exhibit large fluctuations, big jumps, or wild oscillations.

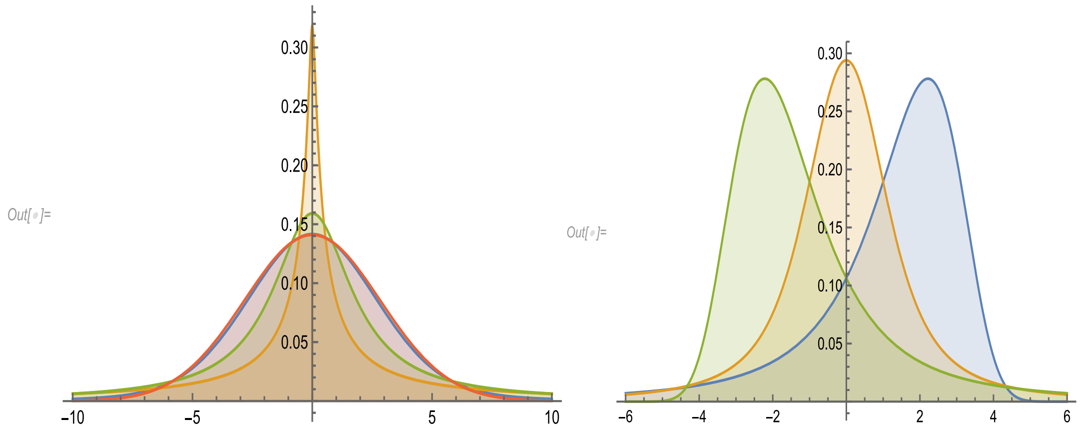

Figure 1. Left Panel: α-stable with α = 1.8, 1.5, 1, 2, β = 0, μ = 0, σ =2. Right Panel: α-stable with α = 1.3, β = -1, 0, 1, μ = 0, σ = 1.

Fluctuations of Maxima

The central result of the classical extreme value theory is the Fisher-Tippett theorem which specifies the form of the limit distribution for centred and normalized maxima.

The asymptotic theories from maxima and sums complement as well as contrast each other. Corresponding results exist for affinely transformed sums and maxima: stable distributions correspond to max-stable distributions, domains of attraction to maximum domains of attraction. Limit theorems for maxima and sums require appropriate normalizing and centering constants. For maxima, the latter are chosen as some quantile (or a closely related quantity) of the underlying *marginal distribution*. Empirical quantiles open the way up for *tail estimation*.

Theory of regular variation plays a fundamental role since the maximum domain of attraction of an extreme value distribution can be characterized via regular variation and its extensions.

Limit Laws for Maxima

We focus on the fluctuations of the *sample maxima*:

$M_1 = X_1, \quad M_n = \max(X_1, \ldots, X_n), \quad n \geq 2,$

with the exact distribution function (df) of the maximum M_n:

$$Pr(M_n \leq x) = Pr(X_1 \leq x, \ldots, X_n \leq x) = F^n(x), \quad x \in \mathbb{R}, \;\; n \in \mathbb{N}.$$

As extremes happen *near the upper end of the support of the distribution*, the asymptotic behavior of M_n must be related to the df F in its right tail near the *right endpoint*, which we denote as:

$$x_F = sup\{x \in \mathbb{R} : \; F(x) < 1\},$$

and for all $x < x_F$,

$$Pr(M_n \leq x) = F^n(x) \to 0, \;\; n \to \infty,$$

And in the case $x_F < \infty$, we have for $x \geq x_F$ that:

$Pr(M_n \leq x) = F^n(x) = 1.$

Thus, $M_n \xrightarrow{P} x_F$ as $n \to \infty$, where $x_F \leq \infty$.

Since the sequence (M_n) is non-decreasing in n, it converges a.s., and hence we conclude that

$M_n \xrightarrow{a.s.} x_F, \;\; n \to \infty.$

Weak Convergence of Maxima under Affine Transformations, or which classes of distributions *F* are closed (up to affine transformations) for maxima, and more formally, which distributions satisfy for all $n \geq 2$ the identity in law:

$max(X_1,\ldots,X_n) \stackrel{d}{=} c_nX \;+\; d_n$

for appropriate constants $c_n > 0$ and $d_n \in \mathbb{R}$.

Those distributions are the only possible limit laws for sums of normalized and centred iid rvs.

Max-Stable Distributions

Max-stable distribution functions are the counterpart of the α-stable distributions.

A non-degenerate random variable (r.v.) *X*, the corresponding distribution or df, is called *max-stable* if it satisfies:

$max(X_1, X_2,\ldots,X_n) \stackrel{d}{=} c_nX + d_n,$

for i.i.d. $X, X_1, \ldots., X_n$, appropriate norming constants $c_n > 0$, $d_n \in \mathbb{R}$ and every $n \geq 2$.

If $\{X_n\}$ is a sequence of i.i.d. max-stable rvs then the above sum can be rewritten as:

$c_n^{-1}(M_n \;-\; d_n) \stackrel{d}{=} X.$

So, max-stable distributions are the only limit laws for normalized maxima.

The extreme value distributions are precisely max-stable distributions.

From the above definition we can see that every max-stable distribution is a limit distribution for maxima of i.i.d. r.v.'s, by using the theorem as follows.

Theorem: Limit Property of Max-Stable Laws (Embrechts et al., 1997)

The class of max-stable distributions coincides with the class of all possible (nondegenerate) limit laws for (properly normalized) maxima of i.i.d. r.v.'s

Theorem: Fisher-Tippett theorem, limit laws for maxima (Fisher and Tippet, 1928)

Let (X_n) be a sequence of i.i.d. r.v.'s and $M_n = \max(X_1, \ldots, X_n)$.

If there exist norming constants $c_n > 0$, $d_n \in \mathbb{R}$ and some non-degenerate d.f. *H* such that:

$c_n^{-1}(M_n - d_n) \xrightarrow{d} H,$

then *H* belongs to the type of one of the following three dfs:

- Fréchet: $\Phi_\alpha(x) = \begin{cases} 0, & x \le 0 \\ exp[-x^{-\alpha}], & x > 0 \end{cases}, \quad \alpha > 0,$
- Weibull: $\Psi_\alpha(x) = \begin{cases} exp[-(-x)^{\alpha}], & x \le 0 \\ 1, & x > 0 \end{cases}, \quad \alpha > 0,$
- Gumbel: $\Lambda(x) = exp[-e^{-x}], \quad x \in \mathbb{R},$

In simple words, this theorem states that if the maximum value of a d.f. tends (in distribution) to a non-degenerate d.f., then this limiting d.f. can only be one of the three forms given above.

By the *limit property of max-stable laws*, the extreme value distributions are precisely the max-stable distributions. If *X* is an *extremal r.v.* it satisfies the above equation.

Proposition: The extreme value distributions (Fréchet, Weibull and Gumbel) are max-stable distributions, and they are the only distributions with this property.

We can easily show that extreme value distributions satisfy the defining property of max-stable distributions.

Let $X_1, \ldots, X_n$ be i.i.d. and $M_n = \max(X_1, \ldots, X_n)$ with distribution function $F_{M_n}(x) = (F(x))^n$, then:

Fréchet case: $F_{M_n}(x) = (0)^n = 0 = F(n^{-1/\alpha}x), \; or \; max(X_1, \ldots, X_n) \overset{d}{=} n^{1/\alpha}X$

Weibull case: $F_{M_n}(x) = (1)^n = 1 = F(n^{1/\alpha}x), \; or \; max(X_1, \ldots, X_n) \overset{d}{=} n^{-1/\alpha}X$

Gumbel case: $F_{M_n}(x) = F(x - \ln n), \; or \; max(X_1, \ldots, X_n) \overset{d}{=} X + \ln n$

To show that they are the only max-stable distributions, we just have to combine the theorem of Limit property of max-stable laws with the Fisher-Tippet Theorem. Indeed, according to the Fisher-Tippet Theorem, extreme value d.f.'s are the only possible limit laws of (properly normalized) maxima. On the other hand, the limit property of max-stable laws dictates that the class of max-stable distributions coincides with the class of all possible (non-degenerate) limit laws for (properly normalized) maxima of i.i.d.

We introduce the notion of *maximum domain of attraction*, in order to explore the necessary conditions for the existence of a limiting d.f. *H*.

Maximum Domain of Attraction

We identified the extreme value distributions as the limit laws for normalized maxima of iid rvs. But:

- Given an extreme value distribution *H*, what conditions on the df *F* imply that the normalized maxima M_n converge weakly to *H*?
- How may we choose the norming constants $c_n > 0$ and d_n such that:
  $c_n^{-1}(M_n - d_n) \overset{d}{\to} H$?
- Can it happen that different norming constants imply convergence to different limit laws?

We can immediately answer the last question. According to the *convergence to types theorem* the limit law is uniquely determined up to affine transformations.

Concerning the other two questions, we have to proceed with the following definition analogously to the sums S_n of iid rvs.

Definition: Maximum Domain of Attraction (Embrechts et al., 1997)

The r.v. *X* (the d.f. *F* of *X*, or the distribution of *X*) is said to belong to the *maximum domain of attraction* of the extreme value distribution *H* if there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H,$$

We write $X \in$ MDA(*H*) (or $F \in$ MDA(*H*)).

That is, the MDA of the extreme value distribution *H* is the family of distributions whose maxima tends, in distribution, to *H*.

The extreme value dfs are continuous on $\mathbb{R}$, hence $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$ is equivalent to:

$$\lim_{n\to\infty} Pr(M_n \le c_n x + d_n) = \lim_{n\to\infty} F^n (c_n x + d_n) = H(x), \ x \in \mathbb{R}.$$

Proposition (Characterization of MDA(H))

The df F belongs to the maximum domain of attraction of the extreme value distribution *H* with norming constants $c_n > 0$, $d_n \in \mathbb{R}$ if and only if

$$\lim_{n\to\infty} n\,\bar{F}(c_n x + d_n) = -ln\, H(x), \ \ x \in \mathbb{R}.$$

When *H*(*x*) = 0 the limit is interpreted as ∞.

Definition (Regularly varying distribution tail)

A distribution tail $\bar{F}$ is regularly varying with index -α for some α ≥ 0, we write $\bar{F} \in \mathcal{R}_{-\alpha}$, if

$$\lim_{x\to\infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-\alpha}, \ \ t > 0.$$

Definition (Tail-equivalence)

Two dfs *F* and *G* are called tail-equivalent if they have the same right endpoint, i.e. if $x_F = x_G$, and

$$\lim_{x\uparrow x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = c, \ \ for\ some\ constant\ 0 < c < \infty.$$

It can be shown that every maximum domain of attraction is closed with respect to tail-equivalence, i.e. for tail-equivalent *F* and *G*, $F \in$ MDA(*H*) if and only if $G \in$ MDA(*H*). Moreover, for any two tail-equivalent distributions one can take the same norming constants.

Definition (Generalized inverse of a monotone function)

Suppose *h* is a non-decreasing function on $\mathbb{R}$. The generalized inverse of h is defined as

$h^{\leftarrow}(t) = inf\{x \in \mathbb{R} : h(x) \geq t\},$

by using the convention that the infimum of an empty set is ∞.

Definition (Quantile function)

The generalized inverse of the df *F*

$F^{\leftarrow}(t) = inf\{x \in \mathbb{R} : F(x) \geq t\}, \quad 0 < t < 1,$

Is called the quantile function of the df *F*. The quantity $x_t = F^{\leftarrow}(t)$ defines the *t* – quantile of *F*.

- **The Maximum Domain of Attraction of the Fréchet Distribution $\boldsymbol{\Phi_\alpha(x) = exp\{-x^{-\alpha}\}}$**

By Taylor expansion, for α > 0,

$1 - \Phi_\alpha(x) = 1 - exp\{-x^{-\alpha}\} \sim x^{-\alpha}, \ x \to \infty,$

Hence the tail of Φ_α decreases like a power law.

But how far away can we move from a power tail and still remain in MDA (Φ_α)?

Theorem (Maximum domain of attraction of $\boldsymbol{\Phi_\alpha}$)

The df *F* belongs to the maximum domain of attraction of Φ_α, *α* > 0, if and only if $\bar{F}(x) = x^{-\alpha}L(x)$ for some slowly varying function *L*.

If F ϵ MDA (Φ_α) then

$c_n^{-1}M_n \xrightarrow{d} \Phi_\alpha,$

This result implies that every F ϵ MDA (Φ_α) has an infinite right endpoint $x_F = \infty$. Furthermore, the norming constants c_n form a regularly varying sequence, $c_n = n^{1/\alpha}L_1(n)$ for some slowly varying function L_1.

Since now we found that:

$F \in MDA\ (\Phi_\alpha) \Leftrightarrow \bar{F} \in \mathcal{R}_{-\alpha}.$

This class of dfs contains "very heavy-tailed distributions" in the sense that $E(X^+)^\delta = \infty \ \ for\ \delta > \alpha.$

Thus, they may be appropriate distributions for modeling large fluctuations of prices, log-returns etc.

Example:

Pareto, Cauchy, Burr, Stable with exponent α < 2, are all Pareto-like in the sense that their right tails are of the form:

$\bar{F}(x) \sim Kx^{-\alpha}, \ \ x \to \infty, for\ some\ K, \alpha > 0.$

As norming constants, we can choose $c_n = (Kn)^{1/\alpha}$. Then

$(Kn)^{-1/\alpha}M_n \xrightarrow{d} \Phi_\alpha.$

Example (Log gamma distribution)

The log gamma distribution has tail:

$$\bar{F}(x) \sim \frac{\alpha^{\beta-1}}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha}, \quad x \to \infty, \quad \alpha, \beta > 0.$$

Hence $\bar{F} \in \mathcal{R}_{-\alpha}$ which is equivalent to $F \in MDA\ (\Phi_\alpha)$.

- **The Maximum Domain of Attraction of the Weibull Distribution $\Psi_\alpha(x) = exp\{-\ (-x)^\alpha\}$**

As before, we have to remark that Ψ_α and Φ_α are closely related as:

$$\Psi_\alpha(-x^{-1}) = \Phi_\alpha(x), \quad x > 0,$$

therefore, we may expect that also MDA (Ψ_α) and MDA (Φ_α) will be closely related.

We have also to notice that all dfs *F* in MDA (Ψ_α) have finite right endpoint x_F.

Theorem (Maximum domain of attraction of Ψ_α)

The df *F* belongs to the maximum domain of attraction of Ψ_α, $\alpha > 0$, if and only if $x_F < \infty$ and

$$\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x),$$

for some slowly varying function *L*.

If F ϵ MDA (Ψ_α), then

$$c^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha,$$

Where the norming constants c_n can be chosen as

$$c_n = x_F - F^{\leftarrow}(1 - n^{-1}), \ \ and \ \ d_n = x_F.$$

Consequently,

$$F \in MDA\ (\Psi_\alpha) \Leftrightarrow x_F < \infty, \bar{F}(x_F - x^{-1}) \in \mathcal{R}_{-\alpha}.$$

Thus MDA (Ψ_α) consists of dfs *F* with support bounded to the right. Because $x_F < \infty$, this is not the best choice for modeling financial data. Distributions with $x_F = \infty$ should be preferred since they allow for arbitrarily large values in a sample. Such distributions typically belong to MDA (Φ_α) or MDA (Λ).

Examples are the Uniform Distribution or the Beta Distribution with $x_F = 1$,

- **The Maximum Domain of Attraction of the Gumbel Distribution $\Lambda(x) = exp\{-\ exp(-x)\}$**

A Taylor expansion argument yields

$$1 - \Lambda(x) \sim e^{-x}, \ x \to \infty,$$

hence $\bar{\Lambda}(x)$ decreases to zero at an exponential rate.

MDA (Λ) contains dfs with very different tails, ranging from *moderately heavy*, as the lognormal distribution to *light* as the normal distribution. Also, it is possible to be $x_F < \infty$ and $x_F = \infty$. As a natural consequence of the variety of tails in MDA (Λ), the norming constants also vary considerably.

Whereas in MDA (Φ_α) and MDA (Ψ_α) the norming constants are calculated by straightforward application of regular variation theory, more advanced results are needed for MDA (Λ).

Notice, that now there is no direct linkage with regular variation notion. However, MDA(Λ) covers a wider collection of d.f.'s with very different tails ranging from moderately heavy (such as the lognormal) to light (e.g., normal distribution). D.f.'s of this class can be either unbounded or bounded on the right.

Still, the above conditions are difficult to verify in practice. Von Mises (1936) has given some simpler sufficient conditions which can be used for practical purposes.

Corollary: Von Mises Condition for MDA(Φ_α) (Embrechts et al., 1997

Let *F* be an absolutely continuous d.f. with density *f* satisfying:

$$\lim_{x\to\infty} \frac{xf(x)}{\bar{F}(x)} = \alpha > 0,$$

then $F \in$ MDA(Φ_α).

Corollary: Von Mises Condition for MDA(Ψ_α) (Embrechts et al., 1997)

Let *F* be an absolutely continuous d.f. with density *f* which is positive on some finite interval (z, x_F). If

$$\lim_{x\to x_{\bar{F}}} \frac{(x_F - x)f(x)}{\bar{F}(x)} = \alpha > 0,$$

then $F \in$ MDA(Ψ_α).

Corollary: Von Mises Condition for MDA(Λ) (Embrechts et al., 1997)

Let *F* be a d.f. with right endpoint $x_F \le \infty$, such that for $z < x_F$ *F* has the representation:

$$\bar{F}(x) = c\,.\,exp[-\int_z^x \frac{1}{a(t)} dt], \;\; z < x < x_F,$$

where *c* is some positive constant, $a(x)$ is a positive, absolutely continuous function (with respect to Lebesgue measure) with density $\alpha'(x)$ having $l \lim_{x\to x_{\bar{F}}} a'(x) = 0$. Then F $\in$ MDA(Λ).

Note that the d.f.'s just described are known as "*von Mises functions*".

Through the procedures used to derive the previous results, proper forms for the norming constants appear which are different for the three different types of limiting d.f.'s.

Table 1. Norming Constants for the Maximum Domains of Attraction

Maximum Domain of Attraction	*Centring Constant d_n*	*Normalizing Constant c_n*
Fréchet	0	$F^{\leftarrow}(1 - n^{-1})$
Weibull	x_F	$x_F - F^{\leftarrow}(1 - n^{-1})$
Gumbel	$a(d_n)$	$F^{\leftarrow}(1 - n^{-1})$

*<u>Note</u>: $F^{\leftarrow}(u) = inf\{x: F(x) \ge u\}$ is the generalized inverse function of the d.f. F, i.e., it is the well-known **quantile function**.*

Limit Laws for Minima

The study of minima is totally equivalent to the study of maxima, since it holds that:

$min(X_1, \ldots, X_n) = -\,max(-X_1, \ldots, -X_n)$,

All the findings mentioned above, concerning maxima have an analogous formula and hold for minima as well. So, there is a one-to-one relationship between limiting d.f.'s of maxima and minima.

More particularly if

$Pr(\max_{i \le n}(-X_i) \le b_n + a_n x) \overset{n\to\infty}{\to} H(x)$, then

$Pr(\min_{i \le n}(X_i) \le d_n + c_n x) \overset{n\to\infty}{\to} 1 - H(-x)$,

Where $c_n = a_n$ and $d_n = -\,b_n$.

This yields that the possible limiting d.f.'s of minima can be only of three types (in the sense of Fisher-Tippet):

- *Converse Fréchet*: $\widetilde{\Phi}_\alpha(x) = 1 - \Phi_\alpha(-x) \begin{cases} 1, & x > 0 \\ 1 - exp[-(-x)^{-\alpha}], & x \le 0 \end{cases}, \quad \alpha > 0,$
- *Converse Weibull*: $\widetilde{\Psi}_\alpha = 1 - \Psi_\alpha(-x) = \begin{cases} 1 - exp[-(x)^{\alpha}], & x > 0 \\ 0, & x \le 0 \end{cases}, \quad \alpha > 0,$
- *Converse Gumbel*: $\tilde{\Lambda}(x) = 1 - \Lambda(-x) = 1 - exp[-e^x], \quad x \in \mathbb{R},$

Some well-known special cases are:

The Converse Gumbel is also called *Gompertz d.f.* This is the d.f. that satisfies the famous *Gompertz law.*

$\widetilde{\Psi}_1(x)$ is the *exponential d.f.* on the positive half-line.

$\widetilde{\Psi}_2(x)$ is the *Rayleigh d.f.* that is also of interest in some statistical applications.

A characteristic relationship is that if the areas of random circles are exponentially distributed, then the diameters have a Rayleigh d.f.

Under the unified representation (ξ-reparameterization) the limiting d.f. of minima is the *Converse Generalized Extreme-Value distribution*:

$$\widetilde{H}_\xi(x) = 1 - H_\xi(-x) = \begin{cases} exp[-(1 - \xi x)^{-1/\xi}] & if\ \xi \ne 0 \\ exp[-exp(x)] & if\ \xi = 0 \end{cases},$$

Where $1 - \xi x > 0$, i.e., the support of $\widetilde{H}_\xi$ is

$x < \xi^{-1}$ for $\xi > 0$,

$x > \xi^{-1}$ for $\xi < 0$

$x \in \mathbb{R}$ for $\xi = 0$.

Another characteristic property of maxima which is transmitted to minima is that of max-stability. More precisely, the limiting d.f.'s of minima are characterized by the min-stability. That is, the d.f.'s converse Fréchet, converse Weibull and converse Gumbel are the only distributions for which it holds that

$$Pr\{\min_{i\leq n}(X_i) \leq d_n + c_n x\} = 1 - (1 - F(d_n + c_n x))^n = F(x),$$

for a certain choice of constants $c_n > 0$, $d_n \in \mathbb{R}$, where (X_i) are i.i.d. random variables with common d.f.

F $(\tilde{\Phi}, \tilde{\Psi}$, or $\tilde{\Lambda})$.

The min-stability can also be expressed in terms of the *survivor function* $\bar{F}$, that is,

$$Pr\{\min_{i\leq n}(X_i) > d_n + c_n x\} = \bar{F}^n(d_n + c_n x) = \bar{F}(x).$$

The below table summarizes the forms of limiting distributions for maxima and minima for seven most widely used continuous distributions (see Kotz et al., 2000):

Table 2. Minima and Maxima Limiting Distributions

Initial Distribution	**Limiting Distribution for Extremes**	
	Maxima	**Minima**
Exponential	Gumbel	Weibull
Gamma	Gumbel	Weibull
Normal	Gumbel	Gumbel
Log-Normal	Gumbel	Gumbel
Uniform	Weibull	Weibull
Pareto	Fréchet	Weibull
Cauchy	Fréchet	Fréchet

The Generalized Extreme Value Distribution (GEV)

According to the fundamental theorem of Fisher-Tippet, the possible limiting d.f.'s of maxima are of three types. These three types can be summarized into a single family of distributions, known as *Generalized Extreme Value distribution*, using a different parametrization. The idea of this unification is attributed to von Mises (1936), though often is also attributed to Jenkinson (1955). The corresponding definition is:

Definition: Generalized Extreme-Value distribution (GEV) (see Embrechts et al., 1997)

First, we introduce the parametric family $(H_\xi)_{\xi\in\mathbb{R}}$ of dfs containing the standard extreme value distributions, namely:

$$H_\xi = \begin{cases} \Phi_{1/\xi} \; if \; \xi > 0, \\ \Lambda \; if \; \xi = 0, \\ \Psi_{-1/\xi} \; if \; \xi < 0 \end{cases}.$$

The df H_ξ above is referred to as the *generalized extreme value distribution* with parameter ξ which can be obtained as a limit.

The condition $F \in MDA\ (H_\xi)$ yields the so-called *ultimate approximation:*

$$F^n(c_n x + d_n) \approx H_\xi(x),$$

for appropriate norming constants $c_n > 0$ and $d_n \in \mathbb{R}$.

The one-parameter representation of the three standard cases in one family of dfs can be represented as follows:

- $\xi = \alpha^{-1} > 0$, corresponds to the Fréchet distribution Φ_α,
- $\xi = 0$, corresponds to the Gumbel distribution Λ,
- $\xi = -\alpha^{-1} < 0$, corresponds to the Weibull distribution Ψ_α.

The standard, by Jenkinson-von Mises representation of the extreme value distributions is defined by:

$$H_\xi = \begin{cases} exp\{-(1 + \xi x)^{-1/\xi}\}, \; if \; \xi \neq 0, \\ exp\{-exp\{-x\}\}, \; if \; \xi = 0, \end{cases} where \; 1 + \xi x > 0.$$

The support for H_ξ corresponds to:

$$\begin{cases} x > -\xi^{-1} \; for \; \xi > 0, \\ x < -\xi^{-1} \; for \; \xi < 0, \\ x \in \mathbb{R} \; for \; \xi = 0. \end{cases}$$

H_ξ is called a *standard generalized extreme value distribution* (GEV). The related *location-scale* family $H_{\xi;\mu,\psi}$ obtained by replacing the argument x above by $(x - \mu) / \psi$ for $\mu \in \mathbb{R}$, $\psi > 0$. Support has to be accordingly adjusted.

The corresponding *standard probability density function (pdf)* is given:

$$h_\xi(x) = \begin{cases} (1+\xi x)^{-\frac{\xi+1}{\xi}} * exp[-(1 + \xi x)^{-\frac{1}{\xi}}], & if \;\; \xi \neq 0 \\ exp[-(x + exp(-x))], \; if \;\; \xi = 0 \end{cases}$$

The parameter ξ is called the *shape parameter* and may be used to model a wide range of tail behavior. The case $\xi > 0$ is that of a polynomially decreasing tail function and therefore corresponds to a *long - tailed* parent distribution. The case $\xi = 0$ is that of an *exponentially decreasing tail*, while $\xi < 0$ is the case of a finite upper endpoint and is therefore *short tailed*.

There are two forms to characterize GEV. Such a one-parameter representation of the three standard cases in one family of d.f.'s will turn out to be particularly useful. Its introduction was mainly motivated by statistical applications.

Theorem: Characterization I of Generalized Extreme-Value distribution (Embrechts et al., 1997)

The d.f. *F* with right endpoint $x_F \leq \infty$ belongs to the maximum domain of attraction of H_ξ ($F \in$ MDA(H_ξ)) if and only if, for $1 + \xi x > 0$,

$$\lim_{u \to x_{\bar{F}}} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1+\xi x)^{-1/\xi} & if \;\; \xi \neq 0 \\ e^{-x} & if \;\; \xi = 0 \end{cases},$$

where *a*(*u*) is a positive, measurable function.

An interesting probabilistic interpretation stems by observing that

$\bar{F}(u + xa(u)) \, / \, \bar{F}(u) = Pr((X - u)/a(u) > x | X > u)$. That is, the above characterization provides a distributional approximation for the scaled excess over the (high) threshold *u*. The appropriate scaling factor is *a*(*u*).

Theorem: Characterization II of Generalized Extreme-value distribution (Embrechts et al., 1997)

The d.f. *F* with right endpoint $x_F \leq \infty$ belongs to the maximum domain of attraction of H_ξ ($F \in$ MDA(H_ξ)) if and only if, for $x, y > 0$, $y \neq 1$,

$$\lim_{s \to \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & if \;\; \xi \neq 0 \\ \frac{\ln x}{\ln y} & if \;\; \xi = 0 \end{cases},$$

where $U(t) = F^{\leftarrow}(1 - t^{-1}), \; t > 0$.

A reformulation of this relation leads to an estimation procedure for quantiles outside the range of the data, while a special case of this formula is also used to motivate the *Pickands estimator* of ξ.

The above theorem provides the necessary and sufficient conditions for a d.f. *F* to have maxima that converge weakly to a non-degenerate d.f. If *F* does not satisfy the above conditions, then its maxima does not converge in any distribution. Another point that we should mention is that the limit d.f.'s is unique only up to affine transformations, i.e., in each case the exact limiting d.f. is not necessarily the standard GEV distribution but it can be a d.f. of the more general location-scale GEV family. Still, this does not cause any identification problem, since in any case, by appropriate normalization of maxima, we can end up in the standard GEV.

Summarizing, we have:

	MDA ($\boldsymbol{\Phi_\alpha}$)	*Examples*
Fréchet	$\Phi_\alpha(x) = exp\{-x^{-\alpha}\}$, $x > 0, \alpha > 0$	• Cauchy
MDA (Φ_α)	$x_F = \infty$, $\bar{F}(x) = x^{-\alpha}L(x)$, $L \in \mathcal{R}_0$	• Pareto
Norming constants	$c_n = F^{\leftarrow}(1 - n^{-1}) = n^{1/\alpha}L_1(n), L_1 \in \mathcal{R}_0, d_n = 0.$	• Burr • Log gamma
Limit result	$c_n^{-1}M_n \xrightarrow{d} \Phi_\alpha$	• Stable with $\alpha < 2$

	MDA ($\boldsymbol{\Psi_\alpha}$)	*Examples*
Weibull	$\Psi_\alpha(x) = exp\{-(-x)^{-\alpha}\}$, $x < 0, \alpha > 0$	• Uniform
MDA (Φ_α)	$x_F < \infty$, $\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$, $L \in \mathcal{R}_0$	• Beta
Norming constants	$c_n = x_F - F^{\leftarrow}(1 - n^{-1}) = n^{-1/\alpha}L_1(n), L_1 \in \mathcal{R}_0, d_n = x_F.$	• Power law behavior at x_F
Limit result	$c_n^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha$	

	MDA (Λ)	*Examples*
Gumbel	$\Lambda(x) = exp(-e^{-x}\}$, $x \in \mathbb{R}$	• Exponential-like
MDA (Λ)	$x_F \leq \infty$, $\bar{F}(x) = c(x)exp\{-\int_{x_0}^{x}\frac{g(t)}{\alpha(t)}dt\}$, $x_0 < x < x_F$, where $c(x) \to c > 0$, $g(x) \to 1$, $\alpha'(x) \to 0$, as $x \uparrow x_F$.	• Weibull-like • Gamma • Normal
Norming constants	$d_n = F^{\leftarrow}(1 - n^{-1}), c_n = \alpha(d_n).$	• Lognormal
Limit result	$c_n^{-1}(M_n - d_n) \xrightarrow{d} \Lambda$	• Exponential behavior as $x_F < \infty$

Distributional Characteristics of Generalized Extreme-Value distribution.

- The mean exists only for $\xi < 1$, and is given by

$$E(H_\xi) = \frac{\Gamma(1-\xi)-1}{\xi},$$

- The variance exists only for $\xi < 1/2$, and equals:

$$Var(H_\xi) = \frac{\Gamma(1-2\xi) \; - \; \Gamma^2(1-\xi)}{\xi^2}$$

- The GEV densities are unimodal with mode:

$$mod(H_\xi) = \frac{(1 \; + \; \xi)^{-\xi} \; - \; 1}{\xi}, for \; \xi \; > \; -1$$

while they are J – shaped for $\xi \leq -1$, and where $\Gamma(.)$ is the Gamma function.

The case $\xi = 0$ is included in the latter formulas by considering limits with ξ tending to zero. That is:

$E(H_0) = \lim_{\xi \to 0} E(H_\xi) = \int_0^\infty (-\, lnx)\, e^{-x} dx = \lambda, \; \lambda = Euler's \; constant,$

$Var(H_0) = \lim_{\xi \to 0} Var(H_\xi) = \frac{\pi^2}{6},$

mod (H_0) = 0.

In the figure below, we visually inspect the form of a limiting d.f. of normilized maxima, depending on the shape parameter ξ.

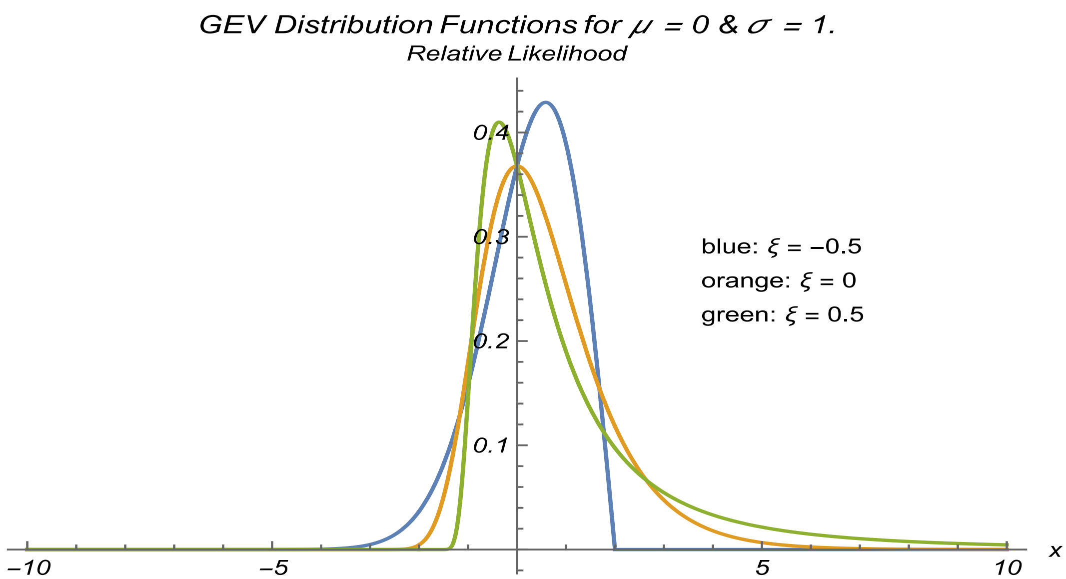

Figure 2. GEV distribution with different values of shape parameter.

Generalized Pareto Distribution (GPD)

The GPD plays a central role in the study of extremes, comparable to the role the normal distribution plays when studying observations in the center of the distribution.

Consider an unknown distribution function *F* of a random variable *X*. We are interested in estimating the distribution function F_u of values of *x* above a certain threshold *u*.

This distribution function F_u is called the *conditional excess distribution function* and is defined as:

$$F_u(y) = Pr(X - u \leq y | X > u), \;\; 0 \leq y \leq x_F - u,$$

where *X* is random variable, *u* is a given threshold, *y* = *x* – *u* are the excesses and $x_F \leq \infty$ is the right endpoint of *F*.

The realizations of the random variable *X* lie mainly between 0 and *u* therefore the estimation of *F* in this interval there is no poses generally problems. But due to the lack of many observations in this area, the estimation of the portion of F_u might be difficult.

EVT provides us with a powerful result about the conditional excess distribution function which is stated in the following theorem:

Theorem: (Pickands (1975), Balkema and de Haan (1974))

For a large class of underlying functions *F* the conditional excess distribution function $F_u(y)$, for *u* large, is well approximated by:

$F_u(y) \approx G_{\xi,\sigma}(y), \;\; u \rightarrow \infty,$

where:

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - (1 + \frac{\xi}{\sigma} y)^{-1/\xi} & if \; \xi \neq 0 \\ 1 - e^{-y/\sigma} & if \; \xi = 0 \end{cases},$$

If *x* is defined as *x* = *u* + *y*, the GPD can also be expressed as a function of *x*, i.e.

$$G_{\xi,\sigma}(x) = 1 - (1 + \frac{\xi}{\sigma}(x - u))^{-1/\xi}.$$

The tail index ξ gives an indication of the heaviness of the tail, the larger the ξ, the heavier the tail. In general, we cannot fix the upper bound for financial losses, and only distributions with shape parameter $\xi \geq 0$ are suited to model financial return distributions.

Embrechts et al. give its standardized form as:

Definition: The Generalized Pareto Distribution (Embrechts et al., 1997)

The Generalized Pareto distribution G_ξ is defined by the formula:

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi} & if \ \xi \neq 0 \\ 1 - e^{-x} & if \ \xi = 0 \end{cases},$$

where the support of G_ξ is

$x \geq 0$ if $\xi \geq 0$, and

$0 \leq x \leq -1/\xi$ if $\xi < 0$.

The corresponding p.d.f. is:

$$g_\xi(x) = \begin{cases} (1 + \xi x)^{-(1 + 1/\xi)} & if \ \xi \neq 0 \\ e^{-x} & if \ \xi = 0 \end{cases},$$

The following figure displays the Generalized Pareto p.d.f. for different values of ξ. Large similarities exist among the graphs of p.d.f. for ξ equal to +0.1, -0.1 and 0. The main difference is that in the case ξ = -0.1, x cannot exceed 10, while in the other two cases x can become infinitely large.

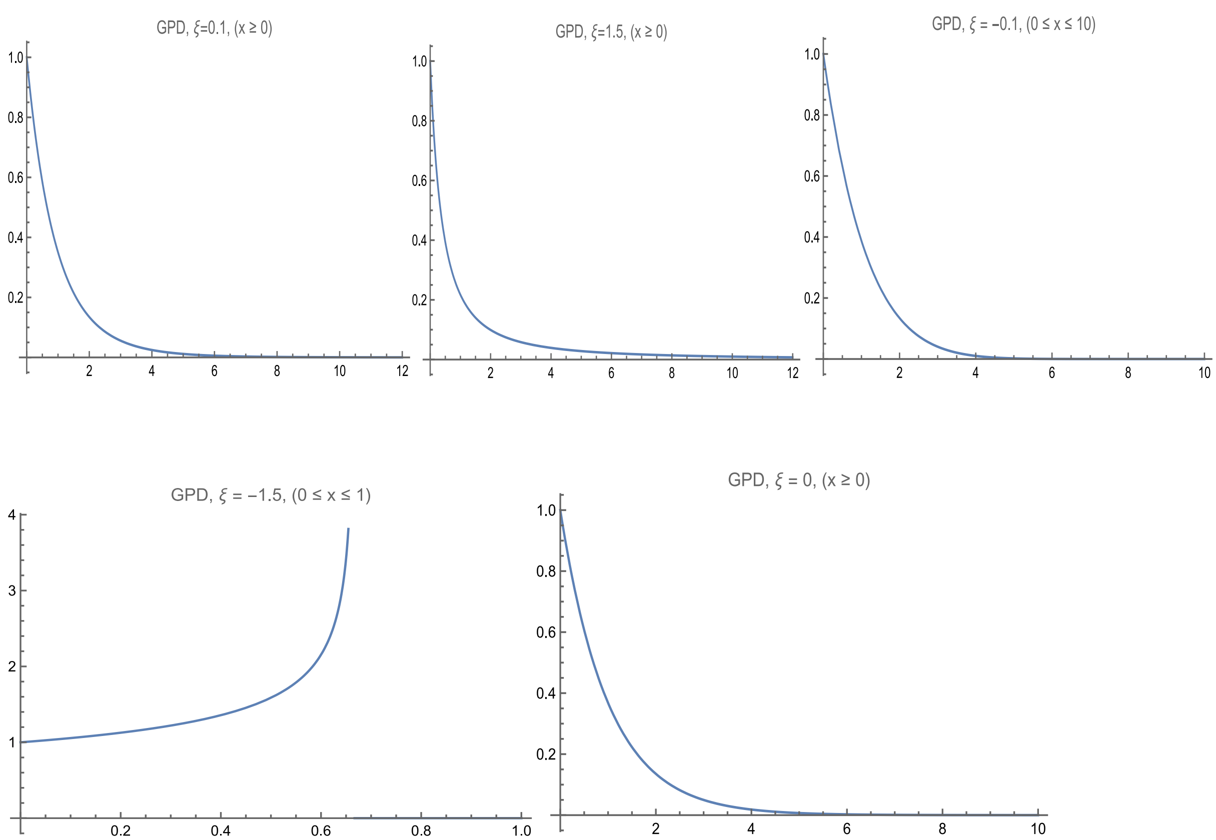

Figure 3. A GP distribution function for different values of ξ.

For $\xi > 0$ the GPD corresponds to *Pareto d.f.*, for $\xi < 0$ to *Beta d.f.*, while for $\xi = 0$ we get the *standard exponential distribution*.

The related location-scale family $G_{\xi;u,\beta}$ can be derived by replacing the argument x above by $(x - u) / \beta$ for $u \in \mathbb{R}$, $\beta > 0$. The support has to be adjusted accordingly.

As in the case of H_0, G_0 can also be interpreted as the limit of G_ξ as $\xi \to 0$.

Summarizing we have:

- GEV H_ξ, $\xi \in \mathbb{R}$, describes the limit distributions of normalized maxima,
- GPD $G_{\xi,\beta}$, $\xi \in \mathbb{R}$, $\beta > 0$, appears as the limit distribution of scaled excesses over high thresholds.

The strong connection of Generalized Pareto d.f. to extreme value theory is revealed in the following considerations. If we take the first characterization property of the maximum domain of attraction of generalized extreme value d.f. (MDA(H_ξ)), we notice that the left-hand side of the relation can be rewritten as:

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = Pr(X > u + xa(u) \mid X > u) = Pr(\frac{X - u}{a(u)} > x \mid X > u) = \bar{F}_u(a(u)x),$$

where F_u stands for the conditional distribution of X given that $X > u$.

That is, if X ϵ MDA(H_ξ), then it holds that for $1 + \xi x > 0$,

$$\lim_{u \to x_{\bar{F}}} Pr(\frac{X - u}{a(u)} > x \mid X > u) = \begin{cases} (1 + \xi x)^{-1/\xi} & if\ \xi \neq 0 \\ e^{-x} & if\ \xi = 0 \end{cases},$$

which gives a distributional approximation for the scaled excesses over the (high) threshold u by the GPD. The appropriate scaling factor is $a(u)$. In this context, F_u is called *excess distribution function*.

So, we can derive the following proposition, which reveals the connection of the Generalized Pareto distribution to extreme-value theory.

Proposition: Limit distribution of scaled excesses over high thresholds (Embrechts et al., 1997)

If the d.f. F with right endpoint $x_F \leq \infty$ belongs to the maximum domain of attraction of H_ξ ($F \in$ MDA(H_ξ)) then, for $1 + \xi x > 0$,

$$\lim_{u \to x_{\bar{F}}} Pr(\frac{X - u}{a(u)} > x \mid X > u) \stackrel{d}{=} \bar{G}_{a(u),\xi}(x).$$

Properties of GPD

Maybe, the most important property of the GPD is the one known as POT-stability, (POT stands for Peaks Over Threshold), i.e., the class of GPD's is closed with respect to changes of the threshold.

Formally, we have the following proposition:

Proposition: POT-Stability of Generalized Pareto Distributions (Reiss and Thomas, 1997)

Generalized Pareto d.f.'s is the only continuous d.f.'s such that for a certain choice of constants b_u and a_u, it holds that:

$$F^{[u]}(b_u + a_u x) = F(x),$$

where ,

$F^{[u]}(x) = Pr(X \leq x \mid X > u) = \frac{F(x) - F(u)}{\bar{F}(u)}$,

is the exceedance d.f. over the threshold *u* (truncation of *F* left at *u*). That is, the truncated version of a GPD remains in the same family of d.f.'s.

The usefulness of this proposition lies on the fact that if we can assume that the d.f. of the scaled excesses over a high threshold *u* is well approximated by GPD, then the same holds for any higher threshold *v* > *u*. This remark will turn out to be of great practical interest in the case of estimation of the extreme-value index ξ.

Let $G_{\sigma,\xi}$, be a GPD with shape parameter ξ and scale parameter σ (when written as $\sigma(u)$ it implies dependence of σ on *u*), it holds:

- For every $\xi \in \mathbb{R}$, $F \in$ MDA(H_ξ) if and only if

$\lim_{u \to x_{\bar{F}}} \sup_{0<x<x_F-u} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0$,

where, $F_u(x) = Pr(X - u \leq x \mid X > u)$, $x \geq 0$, is the excess d.f. over the threshold *u*.

- For every x_i, i = 1,2 (to the appropriate domain of support),

$\frac{\bar{G}_{\xi,\sigma}(x_1 + x_2)}{\bar{G}_{\xi,\sigma}(x_1)} = \bar{G}_{\xi,\sigma + \xi x_1}(x_2)$,

a reformulation of the POT-stability property.

- If $\xi < 1$, then for $u < x_F$, it holds that

$e(u) = E(X - u \mid X > u) = \frac{\sigma + \xi x}{1 - \xi}, \quad \sigma + \xi x > 0$,

where *e*(*u*) is called *mean-excess function*.

This property shows that the mean-excess function is linear with respect to *u*. This remark is a key-point to many statistical techniques (estimation methods for ξ).

- If $\xi < 1$, then it also holds:

$E(X) < \infty$

$E[(1 + \frac{\xi}{\sigma}X)^{-r}] = \frac{1}{1 + \xi r}, \ for\ r > -1/\xi$,

$E[\ (ln\ (1 + \frac{\xi}{\sigma}X))^k] = \xi^k k!, \ for\ k \in \mathbb{N}$,

$E[X\ (\bar{G}_{\xi,\beta}\ (X))^r] = \frac{\beta}{(r + 1 - \xi)(r + 1)}, \quad (r + 1)/|\xi| > 0$,

Moreover,

If $\xi < 1/r$, $r \in \mathbb{N}$, then

$$E(X^r) = \frac{\beta^r \Gamma(\xi^{-1} - r)}{\xi^{r+1}\Gamma(1 + \xi^{-1})} r!.$$

Let N be a Poisson r.v. with parameter λ ($P(\lambda)$), independent of the i.i.d. sequence (X_n) with a $G_{\xi,\beta}$ d.f., and $M_N = \max(X_1, \ldots, X_N)$. Then, it holds that:

$$Pr(M_N \leq x) = exp\left\{-\lambda(1 + \xi\frac{x}{\beta})^{-1/\xi}\right\} = H_{\xi,\mu,\sigma}(x),$$

Where $\mu = \beta\xi^{-1}(\lambda^{\xi} - 1)$ is the location parameter and $\sigma = \beta\lambda^{\xi}$ is the scale parameter.

The essence of this property is that if the number of exceedances over a high threshold is exact Poisson and the excess d.f. is an exact GPD, then the maximum of these excesses has an exact GEV distribution.

The above properties suggest the following approximate model for the exceedance times and the excesses of an i.i.d. sample:

- The number of exceedances of a high threshold follows a Poisson process.
- Excesses over high thresholds can be modelled by a GPD.
- An appropriate value of the high threshold can be found by plotting the empirical mean excess function (and searching for that point from which linearity seems to start).
- The distribution of the maximum of a Poisson number of i.i.d. excesses over a high threshold is a GEV distribution.

Approaches to EVT

Given an unknown distribution *F* (i.e., return time series of financial instruments), extreme value theory (EVT) is interested in *modelling the tail* of *F* only, neglecting any specific distributional assumption concerning the center of the distribution.

Inference consists of fitting the GEV or GP distributions to the observed maxima/minima or threshold exceedances, followed by model verification and extrapolation.

This goal can be attained by means *of three different approaches*, two of which are *parametric* in nature, while the third one is *non-parametric*.

To pursue the above analysis, the *tail index* and the bounds which are associated with certain low probabilities of an excess must be estimated. In other words, to fit the model, one needs to estimate the shifting and scaling parameters, c_n, d_n and the tail index, α. The tail index is the most difficult to estimate rigorously. The advantage of the extreme value approach is that all fat-tailed models are nested with respect to their tail index in one model. The procedure is then to estimate this index directly and use the asymptotic confidence interval to discriminate between the competing hypotheses. The tail index, given several observations X_ι, can be estimated by *maximum likelihood* (see Smith, 1987) or by a *moment estimator*. The traditional approach, maximum likelihood, assumes that each period's maximum exactly follows one of the three limit laws. If the type I limit law (Fréchet) applies, direct estimation by maximum likelihood is consistent.

We will focus on two methods of extreme value analysis. The first approach, GEV, looks at distribution of *block maxima* (a block being defined as a set time period such as a weak, month or year); depending on the shape parameter, a Gumbel, Fréchet, or Weibull distribution will be produced. The second method, GP, looks at values that *exceed a defined threshold*; depending on the shape parameter, an Exponential, Pareto, or Beta distribution will be produced.

Parametric Estimation

Block Maxima Method

The block maxima (BM) approach in extreme value theory (EVT), consists of dividing the observation period into nonoverlapping periods of equal size and restricts attention to the maximum observation in each period [Gumbel, 1958]. The new observations follow approximately an extreme value distribution, H_ξ, for some real ξ and under domain of attraction conditions. Parametric statistical methods are then applied to those observations.

More specific, the *block maxima method* divides the given *N* observations sample into *m* subsamples of *n* observations each (*n*-blocks) and picks the maximum M_k (k = 1, . . . , m) of each subsample, a so-called *block maximum*.

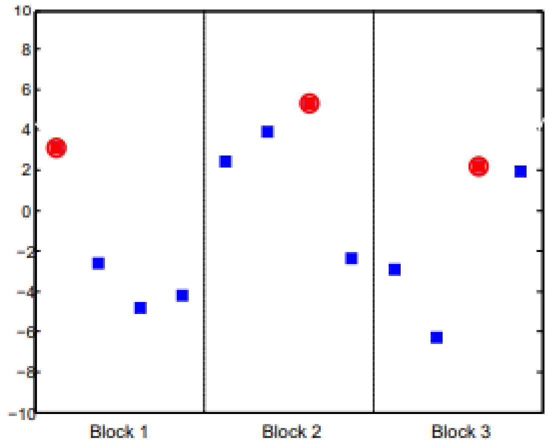

Figure 4. Block Maxima Method

The set of extreme values of *F* is then identified with the sequence (M_k) of block maxima and the distribution of this sequence is studied. The main result of EVT states that, as *m* and *n* grow sufficiently (ideally, as $m, n \rightarrow +\infty$, under some additional assumption), the limit distribution of (adequately rescaled) block maxima belong to one out of three different families. Which of them it belongs to, depends on the behavior of the upper tail of *F*, whether it is *power-law decaying* (the fat-tailed case, of major concern to financial applications), *exponentially decaying* (e.g., if F is the normal distribution), or *with upper bounded support* (the less relevant case for finance).

Fitting a generalized extreme-value distribution to a sample of monthly maxima is the earliest statistical method in extreme-value theory. Various inference procedures have been explored, including quantile or probability matching (Gumbel, 1958), the probability weighted moment method (Hosking et al., 1985), and the maximum likelihood method (Prescott and Walden, 1980; Hosking, 1985).

Threshold Exceedances Method

The second *parametric approach*, the *threshold exceedances method*, defines extreme values as those observations that exceed some fixed high threshold *u*. The aim of this method is therefore to model the distribution of the exceedances over *u* the random variables $Y_j = X_j - u$, for those observations (returns, in our case) X_j that exceed *u* (i.e., such that $X_j > u$).

The *peak over threshold-method* (POT-method) is one of many methods that falls under extreme value analysis and is based around looking at extreme values of some sample that exceeds a certain threshold value. A threshold *u* could be set at the 90% quantile of the data. Using the POT-method all sample points that exceed the threshold are of interest and separated from the rest of the data for further analysis. It has been proven that the values above the threshold can be modeled as a generalized Pareto distribution.

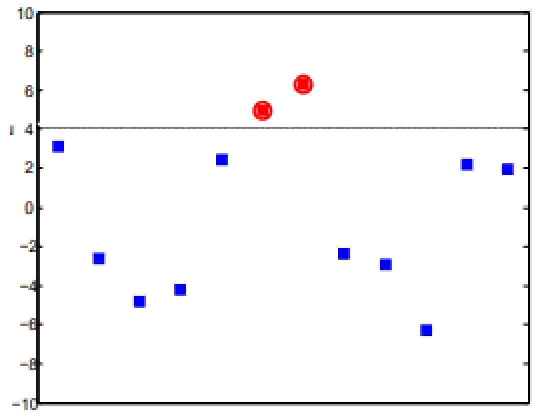

Figure 5. Peaks-Over-Threshold Approach

The main result of EVT following this approach is that, as the threshold u grows to infinity (or to the right endpoint of the support of F, if that point is finite), the distribution of the positive sequence (Y_j), appropriately scaled, belongs to a parametric family, the *generalized Pareto distribution (GPD),* whose main parameter is the same shape parameter ξ of the corresponding GEV distribution. That is to say for instance, financial returns whose block maxima follow a GEV distribution with a certain value $\xi_0 > 0$ of the parameter ξ are such that, for a sufficiently high threshold u, the exceedances over u follow a GPD with $\xi = \xi_0$.

Consider observations $X_1, \ldots, X_T$ that are independent and identically distributed from a distribution function F_X that belongs to a wide class of continuous distributions. For a given high threshold u, the POT method is based on the decomposition of the tail of F_X as

$$1 - F_X(x) = Pr(X > u)\{1 - F_u(x)\}, \;\; x \geq u,$$

Where the (conditional) excess distribution F_u is defined as

$$F_u(x) = Pr(X \leq x \mid X > u), \;\; x \geq u.$$

Letting $E_u = \{t \in \{1, \ldots, T\} : X_t > u\}$ be the set of indices t for which X_t exceeds u, the POT method models the random number $N_u = \text{card}(E_u)$ and the excesses $\{W_s = X_{t_s} - u\}_{t_s \in E_u}$ of all the extreme observations $\{X_{t_s}\}_{t_s \in E_u}$. In mathematical theory, Leadbetter (1991) proposes threshold u to be chosen as the independent modeling of

- The number of exceedances $N_u \sim \text{Poisson}(\lambda)$ as the limiting distribution of the sum of T Bernoulli random variables with success probability λ / T, and
- The excesses $W_1, \ldots, W_{N_u}$, conditional on the number of exceedances $N_u = n_u$, as a sample from the generalized Pareto distribution (GPD)

$$G(w; \sigma, \xi) = \begin{cases} 1 - (1 + \xi w/\sigma)_+^{-1/\xi}, & \xi \neq 0 \\ 1 - exp(-w/\sigma), & \xi = 0 \end{cases},$$

where $\sigma > 0$, and the support is $w \geq 0$ when $\xi \geq 0$ and $0 \leq w \leq -\sigma / \xi$ when $\xi < 0$. The domain of attraction of the GPD includes many common distributions which can be classified based on their tail characteristics, or equivalently, on the shape parameter ξ. The useful distributions for financial applications are for positive ξ which corresponds to heavy tailed (or power-tailed) distributions, e.g., Pareto, Student t or Fréchet. The case $\xi = 0$ corresponds to distributions whose tail essentially decays exponentially, e.g., normal, gamma or lognormal, while $\xi < 0$ is for short-tailed distributions with a finite right endpoint, like uniform or beta.

The advantage of the Block Maxima approach is that it satisfies the assumption of independence. Meanwhile, the disadvantage is that it may include data points which should not be considered extreme due to lack of any extreme value in a given block. Additionally, the method might miss an extreme value because another more extreme already occurred within the same block. By incorporating the Peak-over-Threshold model the issues described can be made redundant. The Peak-over-Threshold model is therefore considered to be an alternative and more suitable model if the whole data set is available (as opposed to having only extreme value data).

Threshold Selection

Inference consists of fitting the generalized Pareto family to the observed threshold exceedances, followed by model verification and extrapolation. This approach contrasts with the block maxima approach through the characterization of an observation as extreme if it exceeds a high threshold. But the issue of threshold choice is analogous to the choice of block size in the block maxima approach, implying a balance between *bias* and *variance*. In this case, too Iow a threshold is likely to violate the asymptotic basis of the model, leading to bias; too high a threshold will generate few excesses with which the model can be estimated, leading to high variance.

The standard approach suggested by Coles (2001) is to select the lowest possible threshold which would provide a reasonable approximation to the proposed model. There are several ways of setting a threshold: Rule of thumb, Graphical approach by using the Mean Residual Life Plot or Parameter Stability Plot.

By using a *rule of thumb* to choose the k largest observations, commonly used is the 90th percentile. Others proposed are as $k = \sqrt{n}$, and $k = n^{(2/3)}/log(log(n))$. All of these are practical but to some level theoretically improper [Scarott et al., 2012].

The POT method picks up all “relevant” high observations. The BM method on the one hand misses some of these high observations and, on the other hand, might retain some lower observations. Hence the POT seems to make better use of the available information.

Maximum Likelihood Estimation

Asymptotic normality of maximum likelihood estimators in regular parametric models is a classic subject, but when the support of the distribution depends on the parameter, the mathematics becomes harder. The family of univariate, three-parameter Generalized Extreme-Value distributions is a case in point.

Let a r.v. X that has the generalized extreme value d.f. with parameter vector $\theta = (\xi, \mu, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$,

$$H_\theta(x) = H_{\xi;\mu,\sigma}(x) = exp\{-\,(1 + \xi\,\frac{x-\mu}{\sigma})^{-1/\xi}\},$$

with the corresponding pdf as follows:

$$h_\theta(x) = h_{\xi;\mu,\sigma}(x) = \begin{cases} \frac{1}{\sigma}(1 + \xi\frac{x-\mu}{\sigma})^{-1/\xi-1}\, exp\{-(1 + \xi\frac{x-\mu}{\sigma})^{-1/\xi}\}, for\ 1 + \xi\frac{x-\mu}{\sigma} > 0 \\ 0, \quad elsewhere \end{cases}.$$

The *Likelihood function* based on data $X = (X_1, \ldots, X_n)$ is given by

$$L(\theta; X) = \prod_{i=1}^{n} h_\theta(x_i) = \prod_{i=1}^{n}\frac{1}{\sigma}(1 + \xi\frac{x_i-\mu}{\sigma})^{-1/\xi-1}\, exp\{-(1 + \xi\frac{x_i-\mu}{\sigma})^{-1/\xi}\}$$

$$= \sigma^{-n}[\prod_{i=1}^{n}(1 + \xi\frac{x_i-\mu}{\sigma})]^{-1/\xi-1}\, exp\{-\sum_{i=1}^{n}(1 + \xi\frac{x_i-\mu}{\sigma})^{-1/\xi}\},$$

For $1 + \xi\frac{x_i-\mu}{\sigma} > 0\ for\ all\ x_i, and\ 0\ elsewhere.$ (*)

The corresponding *log-likelihood* is:

$$\ell(\theta; X) = -n\ ln\ \sigma - (1 + \xi)\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} e^{-y_i},\ where\ y_i = \xi^{-1}ln(1 + \xi\frac{x_i-\mu}{\sigma}).$$

By definition, the maximum likelihood estimator (MLE) $\hat{\theta} = (\hat{\xi}, \hat{\mu}, \hat{\sigma})$ for the unknown parameters θ = (ξ, μ, σ) equals:

$$\hat{\theta} = arg\ \max_{\theta\in\Theta} \ell(\theta; X).$$

As long as the estimation of large quantiles x_p is concerned, the equi-variance property of maximum likelihood implies that the MLE of a quantile is obtained by substitution of the (MLE) $\hat{\theta} = (\hat{\xi}, \hat{\mu}, \hat{\sigma})$ into the quantile function (generalized inverse d.f.) of the generalized extreme-value distribution,

$$x_p = \mu - \frac{\sigma}{\xi}[1 - (-lnp)^{-\xi}].$$

Maximization of the pair of likelihood equations with respect to the parameter vector (μ, σ, ξ) leads to the maximum likelihood estimate with respect to the entire GEV family. There is no analytical solution, but for any given dataset the maximization is straightforward using standard numerical optimization algorithms.

At parameter combinations for which the above inequality (*) is violated, corresponding to a configuration for which at least one of the observed data falls beyond an endpoint of the distribution, the likelihood is zero and the log-likelihood equals minus infinity.

Subject to the limitations on ξ, the approximate distribution of $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ is multivariate normal with mean and variance-covariance matrix equal to the inverse of the observed information matrix evaluated at the maximum likelihood estimate. Though this matrix can be calculated analytically, it is easier to use numerical differencing techniques to evaluate the second derivatives, and standard numerical routines to carry out the inversion. Confidence intervals and other forms of inference follow immediately from the approximate normality of the estimator.

In general, deriving large sample asymptotics of the maximum likelihood estimator for a distribution family with varying support is a difficult problem. The classical regularity conditions of Cramer (1946) are not fulfilled, and the same is true for the weaker Lipschitz conditions stemming from empirical process theory (van der Vaart, 1998, Theorem 5.39).

Setting y_i, …, y_k to be only those values of x_1, …, x_n that exceed the threshold, the *GP log-likelihood* is given by:

$$\ell(\sigma_u, \xi;\ y_1, \ldots, y_k) = -\,k\ ln\ \sigma_u\ -\ (1\ +\ 1/\xi)\sum_{i=1}^{k} ln\,\{[1\ +\ \xi(\frac{y_i - u}{\sigma_u})]_+\},$$

where $y_+ = max\{y, 0\}$ so that any value y_i with parameter combination such that $1\ +\ \xi(y_i\ - u)/\sigma_u \leq 0$ would yield a log-likelihood of - ∞. Similar to the case for the Gumbel distribution, the shape parameter of the exponential one is also a single point in a continuous parameter space.

Smith (1985) was the first to consider maximum likelihood estimation in a large class of non-regular parametric families on the real line. Smith's paper may be applied to two particular distributions which are of importance in the analysis of extreme values. These are the generalized extreme value distribution, which includes the three-parameter, Weibull as a special case, and the generalized Pareto distribution introduced by Pickands (1975).

Smith (1985) studied this problem in detail and obtained the following results:

• when ξ > - 0.5, maximum likelihood estimators are regular, in the sense of having the usual asymptotic properties since the *information matrix* is finite.

• when -1 < ξ < - 0.5, MLE has a non-normal limiting distribution with rate of convergence $O(n^{-\xi})$.

• when ξ < - 1, maximum likelihood estimator is J-shaped, it is globally maximizing at sample maximum.

The case ξ < - 0.5 corresponds to distributions with a very short bounded upper tail. This situation is rarely encountered in applications of extreme value modeling, so the theoretical limitations of the maximum likelihood approach are usually no obstacle in practice.

Inference for extremes based on the parametric distributions given by the GEV equation is often based on likelihood methods, which provide a unified approach to modeling, estimation, and uncertainty assessment. The log-likelihood is easily obtained for independent data, and maximum likelihood estimates, and their standard errors may be obtained numerically by, for example, using routines provided in the packages *evd, ismev*, or *Pot* in the statistical computing environment R (R Development Core Team 2014). However, the regularity conditions that underpin classical large sample likelihood approximations hold only when $\xi > -1/2$ (Smith (1985). Uncertainty concerning high quantiles of the underlying distribution, termed return levels, is typically highly asymmetric; confidence intervals for these quantiles should be based on the *profile log-likelihood*.

Method of Probability-Weighted Moments (PWM)

Probability-weighted moments are generalization of the usual moments of a probability distribution, which give increasing weight to the tail information. The Gumbel distribution is a special case of the generalized extreme-value distribution, thus implying that the PWM may be useful in the generalized extreme-value d.f., too. The estimation of the parameters of the generalized extreme-value distribution by the method of PWM is described in detail by Hosking et al. (1985).

The probability-weighted moments of a r.v. *X* with distribution function F are the quantities:

$$M_{p,r,s} = E[\,X^p\{F(x)\}^r\{\,1\,-\,F(x)\}^s],\quad p,r,s\,\in\,\mathbb{R}.$$

Probability-weighted moments are likely to be most useful when the inverse d.f. $F^{\leftarrow}(.)$ can be written in closed form. Then the most convenient way of evaluating the moments is:

$$M_{p,r,s} = \int_0^1 \{F^{\leftarrow}(u)\}^p\, u^r (1\,-\,u)^s du$$

In the context of generalized extreme-value estimation we use the form:

$$\beta_r = M_{1,r,0} = E[X\{F(x)\}^r],\;\; r\,=\,0,1,2,\ldots$$

Given a random sample of size *n* from the distribution *F*, estimation of β_r is most conveniently based on the ordered sample $X_{1:n} \geq X_{2:n} \geq \ldots \geq X_{n:n}$. An unbiased estimator of β_r is provided by the statistic:

$$b_r = \frac{1}{n}\sum_{j=1}^{n}[\frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} X_{n\,-\,j\,+\,1:n}].$$

Other asymptotically equivalent and consistent estimators are provided by the formulae:

$$\hat{\beta}_r[p_{j,n}] = \frac{1}{n}\sum p_{j,n}^r X_{n\,-\,j+1:n},$$

Where $p_{j,n}$ is a plotting position, such as

$$p_{j,n} = \frac{j-a}{n}, with\; 0 < a < 1, or\; p_{j,n} = \frac{j-a}{n+1-2a}\; with\; -0.5\, < a\, <\, 0.5.$$

The estimators b_r are closely related to U-statistics, which are widely used in non-parametric statistics. Their desirable properties of robustness to outliers in the sample, high efficiency, and asymptotic normality may be expected to extend to the PWMs estimators b_r and other quantities calculated from them.

The PWMs for the generalized extreme value d.f. is:

$$\beta_r = \frac{1}{r+1}\{\mu\,-\,\frac{\sigma}{\xi}[1\,-\,(r\,+\,1)^{\xi}\,\Gamma(1\,-\,\xi)]\},\;\; for\;\xi\,<\,1\; and\;\xi \neq 0.$$

Note that the PWMs for generalized-extreme value distribution exist only for $\xi < 1$.

According to Hosking et al. (1985) the approximate PWM estimators for the parameters of the generalized extreme-value distribution are:

$$\hat{\xi}\;=\;7.8590c\;-\;2.9554c^2,\;\; where\; c\;=\;\frac{ln\,2}{ln\,3}\;-\;\frac{2b_1-b_0}{3b_2-b_0},$$

$$\hat{\sigma}\;=\;\frac{(2b_1-b_0)\hat{\xi}}{\Gamma(1-\hat{\xi})(2^{\hat{\xi}}-1)},\text{ and}$$

$$\hat{\mu}\;=\;b_0\;-\;\frac{\hat{\sigma}}{\hat{\xi}}\{\Gamma(1\,-\,\hat{\xi})-1.$$

The two methods are summarized below:

	Generalized Extreme Value (GEV)	*Generalized Pareto (GP)*
Description	Block Maxima/Minima Approach – Distribution Function of standardized maxima (or minima)	Peaks over Threshold Approach – Probability of exceeding pre-determined threshold
Parameters	Location μ: position of the GEV mean	Threshold u: reference value for which GP excesses are calculated
	Scale σ: multiplier that scales function, Shape ξ: Parameter that describes the relative distribution of the probabilities.	
General Function (CDF)	For extreme value *x*, $G(x) = exp[-\{1+\xi(\frac{x-\mu}{\sigma})\}_{+}^{-1/\xi}]$	For threshold excess *x*, $H(x) = 1 - [1+\xi(\frac{x-u}{\sigma_u})]_{+}^{-1/\xi}$
Limit as $\xi \to 0$	Gumbel: $G(x) = exp[-exp\{-(\frac{x-\mu}{\sigma})\}]$	Exponential: $H(x) = 1 - exp(-\frac{x-u}{\sigma})$
$\xi > 0$	Fréchet	Pareto
$\xi < 0$	Weibull	Beta
Interpretation of Results	Return Level: value x_p that is expected to be exceeded on average once every 1 / *p* period, where 1 – *p* is the probability associated with the quantile. Find x_p such that $G(x_p) = 1 - p$	Return Level: value x_m that is exceeded every *m* times. Begin by estimating ζ_u, the probability of exceeding the threshold. Then, x_m is $x_m = \begin{cases} u + \frac{\sigma_u}{\xi}[(m\zeta_u)^{\xi} - 1] & \xi \neq 0 \\ u + \sigma_u ln(m\zeta_u) & \xi = 0 \end{cases}$

Notes: (1) Weibull distribution has a finite right endpoint; Gumbel and Fréchet have infinite right endpoints. (2) GP function can be approximated as the tail of a GEV; the scale parameter σ_u is a function of the threshold and is equivalent to $\sigma_g + \xi\,(u - \mu)$, where σ_g, ξ and μ are all parameters of a corresponding GEV distribution.

Non-parametric Approaches and Tail Index Estimators

Both previous approaches are parametric (or semi-parametric), as they fit a parametric model (usually via *maximum likelihood estimation*) to the upper tail of the distribution (analogous results hold for the lower tail as well), though neglecting what happens at the center of the distribution. Anyway, if one does not want to fit a model to the tail either, the possibility exists to directly estimate the shape parameter ξ, pursuing a *non-parametric approach*.

The *non-parametric approach* does not assume that the observations of extremes follow exactly the asymptotic distribution.

Several estimators accomplishing this task have been proposed, but the most frequently employed one is by far the *Hill estimator*.

Hill Estimator

Hill proposed a consistent estimator, however, to the Fréchet case where $\xi > 0$. This estimator is applicable only to regularly varying tail. But it is the most popular tail index estimator. The original derivation of Hill estimator relied on the notion of conditional maximum likelihood estimation method.

In this context, when considering fat-tailed distributions, *the inverse of* ξ, namely $\alpha = 1/\xi$, is often the quantity of interest to be estimated. As previously, this quantity is known as the *tail index of the distribution* and is the exponent of the *power-law* that governs the decaying of the tail.

The problem we are facing here, can be simply described as follows: We have a data-set (i.i.d. sample) $X_1, X_2, \ldots, X_n$ (let $X_{1:n} \geq X_{2:n} \geq \ldots \geq X_{n:n}$ be the corresponding descending order statistics) from an unknown d.f. *F* and we are interested in the tail (upper or lower) behavior of the d.f. *F*, i.e. in the behavior of the extreme values.

After considering the order statistics $X_{1,n} \geq \ldots \geq X_{n,n}$ of the data, one can write the Hill estimator assuming positive losses, in order to take logarithms of the data, as

$$\hat{\alpha}^H_{k,n} = \left[\frac{\sum_{j=1}^{k}(\ln X_{j,n} - \ln X_{k,n})}{k}\right]^{-1},$$

i.e. the inverse of the average log-exceedance above the threshold $\ln X_{k,n}$.

The Hill estimator for the shape parameter $\xi > 0$, is obtained by inversion of the above equation:

$$\hat{\xi}^H_{k,n} = \left[\frac{\sum_{j=1}^{k}(\ln X_{j,n} - \ln X_{k,n})}{k}\right].$$

The estimator thus depends on the parameter $k \in \{2, \ldots, n\}$, which represents the cut-off between the observations considered as belonging to the center of the distribution and those pertaining to the upper tail, so that order statistics $X_{j,n}$ with $1 \leq j < k$ can be considered as extreme realizations.

The dependence of the Hill estimator on *k* is a critical issue for the application of the method to empirical studies.

From a theoretical viewpoint, the favorable consideration towards the Hill estimator is justified by its *asymptotic properties*, which are summarized in Embrechts et al. (1997, Theorem 6.4.6):

- *Weak consistency* – if $k \to +\infty$ and $k / n \to 0$ for $n \to +\infty$, then $\hat{\alpha}^H_{k,n} \xrightarrow{Pr} \alpha$
- *Strong consistency* – if $k / n \to 0$ and $k / \ln \ln n \to +\infty$ for $n \to +\infty$, then $\hat{\alpha}^H_{k,n} \xrightarrow{a.s.} \alpha$
- *Asymptotic normality* – under additional hypotheses, $\sqrt{k}(\hat{\alpha}^H_{k,n} - \alpha) \xrightarrow{d} N(0, \alpha^2)$.

Since the estimator: $\hat{\xi}^H_{k,n} = \frac{1}{\hat{\alpha}^H_{k,n}}$

can be used as an estimator of the shape parameter ξ, so that we will talk indifferently of the Hill estimator as estimating either the tail index or the shape parameter.

For the Hill estimator to have its claimed good *properties*, the number k of upper order statistics used for its calculation, which can be written as k(n) to show its dependence on the sample size n, should satisfy the following conditions:

$\lim_{n\to\infty} k(n) = \infty$, and take advantage of an increasing sample size, but

$\lim_{n\to\infty} \frac{k(n)}{n} = 0$, or the approximation at the population upper tail will not improve indefinitely.

Or, *k* should be large enough, but not too large, in other words *k* must moderate increase. The choice of *k* seems to be an art.

A simple, and sometimes effective, way to choose *k* is through the *Hill plot* of the pairs (*k*, $\hat{\xi}_{k,n}^{H}$), where $\hat{\xi}_{k,n}^{H}$ is the Hill estimator calculated based on *k* upper order statistics. An area of *k* where the graph is almost horizontal indicates a proper range of *k*-values to choose from. Still, as Embrechts et al. (1997) present, there are sometimes where the results of Hill-plots are totally misleading. As Drees et al. (2000) prove, traditional Hill-plot is most effective only when the underlying distribution is Pareto or very close to Pareto. For the Pareto distribution, one expects the Hill plot to be close to extreme-value index ξ in the right side of the plot (since Hill is the MLE). On the other hand, in the case of regularly varying d.f.'s, i.e., when Pareto distribution is only approximated in the tail, Hill is only an approximate MLE based on large observations, and it is less clear what portion of the plot is most accurate.

Apart from the likelihood approach, the Hill estimator can be derived via a regular variation approach (Embrechts, et al., 1997) as well as via a graphical approach using mean excess function or Pareto QQ plot (Beirlant, et al., 1996). In the *Pareto QQ Plot*, we utilize the fact that if 1-*F* is regularly varying with index $-\alpha$, the tail 1-*F* is of Pareto type for large *x*, i.e. $1- F(x) \sim x^{-\alpha}$, for $x \to \infty$. That implies that the corresponding Pareto QQ plot should be linear for $x \to \infty$ (ultimately linear as it is often referred to). It can be shown (Beirlant et al., 1996) that the slope of this line is equal to 1 / α, while an estimator of this slope for the upper part of the plot (for values of *x* larger that $X_{k+1:n}$), where linearity can be regarded valid, is the known Hill estimator. The same reasoning holds for the mean excess plot.

The Hill estimator often outperforms other estimators as far as we are concerned in the estimation of the tail index, i.e., in the case when our cumulative distribution function is in the MDA of a Fréchet distribution. Otherwise, the Hill estimator does not work, and one has to employ some other estimator. This is not usually a major issue, since many financial time series have an extreme value distribution of Fréchet type and the strong connection between the tail index and the parameters of stable laws and the Student-t distribution explains why the earliest applications of EVT to finance were especially focused on estimating the tail index.

Pickands Estimator

Studied by Pickands (1975), this estimator, unlike the Hill estimator, can be used to estimate the shape parameter of any of the three extreme value distributions.

By adopting a percentile estimation method, Pickands ends-up with the estimator for the parameter ξ:

$\hat{\xi}_{k,n}^{P} = \frac{1}{\ln 2} \ln \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}$.

Moreover, Pickands method provides us with an estimate of the scale parameter σ.

A more formal justification of Pickands estimator is provided by Embrechts et al. (1997). According to them, the basic idea behind the estimator is to find a condition equivalent to $F \in$ MDA($H_{\xi;\mu,\sigma}$), involving the parameter ξ in a straightforward way. The key-point is the characterization II property of MDA($H_{\xi;\mu,\sigma}$).

The asymptotic properties of the Pickands estimator are studied in Dekkers and de Haan (1989):

• Weak consistency – if $k \to +\infty$ and $k / n \to 0$ for $n \to +\infty$, then $\hat{\xi}_{k,n}^{P} \overset{Pr}{\to} \xi$

• Strong consistency – if $k / n \to 0$ and $k / \ln \ln n \to +\infty$ for $n \to +\infty$, then $\hat{\xi}_{k,n}^{P} \overset{a.s.}{\longrightarrow} \xi$

• Asymptotic normality – under additional hypotheses, $\sqrt{k}(\hat{\xi}_{k,n}^{P} - \xi) \overset{d}{\to} N(0, \nu(\xi))$

with ν (ξ) depending on ξ in a highly nonlinear way, as

$\nu(\xi) = \xi^2(2^{2\xi+1} + 1)/\{2(2^{\xi} - 1)ln2\}^2$.

Similar conditions as Hill estimator, hold for the choice of k.

A handier way that is used in practice for the choice of k is based on the Pickands-plot (Embrechts, et al., 1997). For each value of k = 1, ..., [n/4], we calculate the Pickands estimator and plot it against k. The range of *k's* that correspond to a "plateau" in the plot (i.e., those values of k that give very similar values of $\hat{\xi}_{k,n}^{P}$) are considered more appropriate, that is, we choose $\hat{\xi}_{k,n}^{P}$ from such a k-region where the plot is roughly horizontal.

Graphical Tools

- ***Order Statistics and Empirical Distribution***

For $x \in \mathbb{R}$ the *empirical df* or *sample df* is defined as:

$$F_n(x) = \frac{1}{n} card\{i : 1 \le i \le n, X_i \le x\} = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}, \ x \in \mathbb{R},$$

Where $\boldsymbol{I}_A$ denotes the indicator function of a set *A*.

- ***Probability and Quantile Plots***

Given a set of data to be analyzed, we usually start with histograms, boxplots, empirical df's plot.

Our problem is to find a df *F* which will be a good approximation for the i.i.d. data $X, X_1, \ldots, X_n$.

Definition: Probability Plot (PP-plot)

$$\{(F(X_{k,n}), \frac{n-k+1}{n+1)}) : k = 1, \ldots, n\},$$

Definition: Quantile Plot (QQ-plot)

$$\{(X_{k,n}, F^{\leftarrow}(\frac{n-k+1}{n+1)})) : k = 1, \ldots, n\},$$

Also note,

$F_n(X_{k,n}) = \frac{(n-k+1)}{n}$, where F_n = the empirical distribution

The approximate linearity of the above plots satisfied by the Glivenko-Cantelli theorem.

The theory of weak convergence of empirical processes forms the basis for the construction of confidence bands around the graphs, leading to hypothesis testing.

QQ-plots properties

- ✓ Comparison of distributions
- ✓ Outliers
- ✓ Location and scale
- ✓ Shape: if the reference distribution has heavier tails (tend to have larger values) the plot will curve down at the left and / or up at the right.

- ***Zipf Plot***

The *Zipf Plot (or Log-Log Plot)* is probably the most used and abused plot for verifying the presence of paretianity in the data. The original plot was constructed on binned observations. Here we use a different version based on the empirical survival function, i.e., a plot in which the logarithm of the empirical survival function is plotted against the logs of the ordered values of *x*. If the data follow a power law, we expect to observe a more or less negative linear relationship in the graph.

More specifically,

$$\bar{F}(x) = (\frac{x}{x_0})^{-\alpha},\ 0 < x_0 \leq x \quad \rightarrow \quad log(\bar{F}(x)) = \alpha \underset{C}{log}(x_0)\ -\ \alpha log(x)$$

From a heuristic point of view a ZIPF Plot can be used to identify the threshold value above which paretianity seems to hold.

- ***Mean Excess Function***

A useful graphical tool for tails discrimination and the selection of the threshold *u*.

The *mean excess function* gives the expected excess of a random variable over a certain threshold given that this random variable is larger than the threshold.

In financial risk management context, switching from the right tail to the left tail, $e(u)$ is referred to as *shortfall*.

Let *X* be a random variable with distribution function *F* and right endpoint x_F, i.e.,

$$x_F = sup\{x \in \mathbb{R} :\ F(x) < 1\}\ .$$

The function

$$e(u) = E[X\ -\ u|X > u] = \frac{\int_u^\infty (t-u)dF(t)}{\int_u^\infty dF(t)},\ \ 0 < u < x_F,$$

is called *mean excess function* of *X*.

From an empirical point of view, the mean excess of a sample $X_1, X_2, \ldots, X_n$ is easily computed as

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)}{\sum_{i=1}^n \boldsymbol{I}_{\{X_i > u\}}}\ ,$$

that is the sum of the exceedances over the threshold *u* divided by the number of data points exceeding *u*, and ***I*** is the indicator function.

Using the definition of e(u) and partial integration we get

$$e(u) = \int_u^{x_F} (x-u)dF(x)/\bar{F}\,(u)$$

$$= \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x)\,dx,\ \ 0 < u < x_F.$$

Whenever *F* is continuous,

$$\bar{F}(x) = \frac{e(0)}{e(x)} exp\{-\ \int_0^x \frac{1}{e(u)} du\},\ \ x > 0.$$

From the above equation follows that a continuous df is uniquely determined by its mean excess function.

If $\bar{F} \in \mathcal{R}_{-\alpha}$ for some $\alpha > 1$, by Karamata's theorem it yields $e(u) \sim u \,/\, (\alpha - 1) \;\; as\, u \to \infty$.

Hence, if we create a graph, in which the points $\{(X_{i,n}, e_n(X_{i,n})) : i = 1, \ldots, n\}$ are plotted,

$X_{1:n}$, $X_{2:n}$, …, $X_{n:n}$ being the order statistics of our data set, what we obtain is called *mean excess plot*, or MEPLOT. Such a plot will not show an infinite slope, rendering the interpretation of such plots problematic for very heavy tailed phenomena.

Regularly varying distributions with tail index $\alpha > 1$, have a mean excess function which is ultimately linear with slope $1 / (\alpha-1)$. If $\alpha < 1$, then the slope is infinite and is not useful.

Given the properties of the mean excess function for Paretian random variables, a MEPLOT showing a linearly increasing trend can be considered a signal of paretianity in the data.

More specifically, long tailed d.f.'s exhibit an upward sloping behavior, exponential type d.f.s are characterized by a constant mean excess plot, whereas the plot for short-tailed data shows a decreasing pattern. A graph representation of $e_n(u)$ is constructed with the mean excess on the vertical axis. This graph can be interpreted as follows: if the slope is rising, it indicates thick tail behavior; but if there is a downward trend, this shows thin tail behavior in the distribution; and finally, if the slope of the line is equal to zero, the behavior of the tail is exponential. If the line is straight and has a positive slope located above the threshold, then it is an example of Pareto-type tail behavior.

Confidence intervals are included in the plot to measure the uncertainty in the estimates of the theoretical sample means which follow approximately a normal distribution. By examining the *mean residual life plot*, the threshold *u* is to be chosen so that there is a linear line up to the chosen threshold before it decays sharply.

- ***Return Level/Period***

The quantiles of the GEV df are of particular interest because of their interpretation as *return levels*. The value expected to be exceeded on average once every 1 / p period, where 1 – p is the specific probability associated with the quantile.

Two common interpretations of a return level with a return period of T are the *waiting time* which is the waiting time until next occurrence of event is T for example years. The second is the *number of events*, or the average number of events occurring within a T – time period is one.

The *return period of events* $\{X_i > u\}$ is defined:

$EL(u) = p^{-1} = (\bar{F}(u))^{-1}$, which increases to ∞ as $u \to \infty$.

A *return level* with a *return period* of $T = 1 / p$ time units is a high threshold $x(p)$ whose probability of exceedance is *p*. For example, if $p = 0.01$, then the return period is $T = 100$-time units.

The return level can be used as a measure of the maximum loss of a portfolio, a rather more conservative measure than the Value-at-Risk.

The quantities of interest are not the parameters themselves, but the quantiles, called return levels, of the estimated GEV or GPD:

$R^p = H^{-1}_{\xi,\sigma,\mu}(1 - \frac{1}{p})$,

Substituting the parameters ξ, σ and μ by their estimates $\hat{\xi}, \hat{\sigma}$ and $\hat{\mu}$, we get:

$$\hat{R}^p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}(1 - (-log(1 - \frac{1}{p}))^{-\hat{\xi}}) & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}\, log\,(-log(1 - \frac{1}{p})) & \hat{\xi} = 0 \end{cases}.$$

A value of $\hat{R}^{10}$ of 7 means that the maximum loss observed during a period of one year will exceed 7% once in ten years on average.

The *return level plot* is derived and graphically depicts different shape parameters. The plot is linear when ξ= 0, convex when ξ< 0 and concave when ξ> 0. The dashed lines in the plot are the confidence intervals which are increasing at larger return levels implying more uncertainty and risk are incorporated at higher return levels.

- ***The Ratio of Maximum and Sum***

A simple tool for detecting heavy tails of a distribution and for giving rough estimates of the order of its finite moments.

Suppose X, X_1, X_2, … are i.i.d. random variables and define for any positive r the quantities:

$S_n(r) = |X_1|^r + \ldots + |X_n|^r$,

$M_n(r) = max(|X_1|^r + \ldots + |X_n|^r)$, with $n \geq 1$, and $M_n = M_n(1)$ and $S_n = S_n(1)$.

Writing $R_n(r) = \frac{M_n(r)}{S_n(r)}$, $n \geq 1, r > 0$, and plotting $R_n(r)$ against n for a variety of r-values we can obtain some preliminary information about $Pr(|X| > x)$.

For large n, if $R_n(r)$ is small provided that $E|X|^r < \infty$. On the other hand, if there are significant deviations of $R_n(r)$ from zero for large n, this is an indication for $E|X|^r$ being infinite.

The limit behavior of M_n / S_n and the corresponding quantities for X_i^r can be used as an explanatory statistical tool for detecting whether EX^r is finite.

Modeling Financial Returns

Financial time series have some characteristic properties and shapes given by the microstructure of the financial market. The basic feature of the financial time series is the high frequency of individual values. This leads to the broadening of the influence of nonsystematic factors to the dynamism of these time series. The result is relatively high volatility which usually changes over time. The systematic factors create a trend and cycle part of time series. Leptokurtosis is usually observed in empirical distributions of price changes.

Financial time series are characterized by the following features:

(1) *clustering of volatility* – large price changes tend to be followed by large price changes and small price changes tend to be followed by small price changes.

(2) *autoregressive behavior* – price changes depend on price changes in the past, e.g., positive price changes tend to be followed by positive price changes.

(3) *skewness* – there is an asymmetry in the upside and downside potential of price changes.

(4) *fat tails* – the probability of extreme profits or losses is much larger than predicted by the normal distribution. Tail thickness varies from asset to asset.

(5) *temporal behavior of tail thickness* – the probability of extreme profits or losses can change through time; it is smaller in regular markets and much larger in turbulent markets.

(6) *tail thickness varies across frequencies* – high-frequency data tends to be more fat-tailed than lower-frequency data.

Mandelbrot (1963) and Fama (1965), had the idea to look for the probability distribution which would catch the properties of the financial time series better than the normal distribution originated many years ago. Both suggested applying stable distributions. As a special case the normal distribution belongs into this class of distributions. This class of distributions was described by Lévy (1924). The characteristic property of stable random variables is that their sum has also stable distribution. The non-normal stable distributions catch the high kurtosis, the non-symmetric shape, and the fat tails of the distributions of the financial time series log returns much better than the normal distribution. This is very closely connected with the properties of the non-normal stable distributions which have infinite second and higher moments. The sample variance and kurtosis of the data generated by the non-normal stable distribution do not converge with the growing sample size.

Mandelbrot's hypothesis states that for distributions of price changes in speculative series, parameter α is in the interval $1 < \alpha < 2$, so that the distributions have means but their variances are infinite.

Heavy or fat tailed distributions?

A *heavy tailed distribution* is defined as a distribution with a tail that is heavier than the exponential distribution, which formally is given by:

Let X be a random variable with cumulative distribution function (cdf) F and $\bar{F}(x) = 1 - F(x)$ is its survival function, then:

$$\forall \lambda > 0, \lim_{x \to +\infty} \bar{F}(x) e^{\lambda x} = +\infty.$$

Examples of heavy tail distributions are the log-normal distribution, the Pareto, the Student's-t, Fréchet, stable or tempered stable distribution.

Distributions with a probability density function (pdf) that behaves like a power law at infinity are sometimes called *fat-tail distributions*.

A heavy tail distribution might have moments of any orders, whereas a fat-tail distribution will have infinite moments at some point.

In this study, we consider distributions that exhibit large *skewness* and *kurtosis* compared with the normal distribution.

Under the assumption of normality, the distribution is characterized entirely by its first two moments. Considering the heavy tail property we need knowledge of higher order moments, particularly the third moment i.e., skewness as a measure of the distribution's asymmetry, and the fourth moment, i.e., kurtosis accounting for the fatness of the tail.

We note that there are other classes of models that have been proposed for financial returns: generalized t-distributions, generalized hyperbolic, generalized inverse Gaussian, geometric stable, tempered stable, etc. While these models can give a good fit to data sets, they lack all the features described above. A different approach is to use extreme value theory as described in Embrechts et al. and McNeil et al.

For many real assets, plots of the returns show a *unimodal*, *mound shaped* distribution. We will show below that some of these distributions have heavier tails than a normal model, and we use a stable distribution to describe the returns.

The relevant probability distribution of future outcomes is the most important assumption in forecasting financial data such as asset returns, inflation percentages, and interest rates.

In most theoretical and empirical work, a normal or lognormal distribution is assumed. However, an important drawback is that the tails of the normal distribution decay exponentially toward zero. As a result, large realizations are very unlikely. This property contradicts the empirical findings on asset returns, which state that these returns generally exhibit *leptokurtic behavior* (they have *fatter tails* than the normal distribution). This implies that extreme returns of either sign occur far more often in practice than predicted by the normal model.

The three fat-tailed models that have received most of the attention are the stable and Student's t distributions and the ARCH process. Embrechts, Kluppelberg and Mikosch (1997) show that the Fréchet distribution is consistent with GARCH models of financial time series which generate Pareto-like distributions.

If one concentrates on the distribution of the extremes, the main parameter of concern, i.e., the tail index, is nested. The tail index is a good indicator for the mass in the tails, and therefore provides a direct measure for how leptokurtic stock market returns are. Hence, the estimation of the tail index of the extreme distribution provides information about the underlying distribution and thereby allows for discrimination between the alternative hypotheses. Moreover, given the value of the tail index, one can calculate the bounds on the returns for low probabilities of an excess.

In this thesis, we applied some of the theory presented above, such as the Block Maxima and POT approaches. We also applied parameter estimation methods such as the MLE method, which uses a parametric approach, and the Hill estimator, which uses a non-parametric approach. The goal is to fit our maximum and minimum return series to the GEV and GPD distributions.

The first step is to block the data into sequences of *n* observations, *n* being sufficiently large.

Then the maxima/minima X_i of each block I is calculated and finally the GEV distribution is fitted to this series of block maxima X_1, X_2, …

Once the GEV distribution has been fitted, we can calculate the quantile function, z_p (or x_p) for the weekly/monthly distribution as:

$$z_p = \begin{cases} \mu - (\sigma/\xi)(1 - (-log(1 - p))^{-\xi} & for\ \xi \neq 0 \\ \mu - \sigma log(-log(1 - p)) & for\ \xi = 0 \end{cases}.$$

Notice that $G(z_p) = 1 - p$. This is the *return level* associated with the *return period* 1 / *p*. That is, z_p is the level that is expected to be exceeded, on average, once every 1 / *p* weeks/months. Equivalently, z_p is the level that is exceeded by the weekly/monthly maximum in any particular week/month with probability *p*.

By using $y_p = -\log(1 - p)$, this quantile function can be expressed as

$$z_p = \begin{cases} \mu - (\sigma/\xi)(1 - y_p^{-\xi}) & for\ \xi \neq 0 \\ \mu - \sigma log y_p & for\ \xi = 0 \end{cases}.$$

If z_p is plotted against $\log y_p$ the plot is *linear* in the case of $\xi = 0$, *convex* in the case of $\xi < 0$ with asymptotic limit as *p* tends to 0 to $(\mu - \sigma)/\xi$ and *concave* for $\xi > 0$ and has infinite bound.

Data-Set Choice

The speculative time series in finance generally displays two empirical properties. The first property is that log (P_t), where P_t is asset's price, exhibits the *martingale property*. The other feature is that the variability of the R_t (asset's return) exhibits a certain *clustering* which makes the variance somewhat more predictable. The hypothesis that logarithmic speculative prices form a martingale series was first raised by Bachelier (1900).

The *martingale property* of the random return series {R_t} derives from economic insight. The returns {R_t} are regarded as martingale innovations, a property which is characterized by the condition that the conditional expectation of the future return *R*, given the past returns value r_{t-1}, ..., r_0 with r_0 = log p_0 (or equivalently, given the prices p_{t-1}, ..., p_0) is equal to zero. Then the returns are necessarily uncorrelated.

When considering asset price returns, we first consider these observations or data to be simply a time series, or a continuous time *stochastic process*. Each instant of time *t* is assigned a real random variable p_t. A usual assumption is that (p_t) or a transformed version of it (for instance, the first-order differences or the log-differences) forms a *stationary process* (strictly stationary or stationary in the wide sense).

Following Longin, (1999), a time series of percentage logarithmic price changes is computed. The price change *ΔP* is defined by $\Delta P_t = 100 * \text{Log}\ [P_t / P_{t-1},]$ where *P*, is the closing price on day *t*. This definition presents several advantages: It provides the econometrician with a stationary time series; it is independent of the unit of measurement; and it is stable under time-aggregation.

The issues concerning the frequency of the data employed in empirical studies and the choice of an appropriate time window for the observations play a significant role when applying EVT. EVT requires a lot of data, since its results are asymptotical, but, on the other hand, it is necessarily faced with a scarcity of data, given that it concentrates on the tails of the distribution and extreme events are, by definition, rare.

The return series utilized in this thesis are the daily closing prices of some representative local equity and bond indices. We consider both the left and the right tail of the return distribution, as the left tail represents losses for an investor with a long position on the index, while the right tail represents losses for an investor being short on the index.

Our primary aim is to fit extreme value distributions to the data. Specifically, we ask is the tail index α < 2, or is $\alpha \geq 2$?

First, we consider the distribution of the block maxima, which allows the determination of the return level. Second, we model the exceedances over a given threshold which enables us to estimate high quantiles of the return distribution.

The parameters of the models are estimated by maximum likelihood method primary, and secondary we use the method of L-Moments. The fitting results are tested by probability plot, quantile plot, density and return level plot.

Concerning the Negative log-likelihood estimates, values that are closer to 0 will indicate a better model fit. AIC (Akaike information criterion) and BIC (Bayesian information criterion) parameters measures the

relative quality of the statistical models, taking into account the tradeoff between the complexity of the model and the goodness of fit (rewards goodness of fit and penalizes increased number of parameters).

AIC is estimated as $AIC = 2k - 2\ln(L)$ and respectively $BIC = (\ln(n) - \ln(2\pi))\, k - 2\ln(L)$

where k is the number of parameters, n is the number of data points, and L is the maximized value of the likelihood function. When comparing models, a model with a smaller AIC/BIC value is considered to be a “better” model.

Data Series Description

We proceed to the application of the GEV and GPD distributions to our database, which is composed of the weekly and monthly maximum and minimum log returns of the ASE Index, an equity series and BEGCGA Index, a bond data series. The goal is to fit our maximum and minimum return series to the GEV and GPD distributions. The weekly and monthly maximum and minimum refers to the maximum and minimum return, extracted from the daily log return, in each week and month for each index.

Both financial series have been tested for the null hypothesis of being normal and have been rejected under Jarque-Bera test.

The methodology applying to right tails, in the left tail case we take the absolute of the returns so that positive values correspond to losses.

The data that we considered was downloaded from the Bloomberg terminal. For the application of the methodology, we used a diversity of tools such as the language R and its extRemes toolkit. Another tool that we used was Matlab, which is a powerful tool in terms of graphs.

In the below table we can see the main characteristics of our data series.

Table 3: Data Series Description

BLP Ticker	Name	Currency	Description
			Equity Index
ASE Index	Athens stock Exchange General Index	EUR	Is a capitalization-weighted index of Greek stocks listed on the Athens Stock Exchange.
			Bond Index
BEGCGA Index	Bloomberg Series-E Greece Govt All > 1Yr Bond Index	EUR, Unhedged	Bloomberg Series-E Greece Govt All > 1 Yr Bond Index

Source: Bloomberg

Application 1 >>> ASE INDEX

In this application we use daily data prices from March 31, 1987, to December 31, 2021, with a total of 8,635 price observations.

Firstly, we proceed to an exploratory data analysis, followed by the application of the Block Maxima and POT approach.

Time Series Plot and Descriptive Statistics

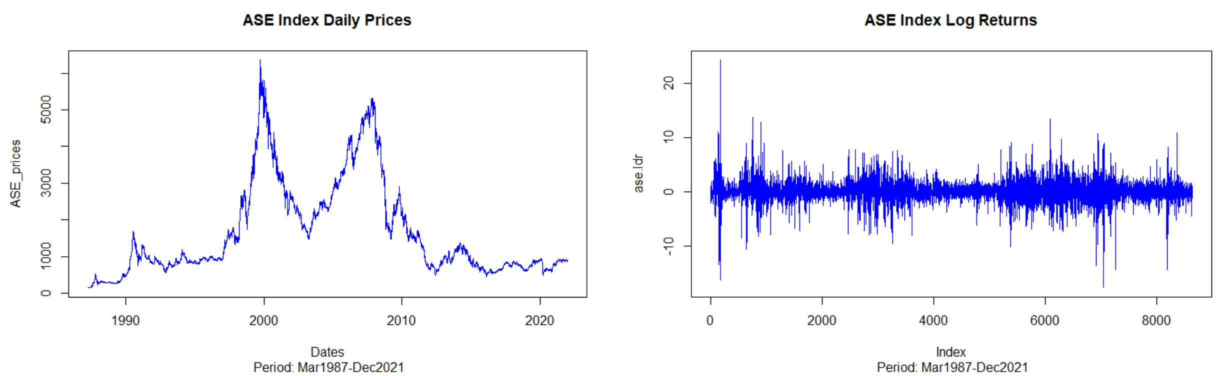

Figure 6. Left Panel: Daily closing prices of the ASE Index. Right Panel: Log-daily returns of the ASE Index

Figure 6 shows the daily closing prices of the ASE Index for almost 34-years period. The right panel of the above figure is a transformation to log-daily returns. It seems reasonable to assume that the data is stationary over the observation period.

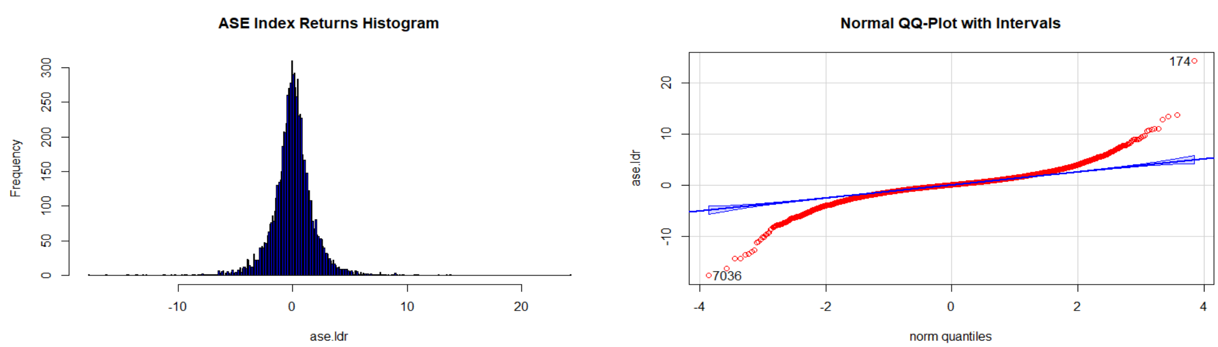

Figure 7. Left Panel: Histogram of ASE Index. Right Panel: Normal QQ-Plot of ASE Index

From the above figures we can see that the data are peaked in middle. QQ-Plot reveals a distribution with fat tails since both ends of the plot deviate from the straight line and its center follows the straight line. We can view in this QQ Plot that this distribution exhibit "heavy tails" versus a normal distribution. This is an over-dispersed distribution, also known as having a leptokurtic distribution with positive excess kurtosis. Notice that the points along a line in the middle of the graph curve off in the extremities. Normal Q-Q plots that exhibit this behavior usually mean that our data have more extreme values than would be expected if they truly came from a normal distribution.

To start, table 4 shows the summary statistics of the daily log returns, weekly and monthly maximum, and minimum log return series of the ASE Index, as also the stats of the left and right tails.

Looking at table 4, it is possible to see that in the daily return the stock index has a mean of 0.02% and a standard deviation of 1.93%, with a maximum of 24.23% and a minimum of 17.71%. In terms of the maximum returns, the weekly series has a mean of 1.84% and a standard deviation of 1.80%, with a maximum return of 24.23% and a minimum of 7%. Concerning monthly maxima, we distinguish a mean of 3.43% with 2.37% standard deviation, 24.23% maximum and 0.67% minimum. In the minimum return space, the weekly series has a mean of -1.78% with a standard deviation of 1.85%, and a minimum of -17,71% and a maximum of 0.00%. Concerning monthly minima, the mean is -3.30% with 2.43% standard deviation, the minimum is 17.71% and the maximum is 0.00%. Looking at left and right tails statistics, we notice the more volatile the left part with a maximum less than the other tail.

Table 4. ASE Index – Explanatory Data Analysis

		Weekly		*Monthly*		*Daily*	
	All Data	*Minima*	*Maxima*	*Minima*	*Maxima*	*Left Tail*	*Right Tail*
N (obs)	8,635	1,809	1,809	417	417	4,192	4,398
Mean Return	0.0201	-1.7824	1.8425	-3.301	3.4363	1.3077	1.2859
Variance	3.7100	3.4372	3.2313	5.9246	5.6323	2.1452	1.9568
St. Dev	1.9261	1.8540	1.7976	2.4340	2.3732	1.4646	1.3989
Min	-17.7129	-17.7129	-7.3717	-17.713	0.6771	0.0054	0.0055
1st Quartile	-0.8296	-2.3786	0.7612	-4.134	1.9064	0.3894	0.4007
Median	0.0328	-1.3339	1.4410	-2.578	2.8134	0.8576	0.8730
3rd Quartile	0.8945	-0.6617	2.4668	-1.662	4.1409	1.7199	1.6794
Max	24.2296	3.1908	24.2296	0.0000	24.2296	17.7129	24.2296
IQR	1.7241	1.7169	1.7056	2.4728	2.2345	1.3305	1.2788
Mode	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Skewness	-0.0801	-2.5800	2.7286	-2.2939	2.8428	3.2376	3.4665
Kurtosis	13.0421	12.1130	18.7541	7.5434	15.5377	18.3570	26.5591
J-B Test (p-value)	36,292 (0.000)	13,100 (0.000)	28,827 (0.000)	1,370.4 (0.000)	4,753.00 (0.000)	66,253 (0.000)	138,207 (0.000)
ADF Test (p-value)	-18.511 (0.01)	-8.917 (0.01)	-7.2113 (0.01)	-5.2531 (0.01)	-4.7485 (0.01)	-14.736 (0.01)	-11.567 (0.01)

Notes: (1) J-B normality testing, (2) ADF stationarity testing
(3) The sign of the left tail has been changed so that positive values correspond to losses

From the descriptive statistics of all data, we can see a *slightly negative skewness* of our full dataset. We can also verify this since the mean of the series is less than the median which indicates a negative skewness. Also, there is a big *positive value of kurtosis* which corresponds to the great extremity of the data, or to a distribution with fat tails.

We *reject the normality assumption* based on the J-B test since the associated probability of the JB test is lower than 0.05 (the default significance level assumed for all tests in this thesis).

Also, *ADF test* indicates *rejection of the unit-root* null model in favor of the alternative model, that is the series is stationary.

Fitting the GEV Distribution in ASE Index Log Returns

First, we conduct a *Likelihood Ratio (LR) test*, for all extreme data series, to conclude if the Gumbel distribution assumption is rejected or not, depending on the probability associated with the test. If it is lower than 0.05, we reject the null and the Gumbel distribution assumption; otherwise, we can admit the Gumbel distribution based on the considered sample. According to the below table, we reject the null hypothesis. This means that we do not assume that the series follows a Gumbel process. Therefore, we can conclude that none of the maximum/minimum return series have been generated by a Gumbel distribution. This means that for the considered series, only the Fréchet or Weibull particular cases of the more general GEV distribution remain.

Table 5. LR test for Gumbel distribution

	Likelihood Ratio Test	*p-value*	*χ^2 – critical value*	*Alternative*
Weekly Min Return	11.136	0.0008	3.8415	Fréchet
Weekly Max Return	141.75	0.0000	3.8415	Fréchet
Monthly Min Return	62.859	0.0000	3.8415	Fréchet
Monthly Max Return	51.66	0.0000	3.8415	Fréchet

Notes: *r-packages "extRemes", "fExtremes", "extremeStat"*

The shape parameter of the Gumbel df (i.e., $\xi = 0$) represents a single point in a continuous parameter space, and there is zero probability of obtaining an estimated shape parameter that is exactly zero. It is possible to fit the Gumbel df separately and test the hypothesis of $\xi = 0$ versus $\xi \neq 0$ but it is arguably preferable to always allow the shape parameter to be nonzero even if such a test supports the Gumbel hypothesis. This can be performed using the likelihood-ratio test (Coles 2001) or by checking *whether or not zero falls within the confidence intervals for ξ*. The latter approach does not require fitting the Gumbel df to the data.

Studying each series individually now and starting with the weekly maximum and minimum series of the ASE Index, following with the monthly maxima and minima we found the optimal values for the parameter estimates, which we present in the next table. To find the parameters of the GEV we use the Maximum Likelihood Estimation (MLE) and the L-moments estimation.

Table 6 summarizes point estimates and the single confidence intervals for the GEV distribution for the ASE Index. Given the positive estimates for the shape parameter, for all our dataset, we conclude that the Fréchet distribution seems to be the most appropriate to describe that kind of data. Confidence intervals for shape parameters reveal that we can accept the alternative hypothesis of the upper unbounding type of distribution, in the confidence level of 0.95%. Consequently, we can conclude that ASE Index return empirical distribution has fat tails. The larger the estimate values for the shape parameter, the more fat-tailed the distribution.

The method of L-moments, also called as the method of probability-weighted moments (PWM), is the linear combinations of the expected values of order statistics. The first L-moment (λ1) is the mean value identical to the first ordinary product moment. The second L-moment (λ2) is a measure of scale or dispersion analogous to standard deviation, and the third (λ3) and fourth (λ4) L-moments are measures of skewness and kurtosis, respectively (DHI, 2007). L-moments estimates of the distribution parameters are computed by substituting the theoretical L-moments for the specified distribution by the L-moment estimators.

The GEV distribution, according to Table 6, yields slightly different results for the method of ML and L-moment.

Based on shape parameter estimation with the MLE method, we can conclude that the tail index *α* will be above 3, or we expect at least the existence of the first 3 moments.

Concerning the return levels, for p = 10, we obtain for our monthly minima $\hat{R}^{10} = 6.03$, meaning that the maximum loss observed during a period of one month exceeds 6.03% in one out of 10 months on average. In the same way we can derive that a loss of $\hat{R}^{50} = 10.91\%$, is exceeded on average only once in 50 months.

The parameters of the models and the fitting results are tested by probability plot, quantile plot, return level plot and density plot.

The *Profile Likelihood function* is introduced to analyze the uncertainty of returns extreme inference and the theory of estimating confidence intervals of the key parameters and quantiles of extreme value distribution. As an effective tool for estimating confidence interval of the key parameters and quantiles of extreme value distribution, profile likelihood function can lead to a more accurate result and help to analyze the uncertainty of extreme values. It is generally important to inspect the plot to be sure that the resulting confidence intervals are not merely the endpoints used.

The shape parameter estimation and the 20-period return level and confidence intervals respectively are then calculated by profile likelihood function. Because the profile-likelihood clearly crosses both vertical dashed lines (where they cross the blue horizontal line), the resulting intervals are believable.

Table 6. ASE Index Block Maxima Method

	Weekly Min Returns	*Weekly Max Returns*	*Monthly Min Returns*	*Monthly Max Returns*
		Method MLE		
Negative Log-Likelihood Value	2,916.113	2,939.227	817.0974	798.1728
Location (SES)	0.9796 (0.0237)	1.0613 (0.0245)	2.1397 (0.0686)	2.2980 (0.0669)
CIs	[0.9332-1.0260]	[1.0132-1.1094]	[2.0053-2.2740]	[2.1669-2.4291]
Scale (SES)	0.8555 (0.0204)	0.8919 (0.0206)	1.2458 (0.0572)	1.1777 (0.0569)
CIs	[0.8155-0.8956]	[0.8516-0.9322]	[1.1336-1.3580]	[1.0662-1.2893]
Shape (SES)	0.3352 (0.0247)	0.2834 (0.0235)	0.2779 (0.0399)	0.3098 (0.0474)
CIs	[0.2869-0.3836]	[0.2374-0.3294]	[0.1996-0.3562]	[0.2169-0.4024]
AIC	5,838.227	5,884.455	1,640.195	1,602.346
BIC	5,854.728	5,900.956	1,652.294	1,614.438
K-S test (two-sample)	0.01572 (0.9673)	0.01572 (0.9673)	0.032506 (0.7704)	0.0388 (0.5583)
		Return Level for Period (*Units*)		
10-units [CI's]	3.85 [3.63-4.07]	3.87 [3.67-4.07]	6.03 [5.46-6.61]	6.13 [5.54-6.72]
20-units [CI's]	5.34 [4.93-5.74]	5.22 [4.86-5.57]	7.89 [6.93-8.85]	8.04 [6.99-9.08]
50-units [CI's]	7.87 [7.05-8.68]	7.42 [6.73-8.12]	10.91 [9.11-12.72]	11.23 [9.17-13.29]
		Method L-Moments		
Location	1.0149	1.0896	2.15	2.3354
CIs	[0.9670-1.0622]	[1.0426-1.1377]	[1.9981-2.2944]	[2.1921-2.4829]
Scale	0.9110	0.9377	1.2744	1.2448
CIs	[0.8654-0.9553]	[0.8955-0.9838]	[1.1459-1.4040]	[1.1182-1.3699]
Shape	0.2520	0.2196	0.2506	0.2398
CIs	[0.2035-0.3057]	[0.1790-0.2660]	[0.1555-0.3603]	[0.1356-0.3428]
K-S test (two-sample)	0.024794 (0.216)	0.021813 (0.3556)	0.03 (0.8472)	0.046184 (0.3374)
		Return Level for Period (*Units*)		
10-units [CI's]	3.77 [3.60-3.95]	3.82 [3.63-4.00]	6.00 [5.47-6.63]	6.05 [5.57-6.64]
20-units [CI's]	5.04 [4.73-5.35]	5.02 [4.68-5.34]	7.77 [6.87-8.90]	7.73 [6.91-8.73]
50-units [CI's]	7.06 [6.49-7.71]	6.88 [6.28-7.52]	10.59 [8.96-12.87]	10.37 [8.91-12.31]
	Sources: r-packages (“evd”, “evir”, “extRemes”, “fExtremes”, “mev”, “lmom”, “extremeStat”, “ismev”, “eva”)			

Fitting the GEV Distribution in Weekly Block Minima

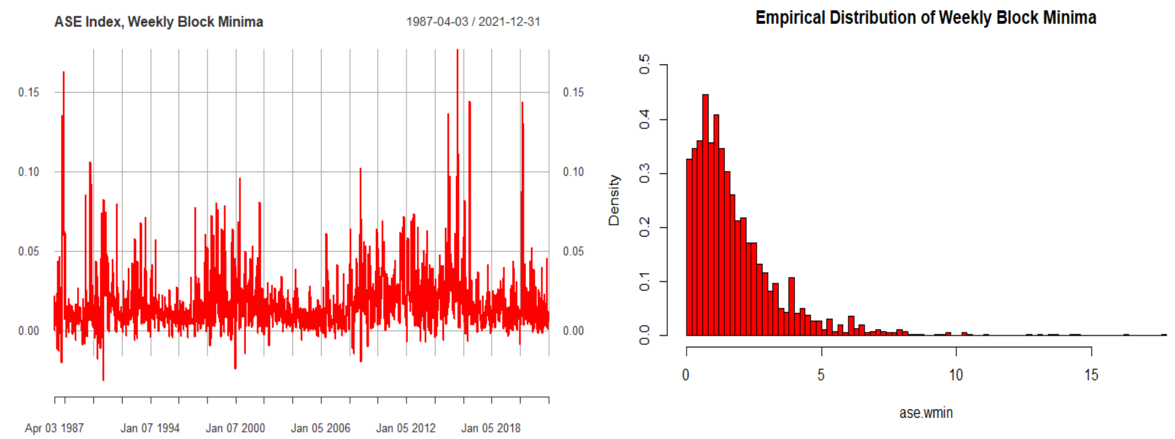

Figure 8. Left Panel: Weekly block minima of the ASE Index. Right Panel: Distribution of weekly block minima of the ASE index

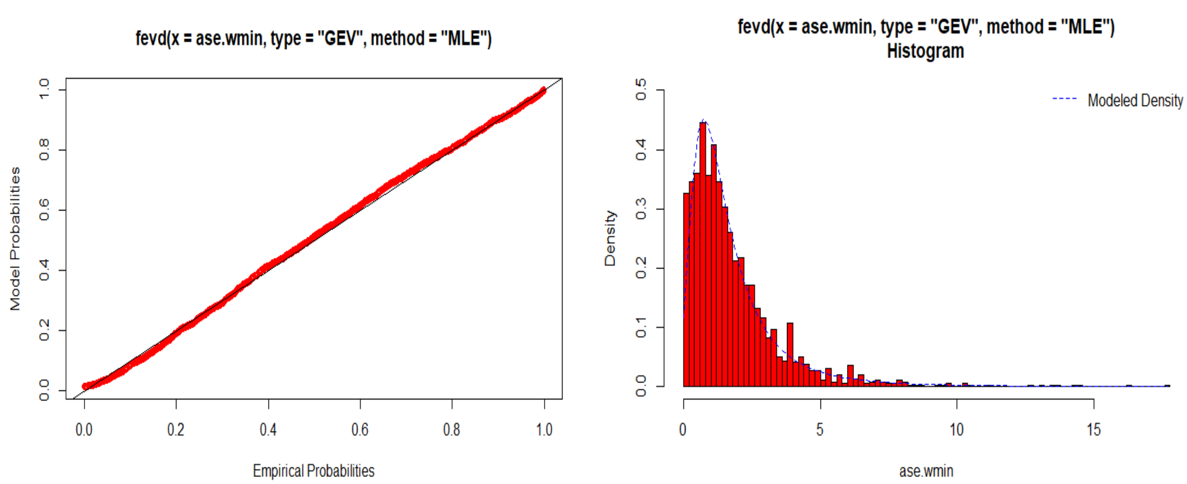

Figure 9. Left Panel: Fit model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

The linearity in the probability plot as Figure 9 and quantile plot as below (Figure 10) indicates that the GEV distribution could be considered an appropriate fit to the underlying ASE data for shorter return periods.

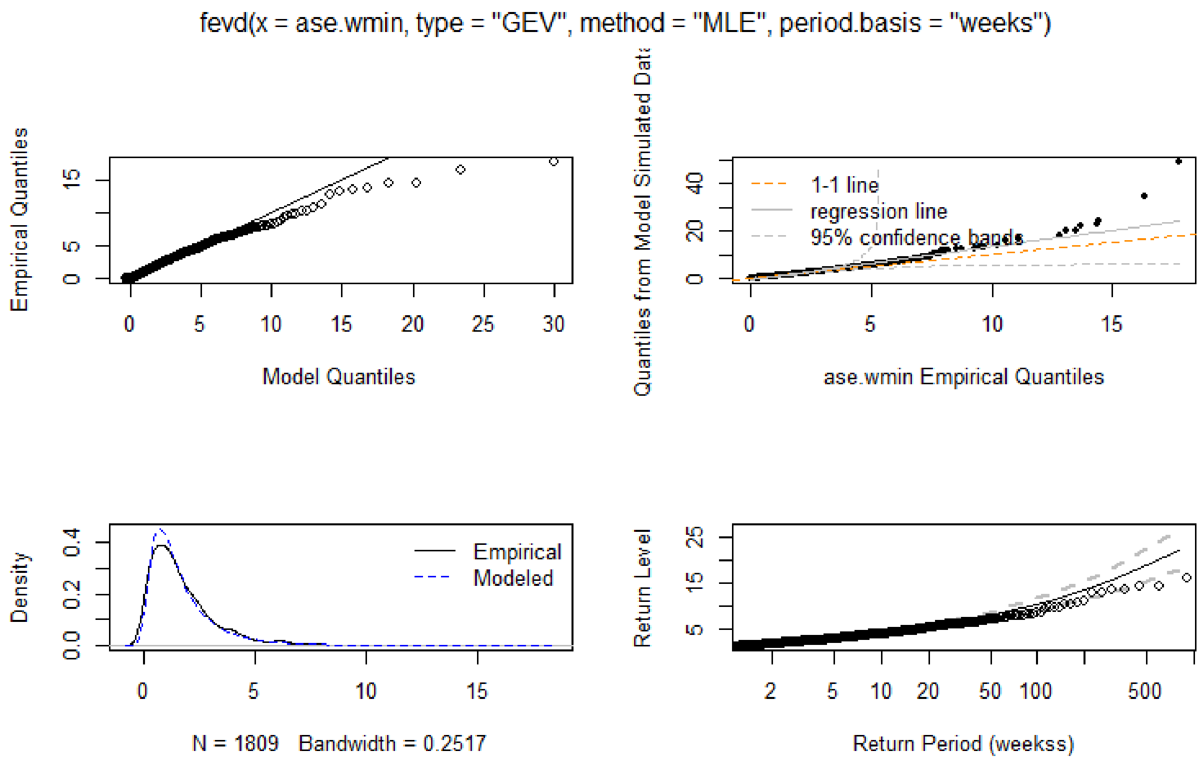

Figure 10. Diagnostic Plots from the GEV df fit to ASE weekly minima; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right).

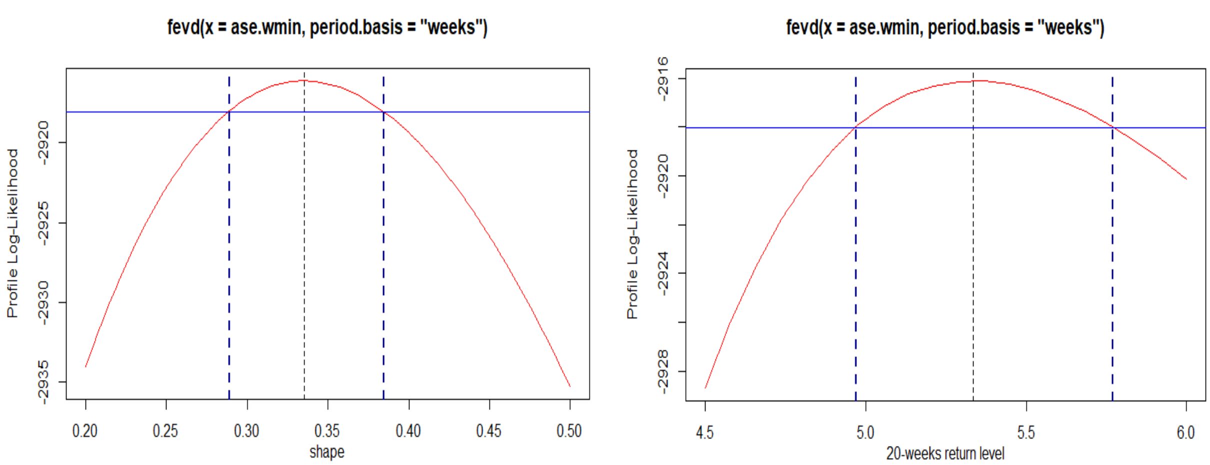

Figure 11. Left Panel: Profile Likelihood for shape parameter. Right Panel: Profile Likelihood for 20-week return level.

The above figures show standard diagnostic plots.

Probability plot's (Figure 9, Left Panel) ordinate is the corresponding theoretical frequency of the sample series value on the theoretical frequency distribution curve and its abscissa is the empirical frequency of the sample series. The ordinate of the *quantile plot* is the actual value of the sample series, and its abscissa is the corresponding quantile value of the empirical frequency of the sample series on the theoretical frequency distribution curve. All the dots that represent sample data in the probability plot and the quantile plot are virtually located on a straight line with a slope of 1.

Return level plot's ordinate is the return level of even scale, and its abscissa is the return period in period with logarithmic scale adopted. In the plot, the upper and the lower curves represent the confidence upper limit and confidence lower limit of the return level respectively. All the dots representing sample data are within/or not the range of the confidence interval, concentrated near the middle convex/concave curve and approaching a certain finite value. Therefore, the return level curve also indicates satisfaction of the GEV model.

The lower right plot in the *diagnostic plot* (Figure 10) is the *density plot* which consists of *histogram* and *distribution density curves*. Its abscissa is series returns and its ordinate is return frequency. The estimation of fitting weekly minimum data to a GEV density curve also fits in with the sample histogram.

In each case, the assumptions for fitting the GEV to the data appear reasonable. For the shape parameter and longer return levels, it is often better to obtain confidence intervals from the profile-likelihood method in order to allow for skewed intervals that may better match the sampling df of these parameters.

By using the profile-likelihood method we get an improved 95% confidence interval for the shape parameter of about (0.2889, 0.3845), but similar to the normal approximation limits (Table 6). We also get that approximate 95% confidence interval for the 20-week return level based on the profile-likelihood is about (4.97, 5.77), which is also similar in this case to the normal approximation limits, though shifted a little bit toward the higher end.

When using the profile-likelihood method, it is necessary to specify a range over which to seek out the confidence intervals. It is generally important to inspect the plot to be sure that the resulting confidence intervals are not merely the endpoints used.

Figure 11 shows the profile-likelihood for the shape parameter and the 20-week return level. Because the profile-likelihood clearly crosses both blue vertical dashed lines (where they cross the blue horizontal line), the resulting intervals are believable.

Applying the L-moment method we can see a more reliable fitting. QQ-plot and the return level plot approximate better the empirical data as shown in the below figure.

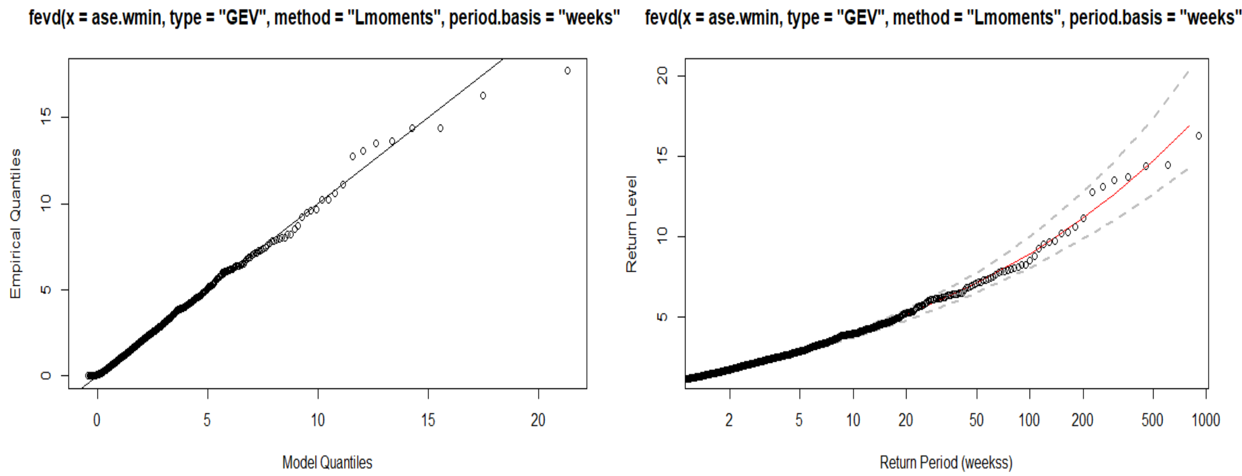

Figure 12. Left Panel: QQ-Plot of ASE weekly minima; Right Panel: Return Level Plot; L-moment method.

Fitting the GEV Distribution in Weekly Block Maxima

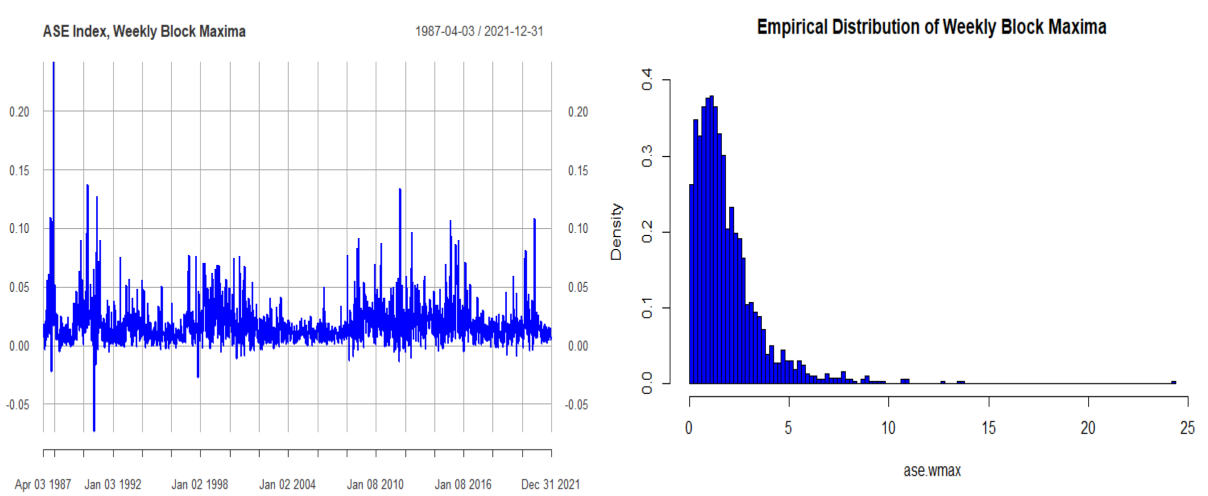

Figure 13. Left Panel: Weekly block maxima of the ASE Index. Right Panel: Distribution of weekly block maxima of the ASE index

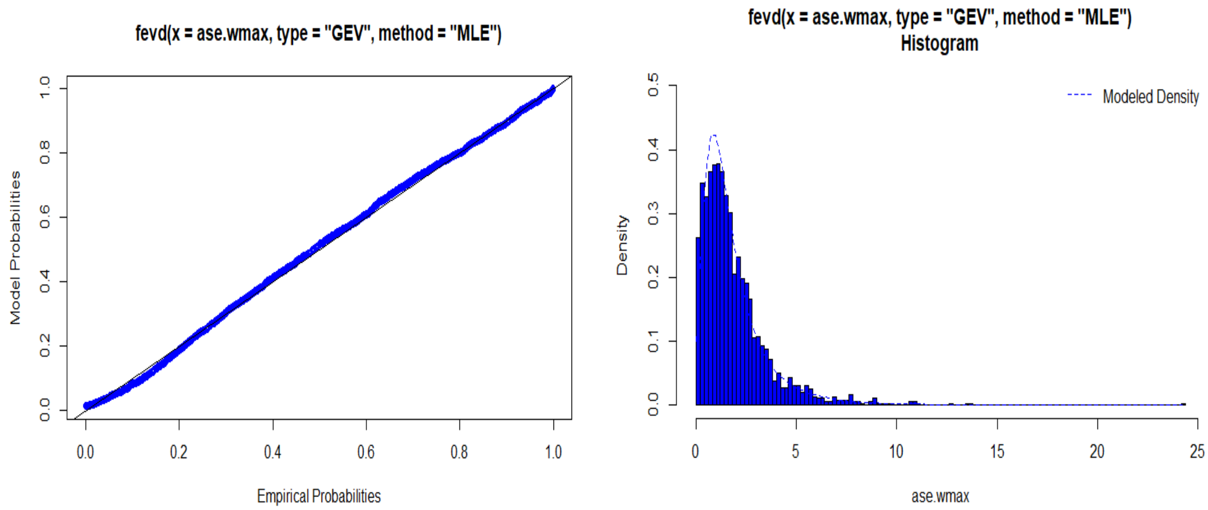

Figure 14. Left Panel: Model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

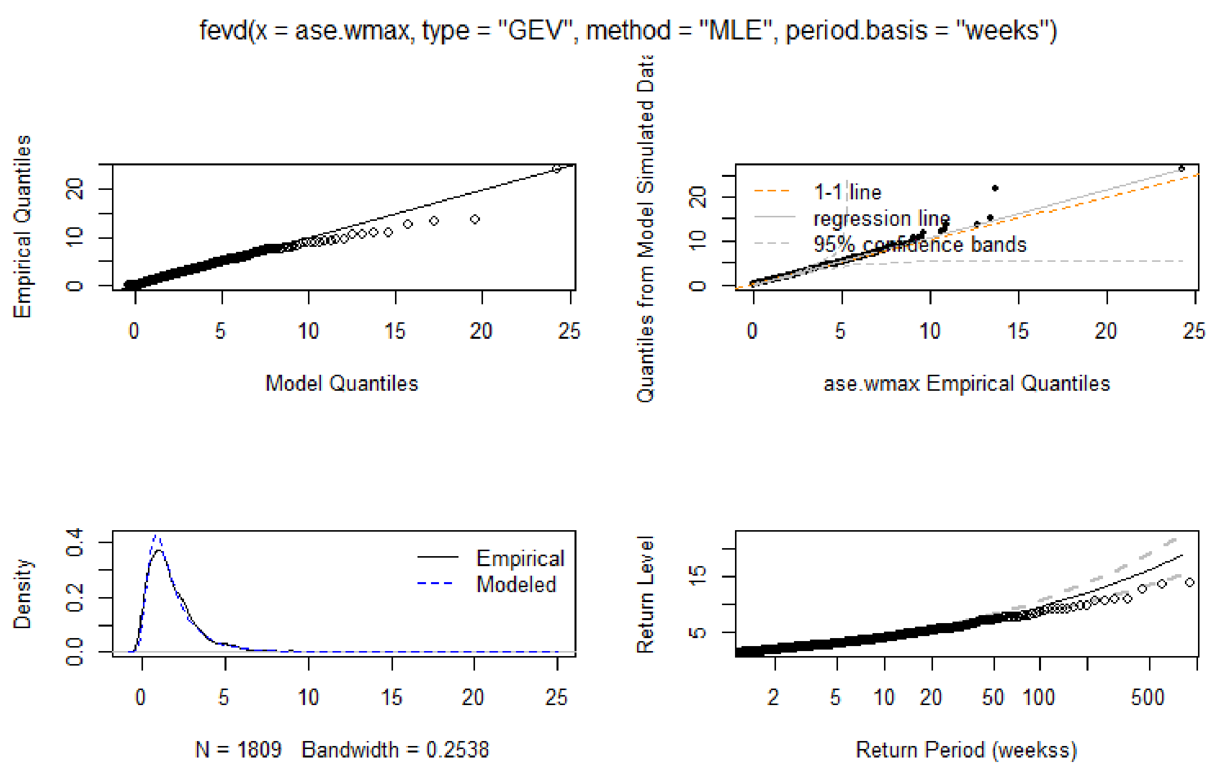

Figure 15. Diagnostic Plots from the GEV df fit to ASE weekly maxima; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right).

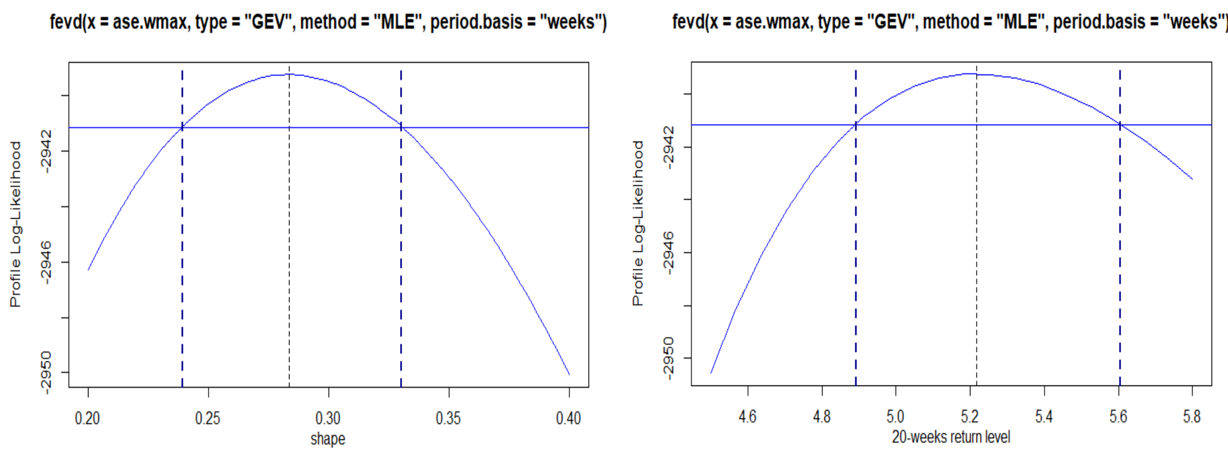

Figure 16. Left Panel: Profile Likelihood for shape parameter. Right Panel: Profile Likelihood for 20-week return level.

By using the profile-likelihood method we get an improved 95% confidence interval for the shape parameter of about (0.2393, 0.3301), close to the normal approximation limits (Table 6). We also get that approximate 95% confidence interval for the 20-week return level based on the profile-likelihood is about (4.89, 5.61), which is also similar in this case to the normal approximation limits. The profile-likelihood clearly crosses both blue vertical dashed lines (where they cross the blue horizontal line), the resulting intervals are also believable.

Similarly, L-moments method applies better to these frequency data period, as shown in the below figure.

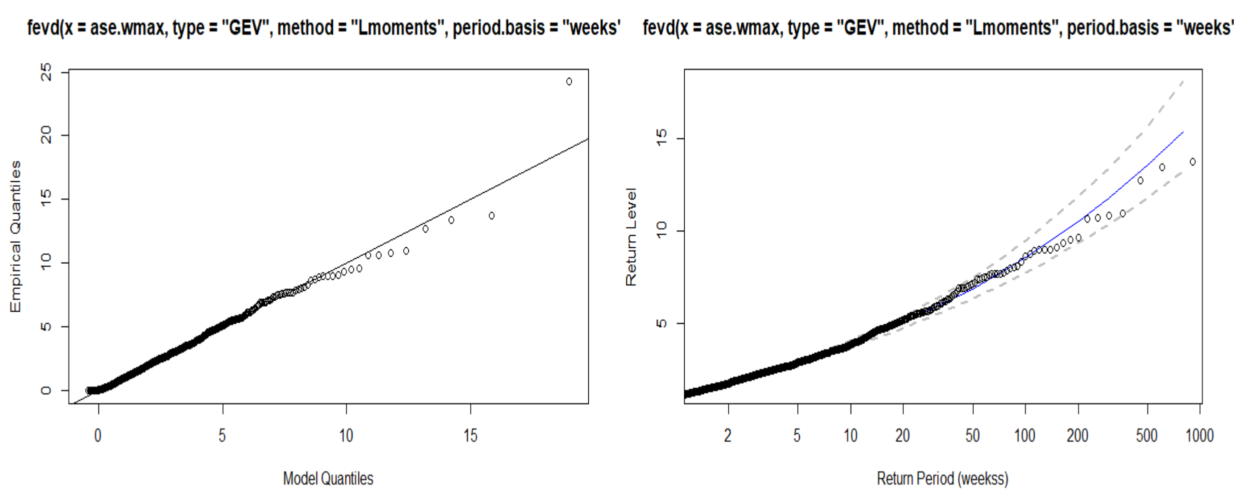

Figure 17. Left Panel: QQ-Plot of ASE weekly maxima; Right Panel: Return Level Plot; L-moment method.

Fitting the GEV Distribution in Monthly Block Minima

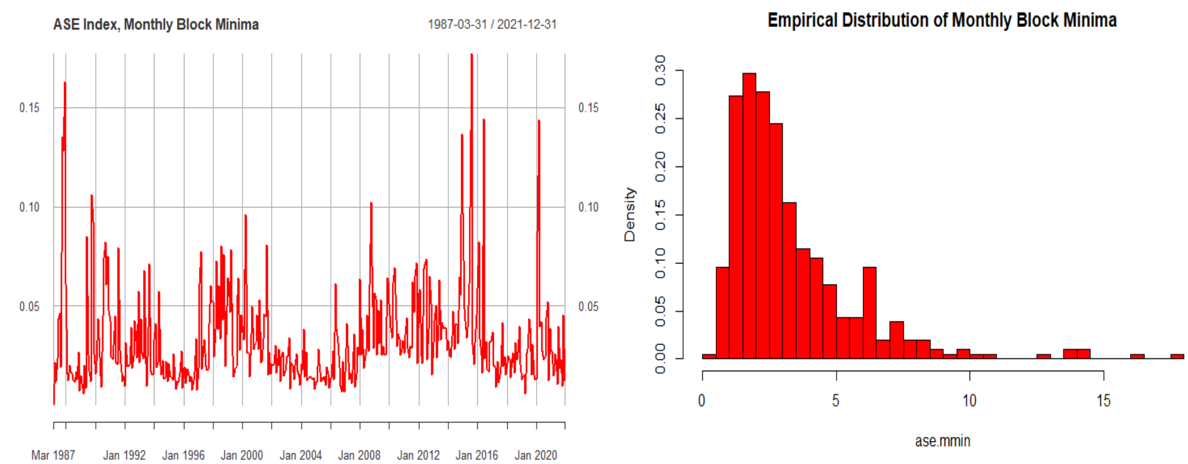

Figure 18. Left Panel: Monthly block minima of the ASE Index. Right Panel: Distribution of monthly block minima of the ASE index

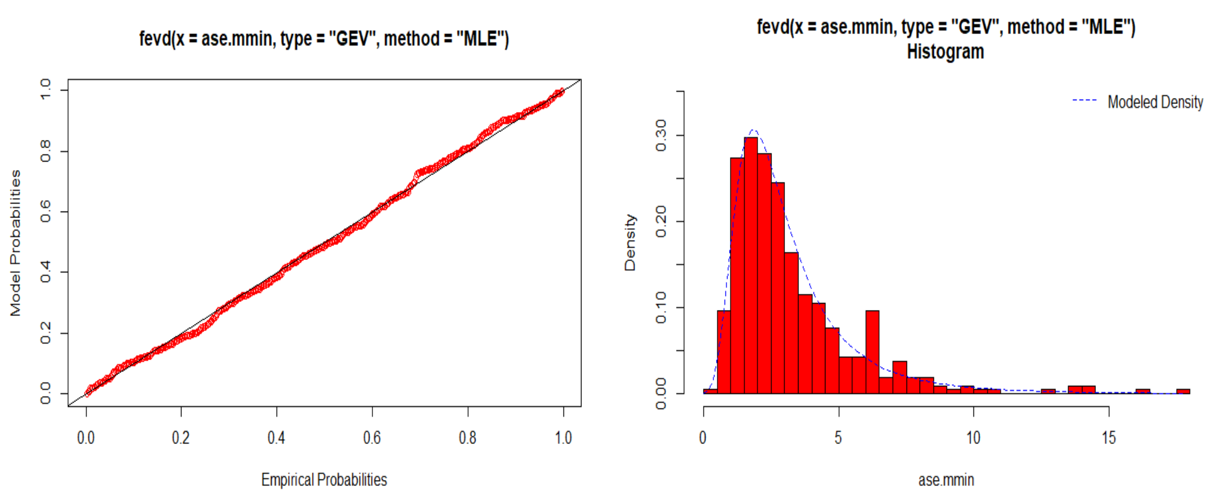

Figure 19. Left Panel: ASE Monthly Minima Model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

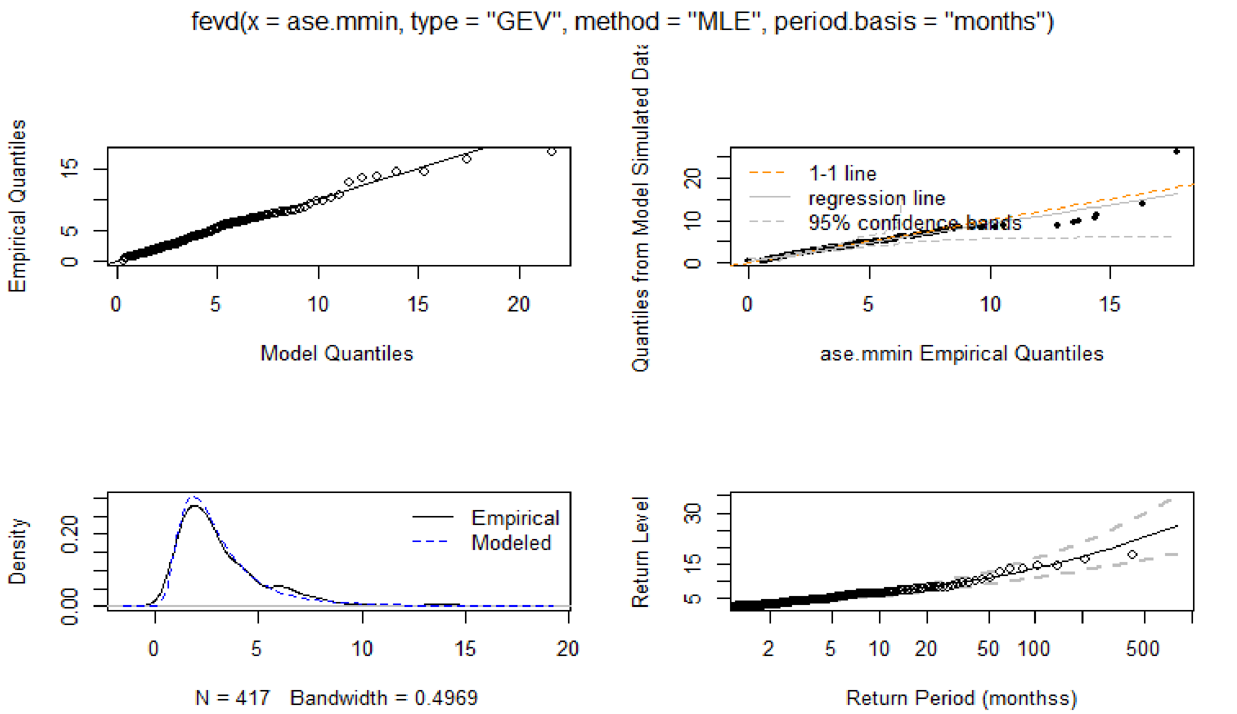

Figure 20. Diagnostic Plots from the GEV df fit to ASE monthly minima; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right)

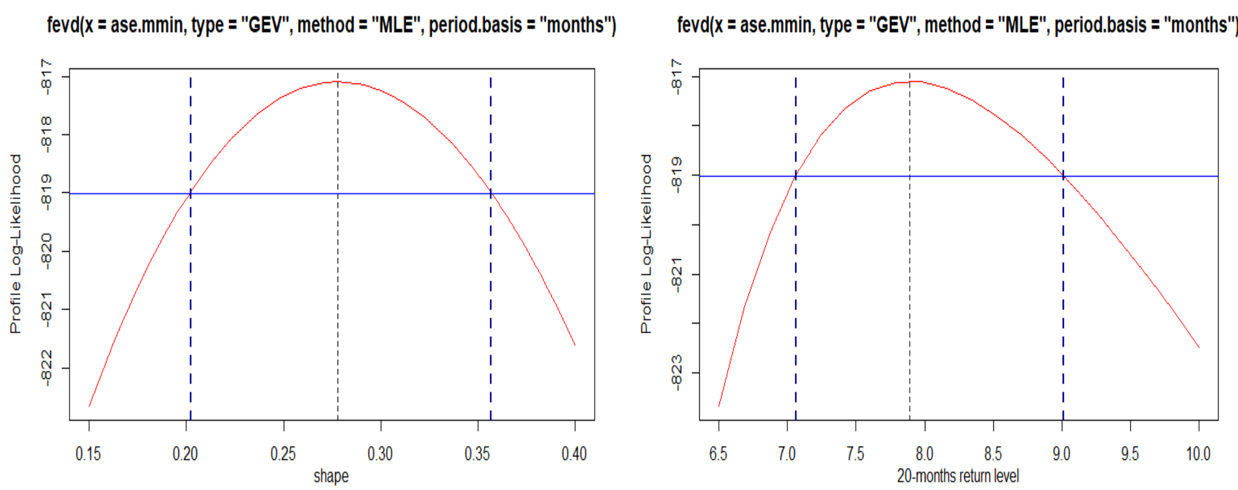

Figure 21. Left Panel: Profile Likelihood for shape parameter. Right Panel: Profile Likelihood for 20-month return level

In this case, the assumptions for fitting the GEV to the data appear reliable. Both MLE and L-moments methods approximate the sample data similarly.

By using the profile-likelihood method we get an improved 95% confidence interval for the shape parameter of about (0.2024, 0.3568), very similar to the normal approximation limits (Table 6). We also get that approximate 95% confidence interval for the 20-week return level based on the profile-likelihood is about (7.1, 9.01), shifted a little bit toward the higher end. The profile-likelihood also crosses both blue vertical dashed lines (where they cross the blue horizontal line), the resulting intervals are believable.

Fitting the GEV Distribution in Monthly Block Maxima

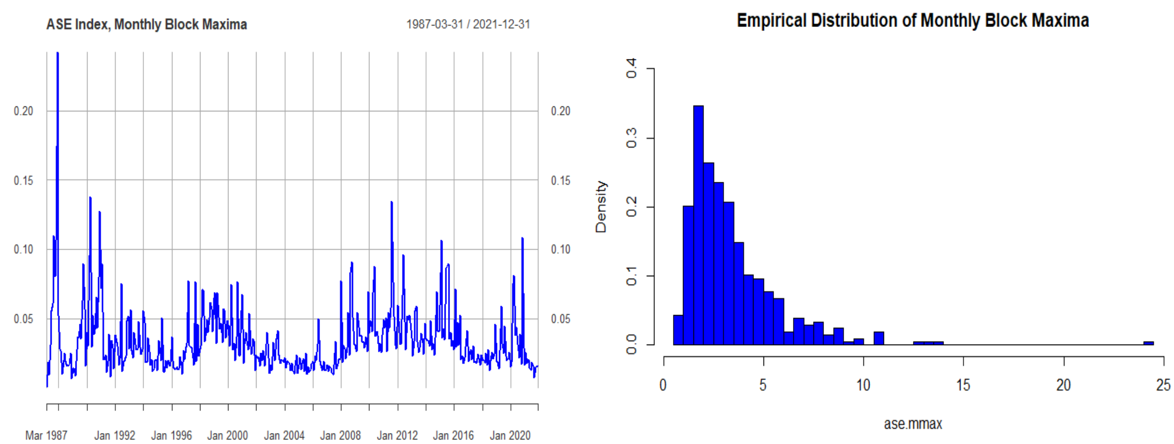

Figure 22. Left Panel: Monthly block maxima of the ASE Index. Right Panel: Distribution of monthly block maxima of the ASE index

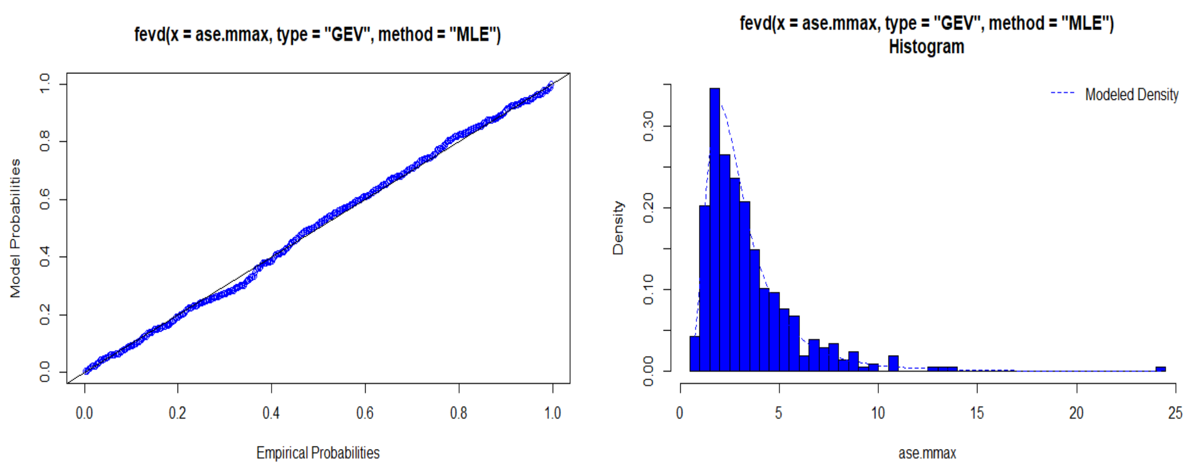

Figure 23. Left Panel: ASE Monthly Maxima Model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

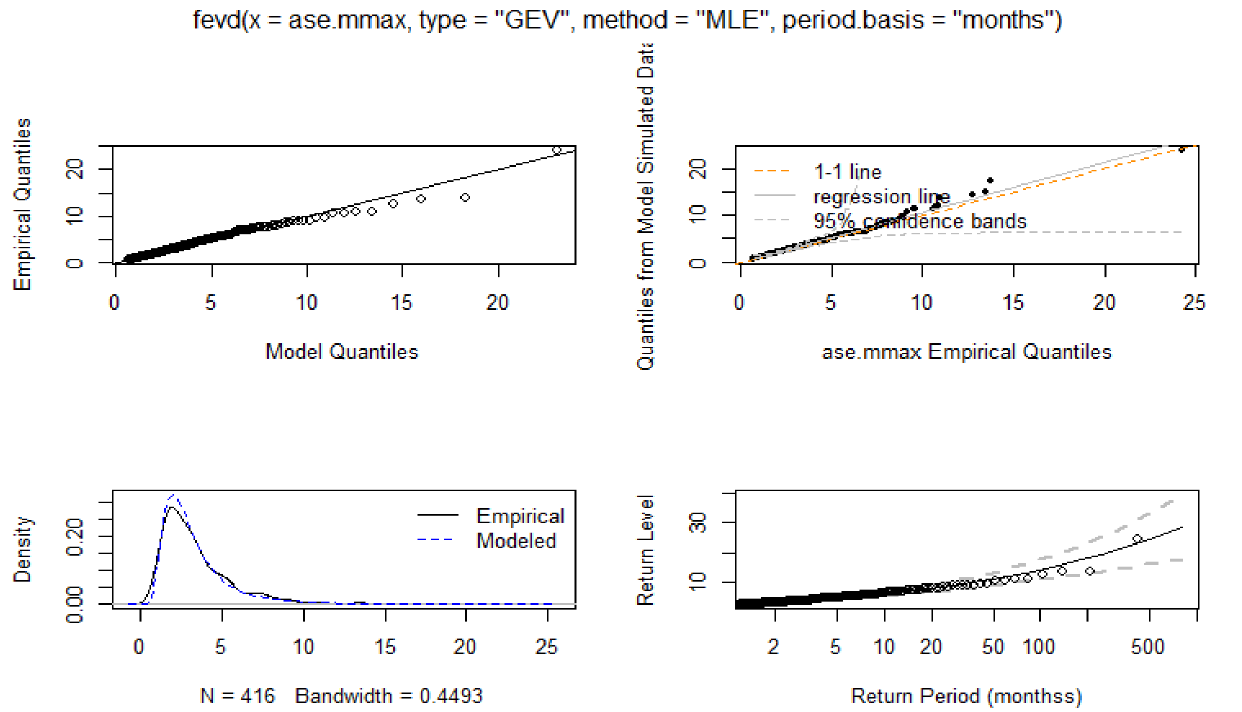

Figure 24. Diagnostic Plots from the GEV df fit to ASE monthly maxima; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right)

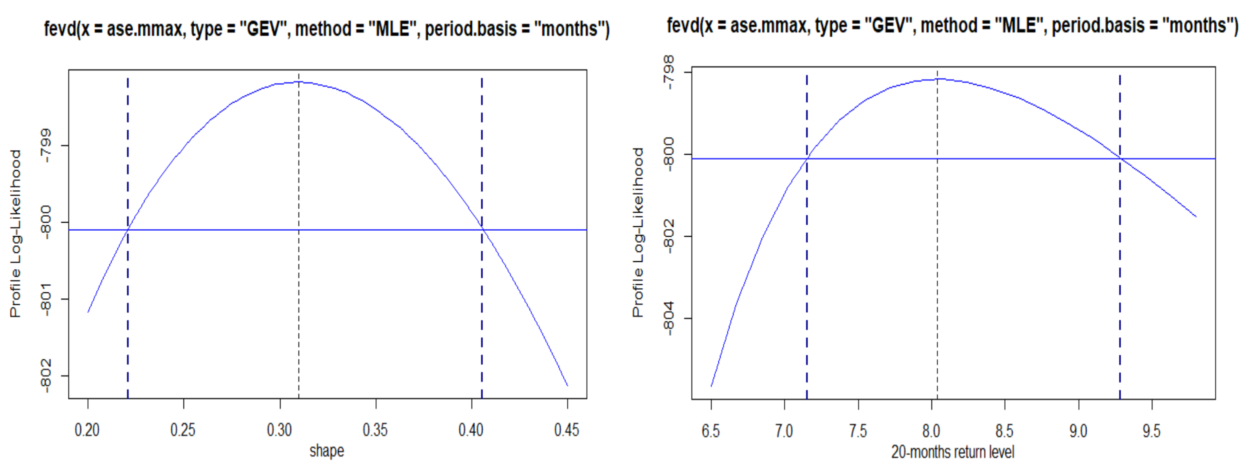

Figure 25. Left Panel: Profile Likelihood for shape parameter. Right Panel: Profile Likelihood for 20-month return level

In this case, the assumptions for fitting the GEV to the data also appear satisfactory.

By using the profile-likelihood method we get an improved 95% confidence interval for the shape parameter of about (0.221, 0.4055), very similar to the normal approximation limits (Table 6). We also get that approximate 95% confidence interval for the 20-week return level based on the profile-likelihood is about (7.15, 9.29), shifted a little bit toward the higher end. The profile-likelihood also crosses both blue vertical dashed lines (where they cross the blue horizontal line), the resulting intervals are believable.

Reviewing the above results, diagnostic plots for the GEV model of the sample series of ASE weekly and monthly minimum/maximum returns, work to reflect the fitting degree of the selected theoretical distribution and the actual sample series.

All above diagnostic plots imply that all series show a good fit to the GEV model. The Probability Plots (PP) and the Quantile Plots (QQ) seem to confirm that the data has been generated by the Fréchet distribution, as the points are near the straight line. Zero in any case falls in the 95% confidence interval, another case of rejecting the null hypothesis of a Gumbel distribution. Kolmogorov-Smirnov statistic in all cases, shows a good fit.

Inspection of the above diagnostic plots show large apparent "outliers" toward the upper end. But such behavior is fairly common for the most extreme observation (in part because of considerable uncertainty) and does not provide strong evidence against the validity of the assumptions for using the GEV df.

Fitting a GP Distribution in ASE Index Log Returns

Parametric Estimation

To start we compute the LR test (Likelihood Ratio test) where in the null we assume the daily returns have been generated by an exponential distribution:

$$LR\ Test\ = \begin{cases} H_0 = Exponential\ Distribution \\ H_1 \neq Exponential\ Distribution' \end{cases}$$

For the LR test (significance level of 5%), we cannot assume the series have been generated by an exponential distribution when we reject the null hypothesis.

The GP distribution can be defined constructively in terms of exceedances. Starting with a probability distribution whose right tail drops off to zero, such as the normal, we can sample random values independently from that distribution. If we fix a threshold value, throw out all the values that are below the threshold, and subtract the threshold off the values that are not thrown out, the result is known as exceedances. The distribution of the exceedances is approximately a GP.

Before fitting the GP df to the data, it is first necessary to choose a threshold. It is important to choose a sufficiently high threshold in order that the theoretical justification applies thereby reducing bias.

However, the higher the threshold, the less available data remains. Thus, it is important to choose the threshold low enough in order to reduce the variance of the estimates.

We can take the largest 10% from the empirical distribution, and then subtract off the 90% quantile to get exceedances. This produces a threshold of 2.8%. We will use this threshold to fit a GP distribution to the left and right tail of our observations, as well to the full data sample.

We can also approximate threshold by visualizing our data to perform diagnostics and draw conclusion based on the results. The threshold level set is more dependent on the data itself and allows for more parameters when deciding relative to using a predetermined rule. However, these methods are far more time-consuming than using a rule-of-thumb.

To inspect the heaviness or not of our full sample size series we examine the *Exponential QQ Plot*. The near concavity in the exponential QQ-plot is a strong signal of the presence of heavy tails.

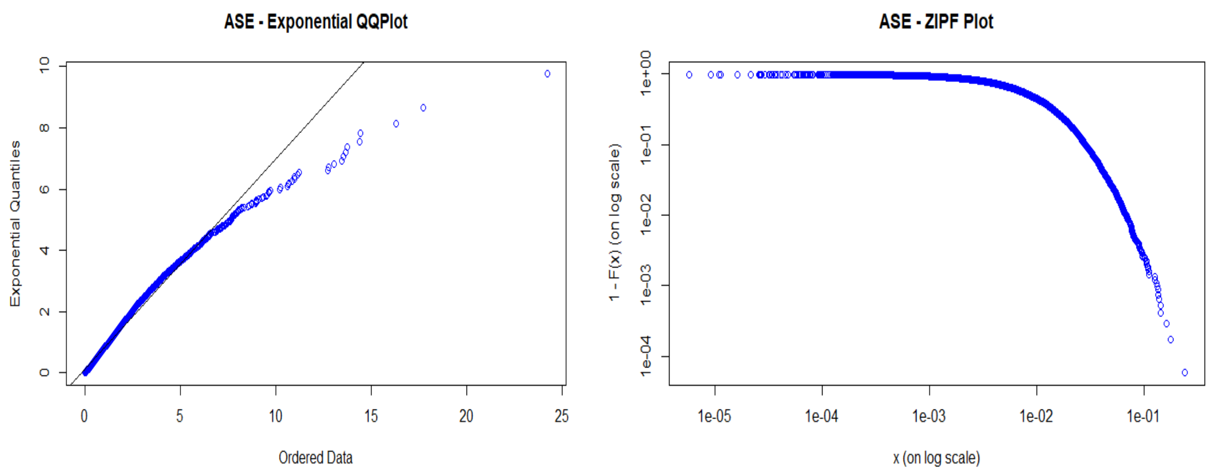

Figure 26. Left Panel: ASE Exponential QQ-Plot. Right Panel: Log-Log Plot (Zipf Plot)

Zipf Plot (or Log-Log Plot) is useful to verify the presence of a power-law decay in the tail of our data. On the x-axis we put the log of our ordered observations and on the y-axis, we put the log of our survival function. A clear negative (decreasing) linear slope is present in the tails. This is the first signal of the fat tailed nature of the data. But a Zipf plot verifies a necessary yet not sufficient condition.

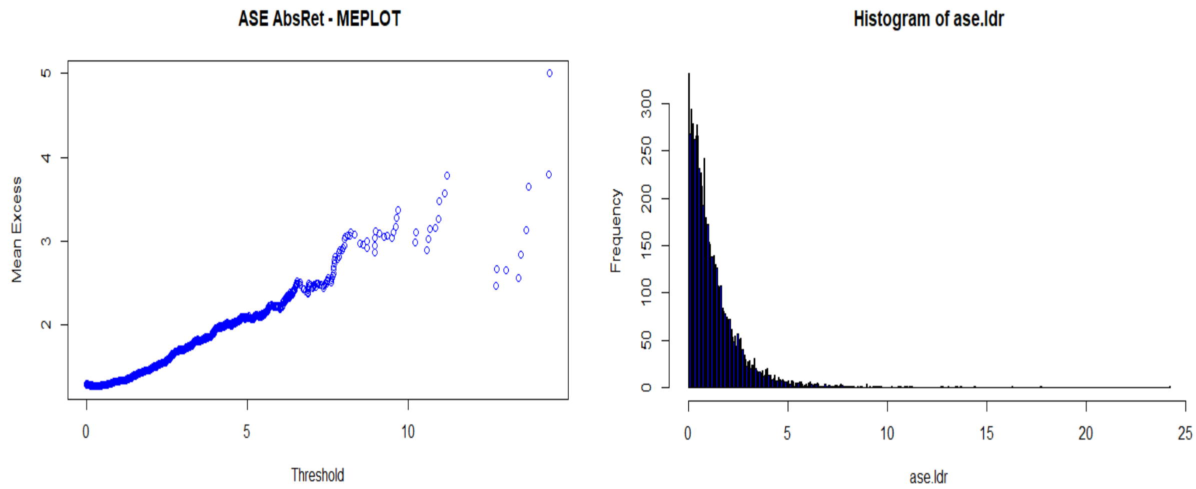

Figure 27. ASE Index Sample Mean Excess Plot and Histogram of absolute returns dataset.

Based on the *Mean Excess Plot* as the Figure 27, right panel, we can infer a fat-tailed behavior for our data since the plot exhibits an upward trend. If the empirical plot appears to follow a reasonably straight line with a positive gradient, then this is an indication that the data follow a GPD with a positive shape parameter ξ in the tail area above u. The plot of Mean Excess above the threshold u = 2.8% suggests that the GPD might provide a reasonable fit to the ASE Index.

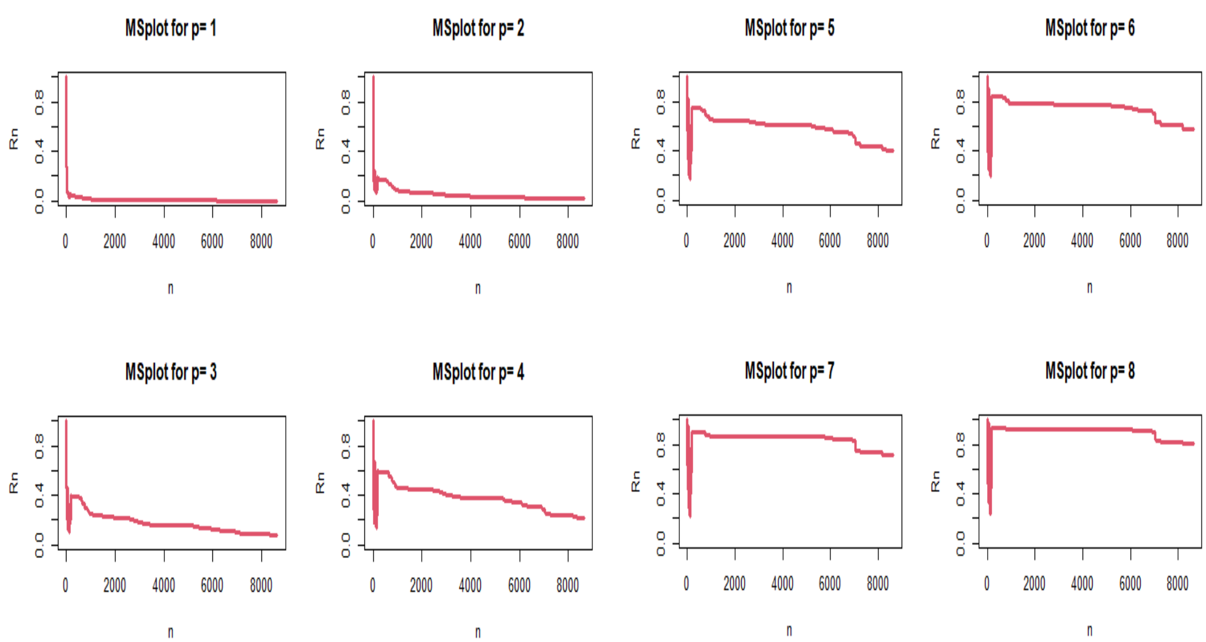

Figure 28. MS Plot: A simple code to check for the first 8 moments.

The *Maximum to Sum Plot (MS Plot)*, Figure 28, *or the ratio of maximum and sum*, R_n (r), for different r, is a simple tool for detecting heavy tails of a distribution and for giving a rough estimate of the order of its finite moments.

In principle, we have that the moment of order *r* exists only if alpha (α) is larger than *r*. So, if the moment of order 2 is finite, it means that alpha is greater than 2. If the moment of order 3 is finite, it means that alpha is larger than 3. But what it means that the theoretical moments are not finite. If we look at our sample data, we always be able to compute the moments of our data. The point is that the theoretical moments might not be finite. And if the theoretical variance is not finite the sample variance, we are computing is essentially useless, since we cannot make inference by using it. On the other hand, if the second moment is finite, but if the fourth moment is not finite, it is then difficult to build confidence intervals for the variance.

In our case as the above Figure 28, we may infer an alpha between 4 and 5 since we can see some kind of convergence to 0, up to 5^{th} moment.

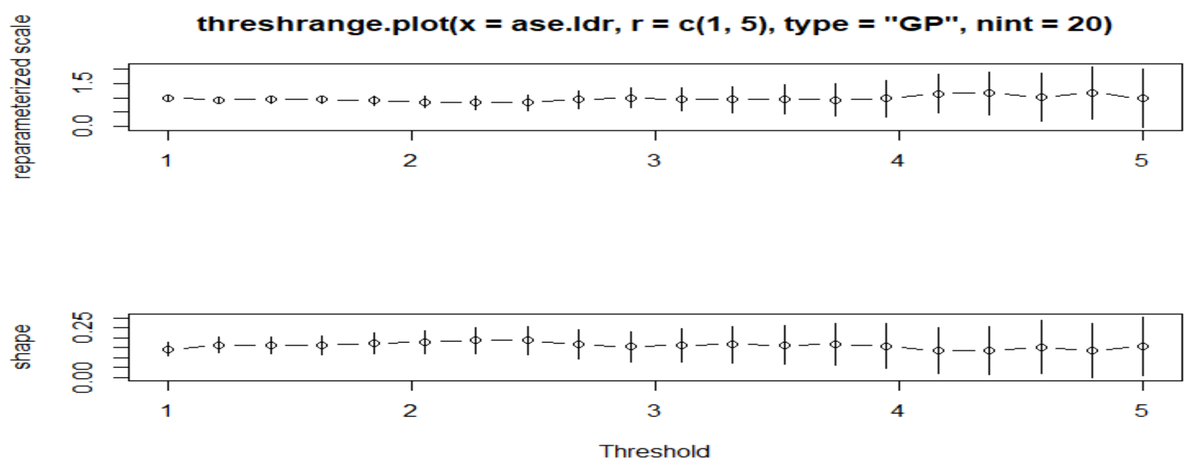

Figure 29. Threshold Stability Plot for ASE Index

The *threshold stability plot* looks at the estimates of the scale and shape parameters to find a suitable threshold, as repeatedly fits the GP df to the data for a sequence of threshold choices along with some variability information. Over some range of *u* and the assumption of being one parameter constant we estimate the other parameter. Two parameter stability plots then can be made by plotting these estimators against the interval *u* with the 95% confidence interval for the estimators. Looking at the plots one can find a suitable threshold at the lowest value where the plots are approximately constant.

Figure 29 shows that the modified scale parameters and shape parameters are, if considering the randomness of sampling, stable near 2.8 (1.7, 2.4 and 3.8) and a certain range of values to its right within the allowable sampling error, which can show that threshold is appropriate, and the corresponding excess complies with the GP distribution.

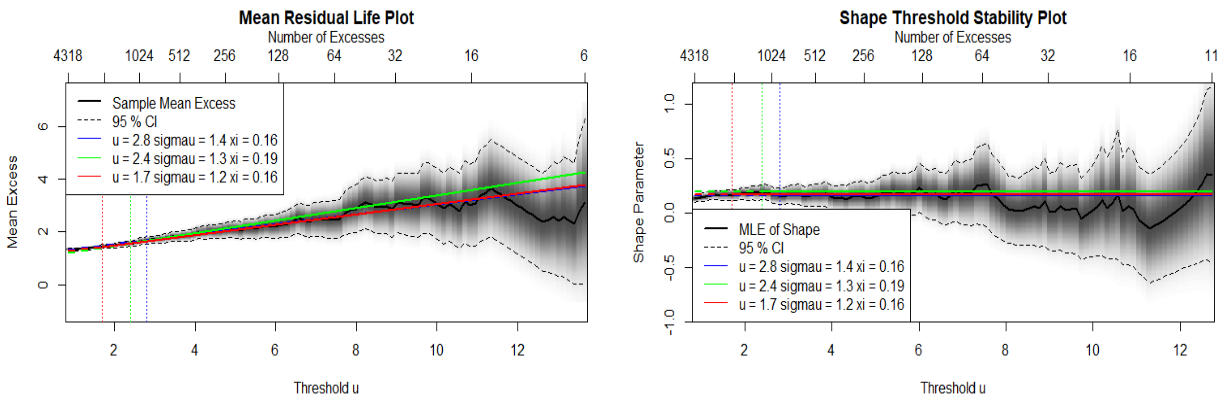

Figure 30. Left Panel: Mean Residual Life Plot. Right Panel: Shape Threshold Stability Plot

The left panel of the Figure 30, Left Panel, represents the MRL plot for the ASE Index returns. Solid jagged line is the empirical MRL with approximate 95% confidence interval (dashed lines). By examining the mean residual life plot, the threshold *u* is to be chosen so that there is a linear line up to the chosen threshold before it decays sharply. The MRL suggested by maximum likelihood parameter estimates for the threshold *u* =1.7, 2.4 and 2.8 are upper, middle, and lower colorful solid straight line. Vertical dashed lines indicate the thresholds selected. The sampling density for the MRL is shown by a greyscale image, where lighter greys indicate low density. The threshold level choice *u* = 2.8 gives about 890 exceedances and is sustained by the empirical MRL being close to linear above this level.

An alternative method to the mean residual life plot is to model multiple thresholds where the estimated parameters are approximately constant and then pick the smallest threshold *u* in which linearity is constant to higher values and within the confidence levels. The *shape threshold stability plot* (Figure 30, Right Panel), exhibits conservative confidence intervals represented by relatively small sampling errors. This strongly indicates that one of the values could be chosen for proceeding with the extreme value analysis as they indicate strong stability on the shape parameter estimated. The selection of a higher threshold would affect the stability criteria as the sample variation becomes too large to make useful inferences. The best choice for threshold value seems to be relative to the objectives of the research. Therefore, prior to a fixed choice, a sensitive analysis may produce insights regarding a consistent threshold level.

Once the threshold is set, parameters can be estimated through the Maximum Likelihood Method (MLE).

Table 7. Fitting GPD to ASE Index – Parametric Estimation

	Method: MLE		
	All Dataset (u = 2.8)	*Left Tail (u = 2.8)*	*Right Tail (u = 2.8)*
Negative Log-Likelihood Value	1,347.911	694.2181	653.4398
	Estimated parameters		
scale (ses)	1.4083 (0.07175)	1.4487 (0.1035)	1.3685 (0.0995)
scale CIs	[1.2677 – 1.5490]	[1.2459-1.6515]	[1.1735-1.5635]
shape (ses)	0.1603 (0.0388)	0.1584 (0.0544)	0.1613 (0.0555)
shape CIs	[0.0841 – 0.2364]	[0.0519-0.2650]	[0.0524-0.2702]
AIC	2,699.821	1,392.436	1,310.88
BIC	2,709.419	1,400.672	1,319.067
	Return Level for Period (units)		
10-units [CIs]	15.46 [13.02-17.90]	17.24 [12.97-21.51]	16.32 [12.29-20.35]
20-units [CIs]	17.98 [14.57-21.39]	19.98 [14.11-25.85]	18.92 [13.36-24.48]
50-units [CIs]	21.77 [16.66-26.88]	24.09 [15.46-32.72]	22.84 [14.62-31.06]
Notes:	*Sources: r-packages (“evd”, “evir”, “extRemes”, “fExtremes”, “mev”, “lmom”, “extremeStat”, “ismev”, “eva”)* *The sign of the left tail has been changed so that positive values correspond to losses*		

Table 7 presents all point estimates along with their confidence intervals and their standard errors. We run GP df concerning complete dataset as the left and right tail also. The results are very close indicating some sort of stability of our model. The shape parameter is well above zero, and confidence intervals for all three cases do not include zero, a sign also for rejecting the exponentiality of our distribution.

There is an approximate relationship between the GEV and GP distribution functions where the GP df is approximately the tail df for the GEV df. In particular, the scale parameter of the GP is a function of the threshold and is equivalent to *scale + shape*(threshold - location*) where scale, shape and location are parameters from the “equivalent” GEV df. Taking the monthly minimum parameter estimations as extracted by GEV, we end up to the same scale estimate resulted by the GP fitting.

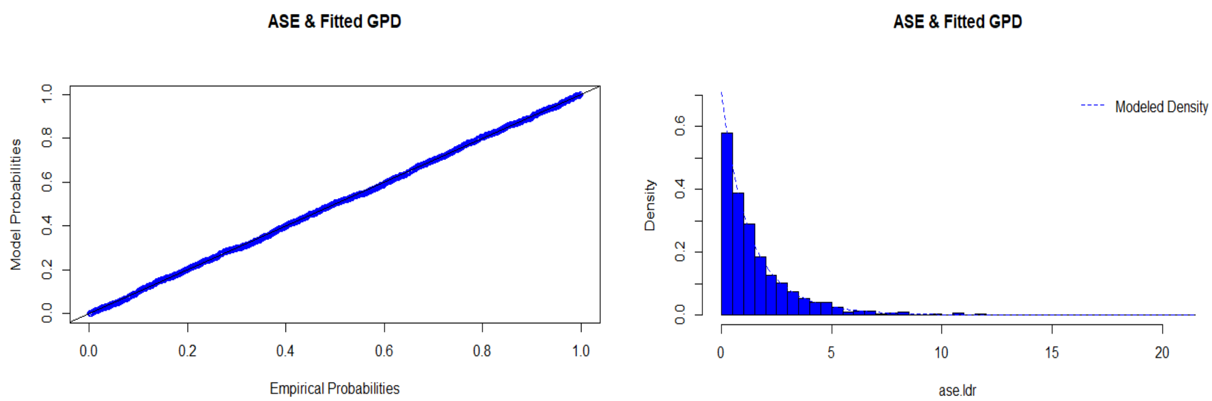

Figure 31. Left Panel: ASE GDP Fitted Model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

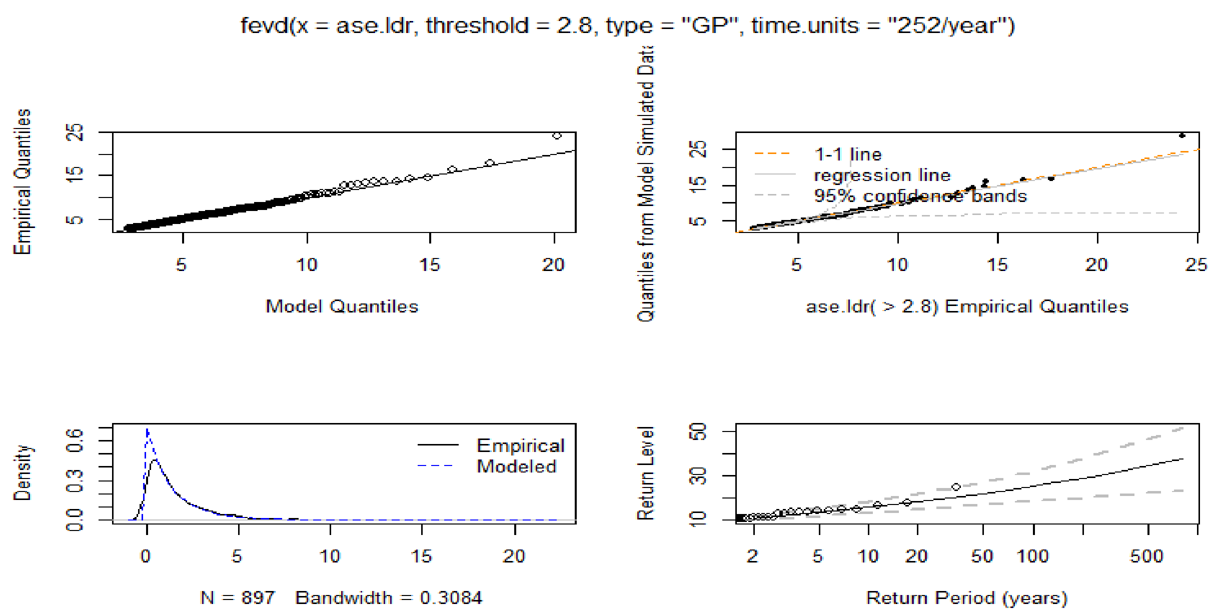

Figure 32. Diagnostic Plots from the GPD df fit to ASE sample data; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right)

Figure 32 includes four diagnostic plots for the GP model of the sample series of all ASE daily returns and they work to reflect the fitting degree of the selected theoretical distribution and the actual sample series. The four diagnostic plots imply that the series shows a good fit to the GP model. In each case, the assumptions for fitting the GPD to the data appear reasonable.

The exceedances, which are the observations above a threshold value of 2.8, is fitted to a GP distribution and depictured as the modeled density function in Figure 31, Right Panel. The relative linearity in the probability plot (Figure 31, Left Panel) is an indication that the exceedances can be fitted to a GP distribution. The linearity in the quantile plot is also substantial there. Hence the POT model could seem ambiguous for higher return levels.

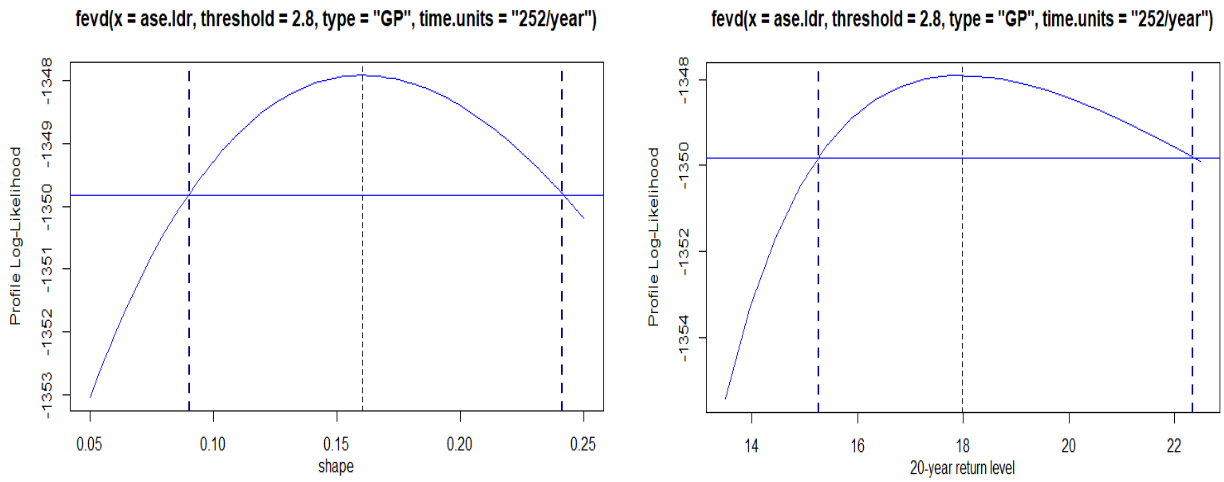

Figure 33. Left Panel: Profile Likelihood for shape parameter. Right Panel: Profile Likelihood for 20-month return level

The shape parameter ξ of the GP model and the logarithmic likelihood function curve for a return period of 20 years are shown in above Figure 33. The ratio of upper limit of confidence interval to 20 increases as the return period increases indicating that the asymmetry of the profile likelihood function curve increases with the increase of the return period. Longer return period leads to higher return level and means less information available from the sample data and hence higher uncertainty of extreme estimation.

Non-parametric estimation

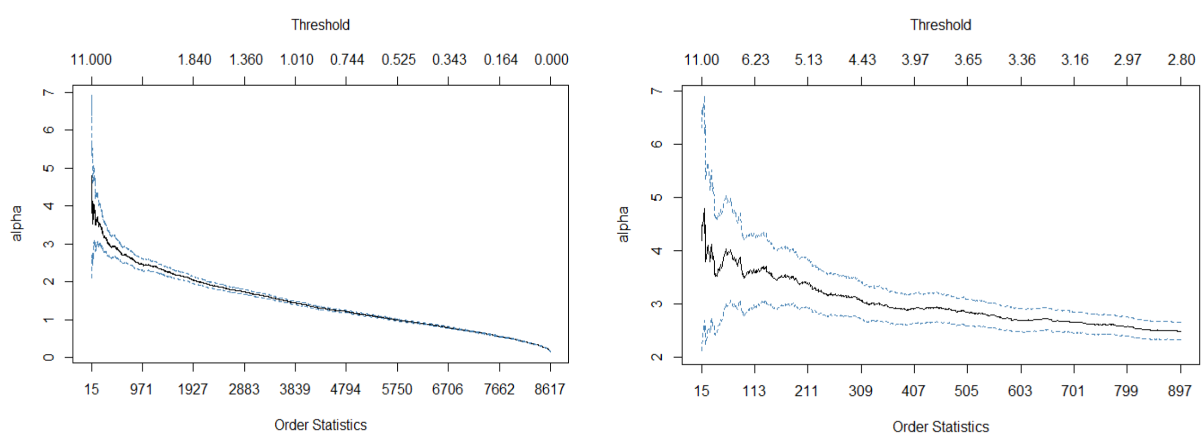

Figure 34. Estimates of the Hill ξ for the Daily ASE Returns. Left Panel: Full data series; Right Panel: Data above threshold u = 2.8

According to McNeil et al. (2005), the GPD method is not the only way to estimate the tails of a distribution as has been performed above. The other methodology for the selection of the threshold is based on the Hill estimator, estimating, in a non-parametric way, the *Hill tail index* $\alpha = 1/\xi$ and the quantile for the stock market returns. This estimator is often a good estimator of α, or its reciprocal ξ. In practice, the general strategy is to graph the Hill estimator for all possible values of k (numbers of excesses through the threshold). Practical experience suggests that the best options for k are relatively small —for example, between 10 and 50 of statistical orders in a sample of size 1000. In Figure 34, the Hill estimator is estimated for market returns of the shape parameter ξ. We expect to find a stable region for the Hill estimator where estimations are constructed based on the different numbers of statistical order. In Figure 34, the upper horizontal axis represents the threshold associated with the possible values of k; in the lower horizontal axis, the number of observations included in the estimation is represented, and finally the confidence interval is observed at 95% (dotted lines). According to the results, it is observed that the estimation of the shape parameter stabilizes at a level of ξ, well above zero, in the area of 0.30. It should be borne in mind that in practice, the ideal situation does not usually occur if the data does not come from a distribution with a tail that changes with regularity. If this occurs, the Hill method is not appropriate. The serial dependence on the data can also impair the performance of the estimator, although this can also be said of the estimator of the GPD.

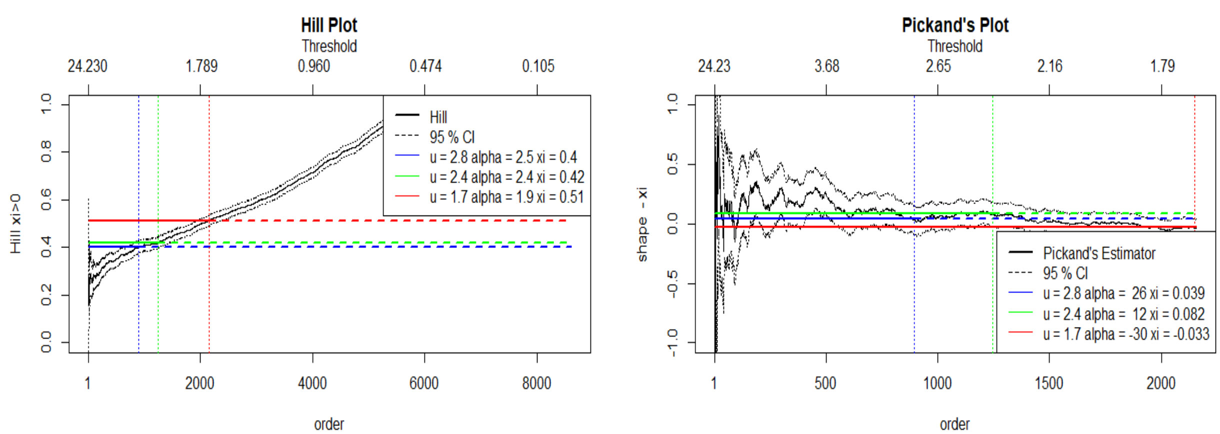

Figure 35. Left Panel: Hill Plot for ASE Index. Right Panel: Pickand's Plot for ASE Index

Predetermining some levels of threshold, we can distinguish that the shape parameters as estimated by Hill estimator in different threshold levels exhibit a very fat distribution behavior. On the other hand, Pickand's estimator infers light tailed behavior, almost exponentially decaying.

Like all tail index estimators, the Hill estimator depends on the threshold, and it is not clear how it should be chosen. A useful heuristic here is that k is usually less than 0.1*n*. Methods exist that choose k by minimizing the asymptotic mean squared error of the Hill estimator. Although it works very well for Pareto distributed data, for other regularly varying distribution functions the Hill estimator becomes less effective.

Application 2 >>> BEGCGA Index

The dataset consists of 5,455 daily logarithmic returns of the BEGCGA Index that cover the 21-year period of 31 December 1999 to 31 December 2021.

In this section, we proceed to an exploratory data analysis, followed by the application of the Block Maxima and POT approach.

Time Series Plot and Descriptive Statistics

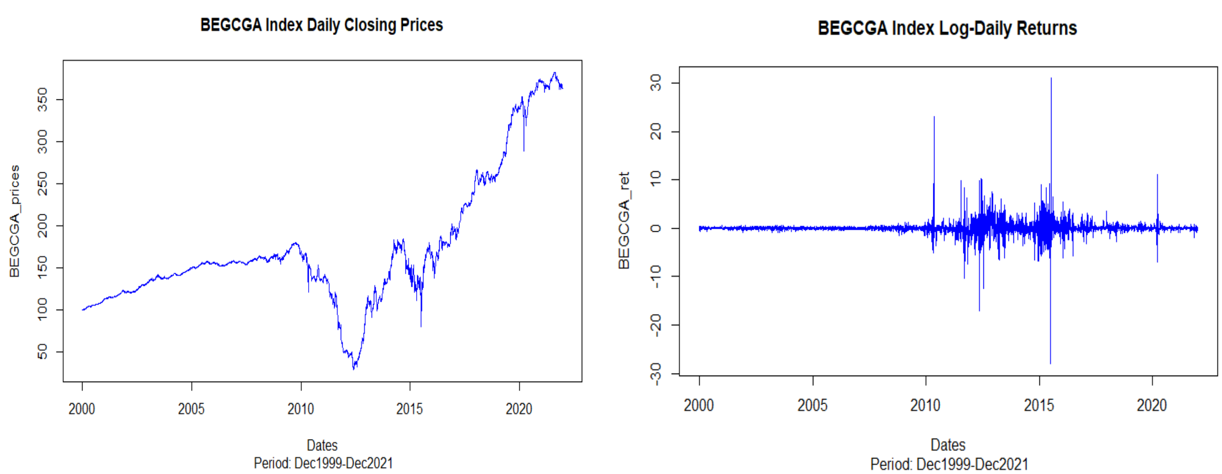

Figure 36. Left panel: Daily closing prices of the BEGCGA Index. Right Panel: Log-daily returns of the BEGCGA Index

Figure 36 shows the daily closing prices of the BEGCGA Index over a 21-year period. The process is non-stationary. Many empirical studies on these types of series have indicated that an approximation to stationarity can be obtained by taking logarithms of ratios of successive observations, or the so-called log-daily returns as in Figure 36, Right Panel.

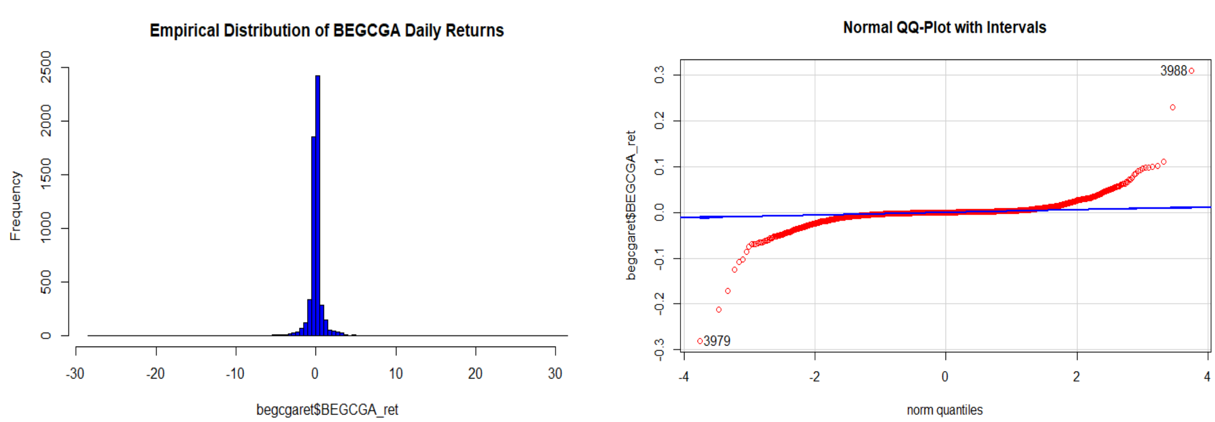

Figure 37. Left Panel: Histogram of BEGCGA Index. Right Panel: Normal QQ-Plot of BEGCGA Index

From the above Figure we can see that the data are extremely peaked in middle and the QQ-Plot reveals a distribution with fat tails since both ends of the plot deviate from the straight.

To start, Table 8 shows the summary statistics of the daily log returns, weekly and monthly maximum, and minimum log return series of the BEGCGA Index, as well as the descriptive statistics of the left and right tail of the sample distribution.

Looking at Table 8, we can see that in the daily return the bond index has a mean of 0.023% and a standard deviation of 1.34%, with a maximum of 31.02% and a minimum of -28.07%. In terms of the maximum returns, the weekly series has a mean of 0.85% and a standard deviation of 1.77%, with a maximum return of 31.02% and a minimum of 1.56%. Concerning monthly maxima, we distinguish a mean of 1.69% with 3.01% standard deviation, 31.02% maximum and 0.15% minimum. In the minimum return BEGCGA, the weekly series has a mean of -0.75% with a standard deviation of 1.65%, and a minimum of -28.07% and a maximum of 1.03%. Concerning monthly minima, the mean is -1.57% with 2.87% standard deviation, the minimum is -28.07% and the maximum is -0.02%.

Table 8. BEGCGA Index – Explanatory Data Analysis

		Weekly		***Monthly***		***Daily***	
	All Data	*Minima*	*Maxima*	*Minima*	*Maxima*	*Left Tail*	*Right Tail*
N (obs)	5,640	1,148	1,148	264	264	2,549	3,091
Mean Return	0.0229	-0.7514	0.8500	-1.5688	1.6952	0.5770	0.5177
Variance	1.7940	2.7185	3.1330	8.2225	9.0723	1.5691	1.4383
St. Dev	1.3394	1.6488	1.7700	2.8675	3.0120	1.2526	1.1993
Min	-28.0688	-28.0688	-1.5627	-28.0688	0.1487	0.0026	0.0028
1st Quartile	-0.1742	-0.7139	0.1633	-1.44367	0.3639	0.0841	0.0781
Median	0.0261	-0.2716	0.3172	-0.6132	0.6006	0.2037	0.1883
3rd Quartile	0.2164	-0.1041	0.7396	-0.3416	1.8273	0.5467	0.4277
Max	31.0219	1.0284	31.0219	-0.0185	31.0219	28.0688	31.0219
IQR	0.3906	0.6099	0.5763	1.1021	1.4633	0.4626	0.3496
Mode	-0.3482	-28.0688	-1.5627	-28.0688	0.1487	0.0026	0.0028
Skewness	0.6151	-7.8885	7.7101	5.345162	5.5042	8.9977	10.1637
Kurtosis	130.1404	101.2445	102.3282	38.92373	45.6575	141.2343	188.6239
J-B Test (p-value)	3799059 (0.000)	489393 (0.000)	498880 (0.000)	17296 (0.000)	21365 (0.000)	2156385 (0.000)	4641572 (0.000)
ADF Test (p-value)	-16.537 (0.01)	-5.2486 (0.01)	-4.7557 (0.01)	-3.3254 (0.067)	-3.6354 (0.03)	-13.582 (0.01)	-19.808 (0.01)

Notes: (1) J-B normality testing, (2) ADF stationarity testing
(3) The sign of the left tail has been changed so that positive values correspond to losses

From the descriptive statistics of all data, we can see a strong *positive skewness* of our full data series. Even though the mean of the series is less than the median which indicates a negative skewness, the mode of the distribution is less than both mean and median, a sign of the right skewness of our data.

Also, there is a big *value of kurtosis* which corresponds to the great extremity of the data, or to a distribution with fat tails.

We can *reject the normality assumption* based on the J-B test since the associated probability of the JB test is lower than 0.05 (the default significance level assumed for all tests in this thesis).

Also, *ADF test* indicates *rejection of the unit-root* null model in favor of the alternative model, that is the series is stationary.

Fitting the GEV Distribution in BEGCGA Index Log Returns

Conducted a Likelihood Ratio (LR) test, it is possible to conclude if the Gumbel distribution assumption is rejected or not, depending on the probability associated with the test. According to the below table, we reject the null hypothesis. This means that we do not assume that the series follows a Gumbel process. Therefore, we can conclude that none of the maximum/minimum return series have been generated by a Gumbel distribution. This means that for the considered series, only the Fréchet or Weibull particular cases of the more general GEV distribution remain.

Table 9. LR test for Gumbel distribution

	Likelihood Ratio Test	*p-value*	*X^2- critical value*	*Alternative*
Weekly Min Return	1437.3	< 2.2e-16	3.8415	Fréchet
Weekly Max Return	1441.6	< 2.2e-16	3.8415	Fréchet
Monthly Min Return	314.92	< 2.2e-16	3.8415	Fréchet
Monthly Max Return	361.39	< 2.2e-16	3.8415	Fréchet

Table 10 summarizes point estimates and the single confidence intervals for the GEV distribution for the BEGCGA Index. Given the large positive estimates for the shape parameter, for all our dataset, we conclude that the Fréchet distribution seems to be the most appropriate to describe that kind of data. Confidence intervals for shape parameters reveal that we can accept the alternative hypothesis of the upper unbounding type of distribution, in the confidence level of 0.95%. Consequently, we can conclude that BEGCGA Index return empirical distribution has fat tails. The larger the estimate values for the shape parameter, the more fat-tailed the distribution. Based on shape parameter estimation, we can estimate a tail index α below 2, or we expect at least the existence of the first 1 moment. The possibility of an infinite theoretical variance makes the model rather useful since we cannot make inferences by using the sample variance in standard models.

Table 10. BEGCGA INDEX Block Maxima Method

	Weekly Min Returns	Weekly Max Returns	Monthly Min Returns	Monthly Max Returns
	Method MLE			
Negative Log-Likelihood Value	636.5258	739.4533	324.2651	318.0868
Location (ses)	0.1751 (0.0075)	0.2335 (0.0083)	0.4712 (0.0305)	0.4834 (0.0269)
CIs	[0.1604-0.1898]	[0.2173-0.2498]	[0.4115-0.5309]	[0.4306-0.5362]
Scale (ses)	0.2144 (0.0094)	0.2460 (0.0098)	0.4459 (0.0361)	0.3921 (0.0363)
CIs	[0.1960-0.2327]	[0.2268-0.2653]	[0.3752-0.5166]	[0.3210-0.4632]
Shape (ses)	0.9328 (0.0424)	0.81776 (0.0352)	0.7857 (0.0645)	0.9906 (0.0744)
CIs	[0.8498-1.0159]	[0.7486-0.8865]	[0.6593-0.9122]	[0.8448-1.1364]
AIC	1,279.052	1,484.907	654.5301	642.1736
BIC	1,294.189	1,500.044	665.258	652.9014
K-S test (two-sample)	0.047725 (0.7126)	0.039199 (0.3410)	0.079545 (0.3738)	0.090909 (0.2253)
	Return Level for Period (Units)			
5-units [CIs]	0.88 [0.78-0.97]	0.96 [0.87-1.05]	1.75 [1.42-2.07]	1.84 [1.45-2.22]
10-units [CIs]	1.82 [1.55-2.09]	1.83 [1.59-2.06]	3.23 [2.43-4.03]	3.77 [2.65-4.88]
20-units [CIs]	3.61 [2.90-4.33]	3.34 [2.79-3.90]	5.76 [3.91-7.61]	7.59 [4.63-10.55]
	Method L-Moments			
Location	0.2133	0.2605	0.5153	0.5731
CIs	[0.1908-0.2378]	[0.2336-0.2862]	[0.4441-0.6119]	[0.4981-0.6785]
Scale	0.2841	0.3020	0.5492	0.5864
CIs	[0.2598-0.3157]	[0.2734-0.3368]	[0.4770-0.6731]	[0.5013-0.7271]
Shape	0.5872	0.5962	0.5807	0.5798
CIs	[0.4793-0.7028]	[0.4884-0.7291]	[0.3926-0.7557]	[0.3795-0.7742]
K-S test (two-sample)	0.083333 (0.1015)	0.068815 (0.0087)	0.10227 (0.1264)	0.14394 (0.0084)
	Return Level for Period (Units)			
5-units	0.89 [0.83-0.97]	0.99 [0.91-1.08]	1.83 [1.55-2.12]	1.97 [1.69-2.30]
10-units	1.54 [1.37-1.71]	1.69 [1.49-1.88]	3.06 [2.43-3.70]	3.29 [2.68-4.06]
20-units	2.50 [2.12-2.98]	2.73 [2.30-3.16]	4.88 [3.57-6.47]	5.22 [3.91-6.95]
	Sources: r-packages ("evd", "evir", "extRemes", "fExtremes", "mev", "lmom", "extremeStat", "ismev")			

Confidence intervals for shape parameters reveal that we can accept the alternative hypothesis of the heavy-tailed distribution, in the confidence level of 0.95%. All confidence intervals do not include zero or negative value and so are statistically compatible with thick tailed distributed extreme returns. Point estimate also indicate fat tailed distributions.

The GEV distribution estimates, according to Table 10, yield slightly different results for the method of ML and L-moments. L-moments method exhibits lower estimates for the shape parameter with lower variability, in all four-time series, relative to the MLE. However, Kolmogorov-Smirnov distance seems less reliable taking account of the standard errors. Confidence intervals also are wider than in the case of MLE. However, L-moments also indicates a tail index $\alpha < 2$, or infiniteness of second moment.

Fitting the GEV Distribution in Weekly Block Minima

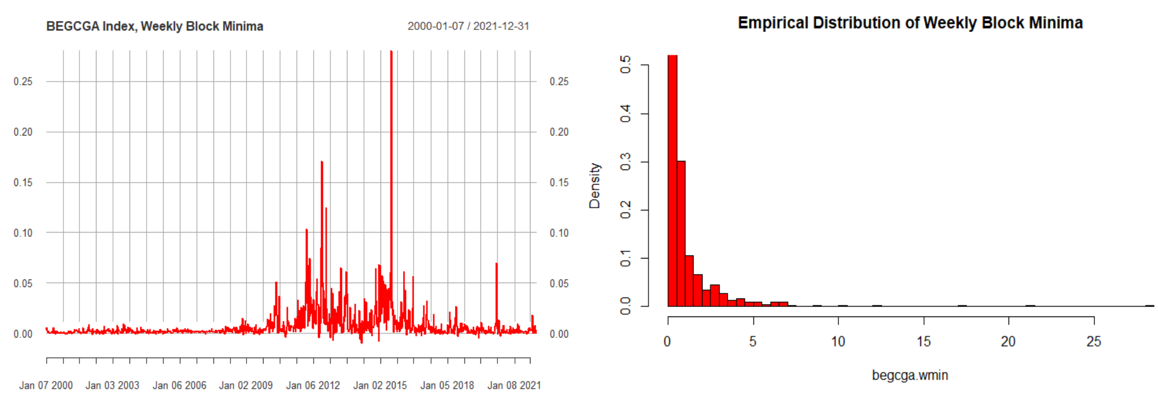

Figure 38. Left panel: weekly block minima of the BEGCGA Index. Right Panel: distribution of weekly block minima of the BEGCGA Index

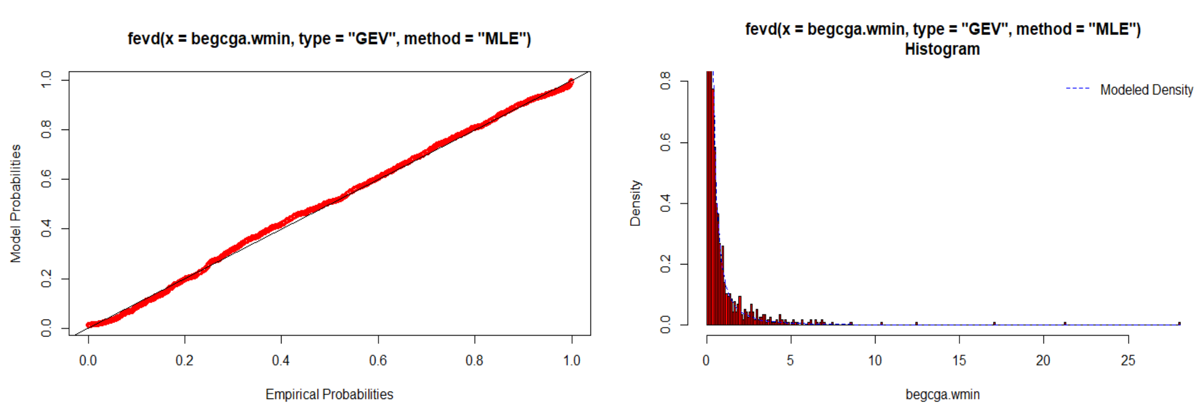

Figure 39. Left Panel: Fit model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

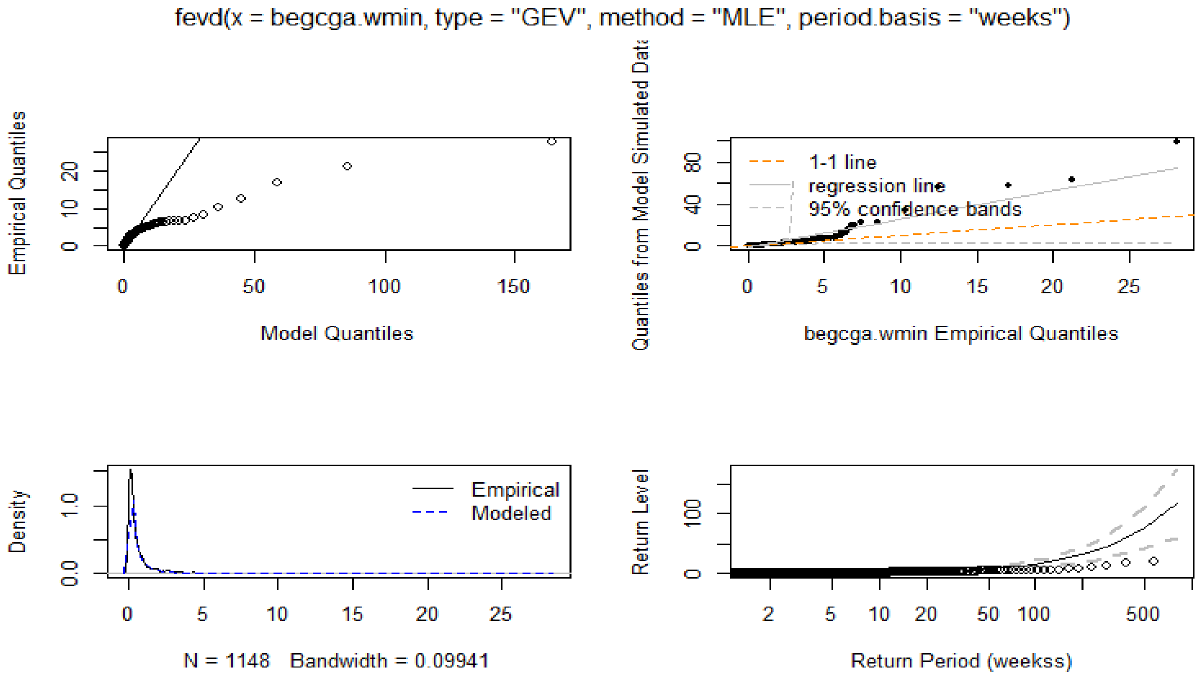

Figure 40. Diagnostic Plots from the GEV df fit to BEGCGA weekly minima; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right)

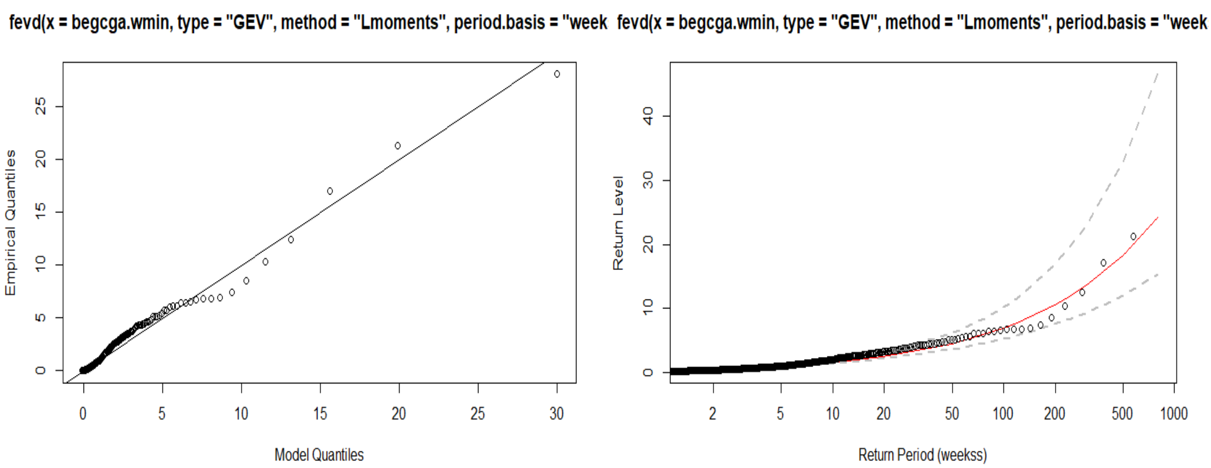

Figure 41. Left Panel: QQ-Plot of BEGCGA weekly minima; Right Panel: Return Level Plot; L-moment method

For the BEGCGA weekly minimum returns, we conclude that the Fréchet distribution seems to be the most appropriate to describe that kind of data, as the shape (ξ) estimate is higher than zero, more precisely $\hat{\xi}$ = 0.932859 by MLE and $\hat{\xi}$ = 0.587179 by L-Moments method or PWM. Consequently, we can conclude that BEGCGA weekly minimum return empirical distribution has fat tails. The larger the estimate values for the shape parameter, the more fat-tailed the distribution.

The above diagnostic plot indicates a rather poor fit for the model. Even though the probability plot gives us good fitness, the quantile plot and the return level plot show high dispersion in extremities. Applying the L-moment method we can see a more reliable fitting. QQ-plot and the return level plot approximate better the empirical data as shown in the above figures. However, in both estimates we distinguish the clear concavity of QQ-Plot.

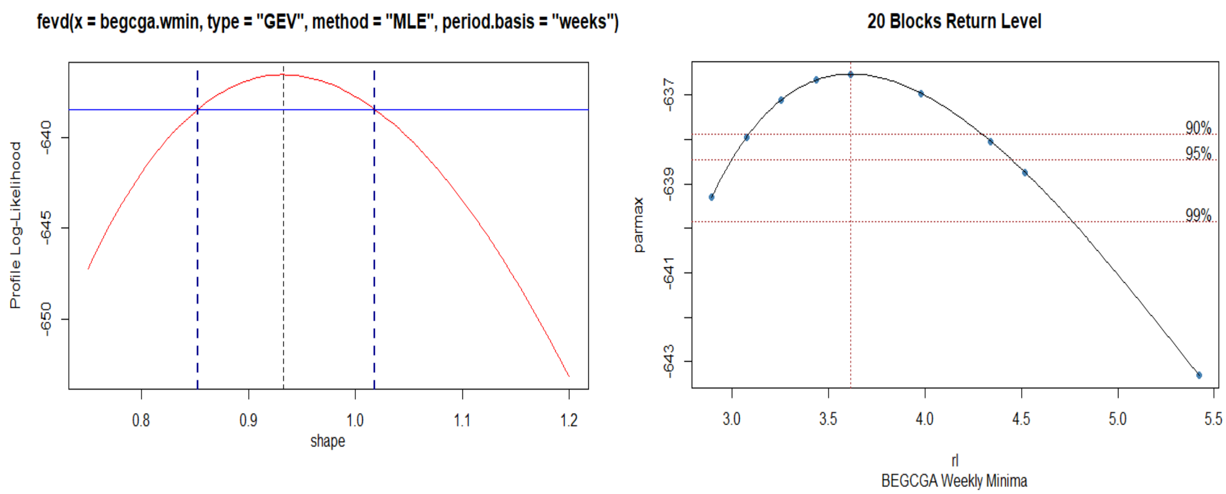

Figure 42. Left Panel: Profile Likelihood for shape parameter. Right Panel: Profile Likelihood for 20-week return level

A profile-likelihood (solid blue line) plot for the return level estimate for the GEV df fit to weekly minimum returns. The red vertical (dashed line) line emphasizes where the location of the MLE for the 20 -period return level, in different confidence levels. By using the profile-likelihood method we get an improved 95% confidence interval for the shape parameter of about (0.8528, 1.0178), very similar to the normal approximation limits (Table 10). We also get that approximate 95% confidence interval for the 20-week return level based on the profile-likelihood is about (3.10, 4.29), which is close to the normal approximation limits.

The asymmetry of the profile likelihood function curve increases with the return period, which can reflect the effect of the length of sample series and return periods on confidence interval.

Fitting the GEV Distribution in Weekly Block Maxima

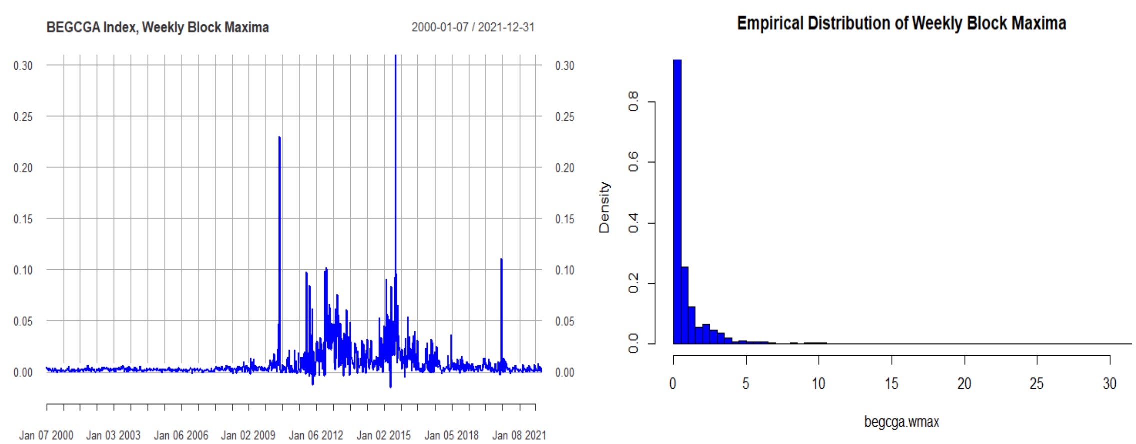

Figure 43. Left panel: Weekly block maxima of the BEGCGA Index. Right Panel: distribution of weekly block maxima of the BEGCGA Index

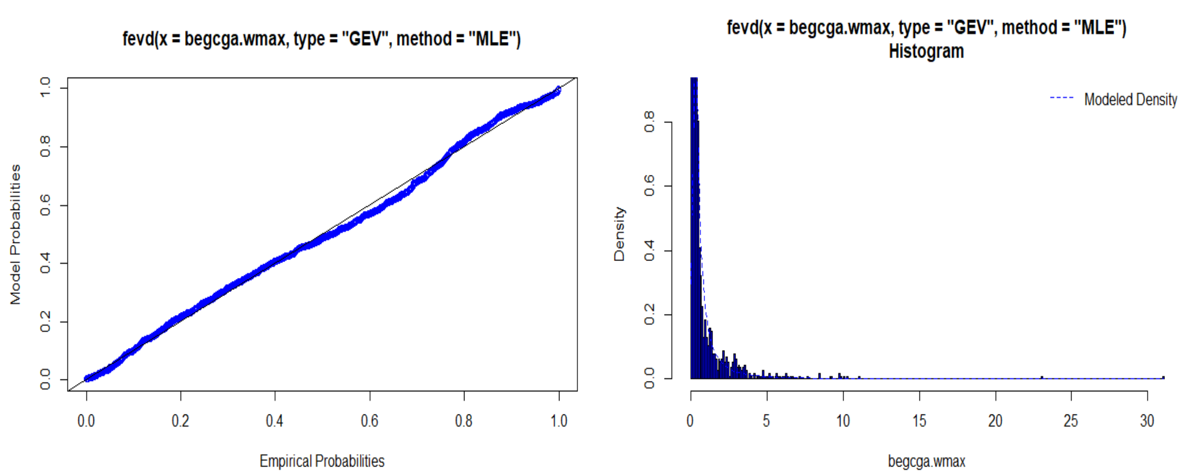

Figure 44. Left Panel: Fit model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

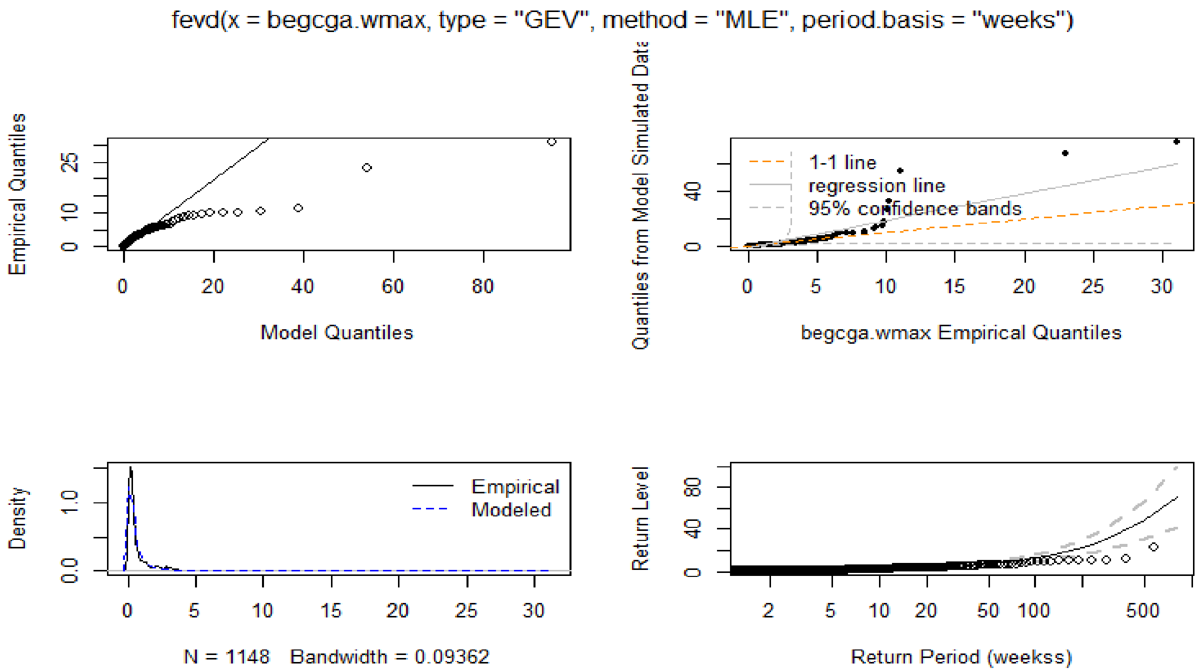

Figure 45. Diagnostic Plots from the GEV df fit to BEGCGA weekly maxima; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right)

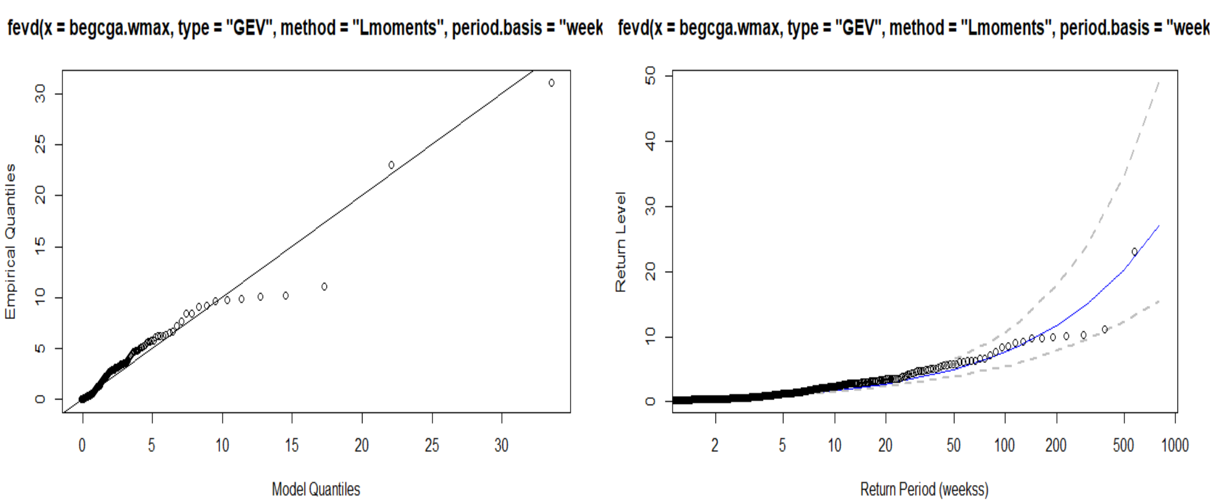

Figure 46. Left Panel: QQ-Plot of BEGCGA weekly minima; Right Panel: Return Level Plot; L-moment method

For the BEGCGA weekly maximum returns, we conclude that the also Fréchet distribution seems to be the most appropriate to describe that kind of data, as the shape (ξ) estimate is higher than zero, more precisely $\hat{\xi}$ = 0.82 by MLE and $\hat{\xi}$ = 0.60 by L-Moments method or PWM. Consequently, we can conclude that BEGCGA weekly maximum return empirical distribution has fat tails. Applying the L-moment method we can see a more reliable fitting concerning the QQ plot and the return level plot.

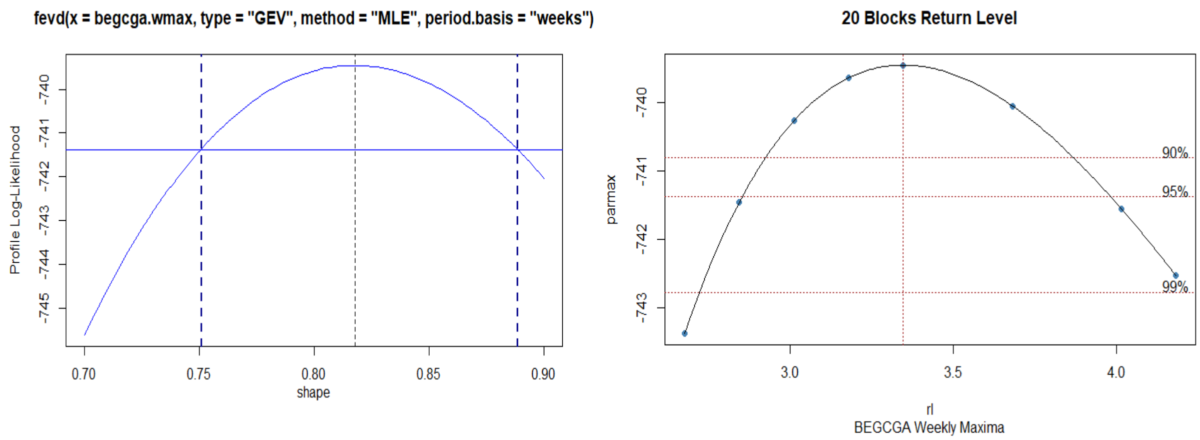

Figure 47. Left Panel: Profile-Likelihood for ξ in the BEGCGA weekly maxima. Right Panel: Profile-Likelihood of 20-period return level

By using the profile-likelihood method we get an improved 95% confidence interval for the shape parameter of about (0.751, 0.8883), like the normal approximation limits (Table 10). We also get that approximate 95% confidence interval for the 20-week return level based on the profile-likelihood is about (2.93, 3.86), slightly different to the normal approximation limits.

Fitting the GEV Distribution in Monthly Block Minima

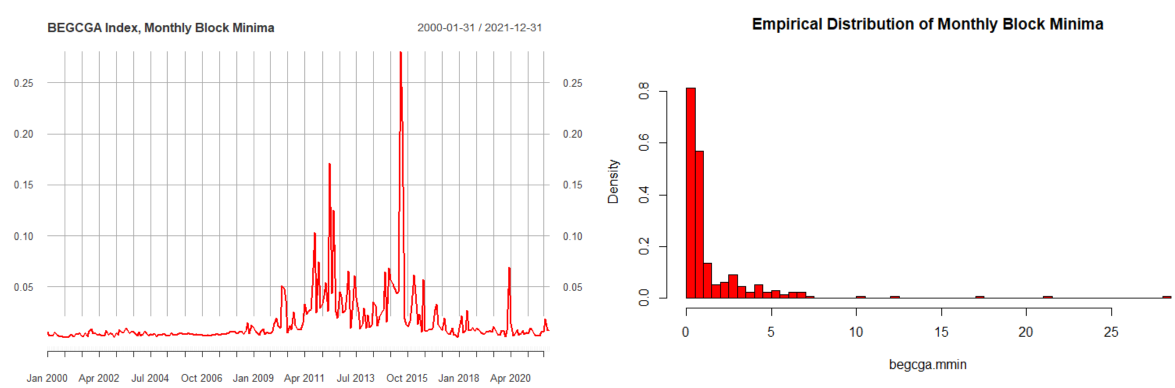

Figure 48. Left panel: monthly block minima of the BEGCGA Index. Right Panel: distribution of monthly block minima of the BEGCGA Index

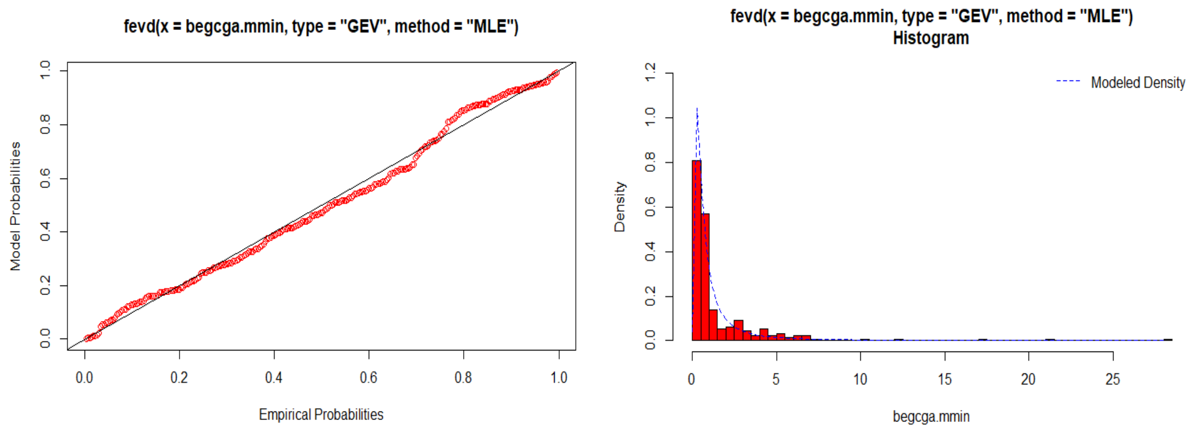

Figure 49. Left Panel: Fit model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

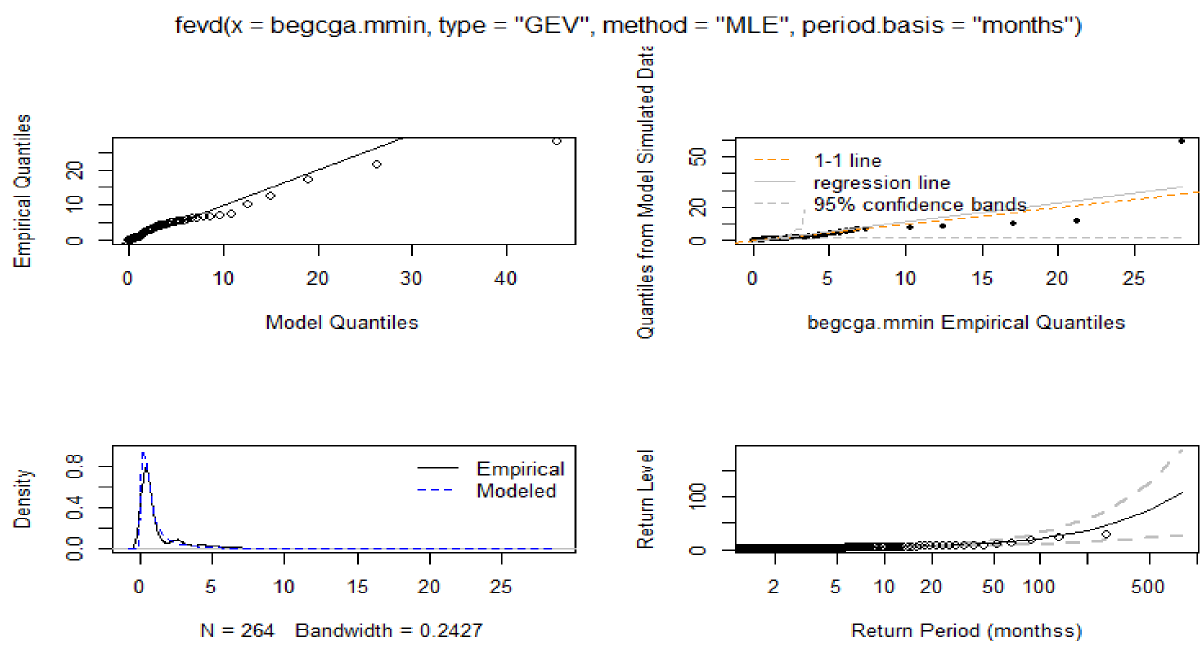

Figure 50. Diagnostic Plots from the GEV df fit to BEGCGA monthly minima; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right)

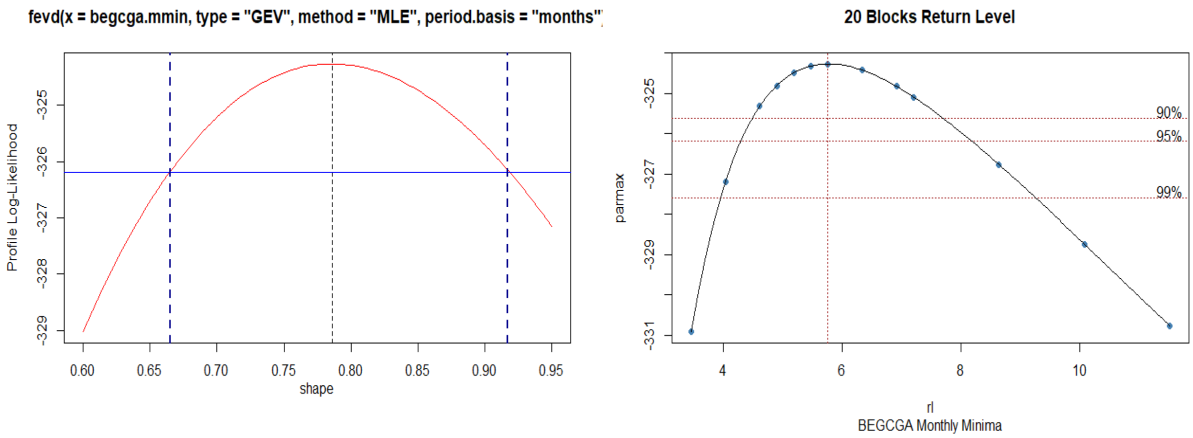

Figure 51. Left Panel: Profile-Likelihood for ξ in the BEGCGA monthly minima. Right Panel: Profile-Likelihood of 20-period return level

The profile-likelihood method gives us an improved 95% confidence interval for the shape parameter of about (0.6654, 0.9168), slightly different only to the left bound. We also get that approximate 95% confidence interval for the 20-week return level based on the profile-likelihood is about (4.51, 7.67).

As in ASE index application, applying the MLE method to monthly data exhibits more reliable fitting, than the weekly ones. Both QQ-Plot and return level plot fit the data well.

Fitting the GEV Distribution in Monthly Block Maxima

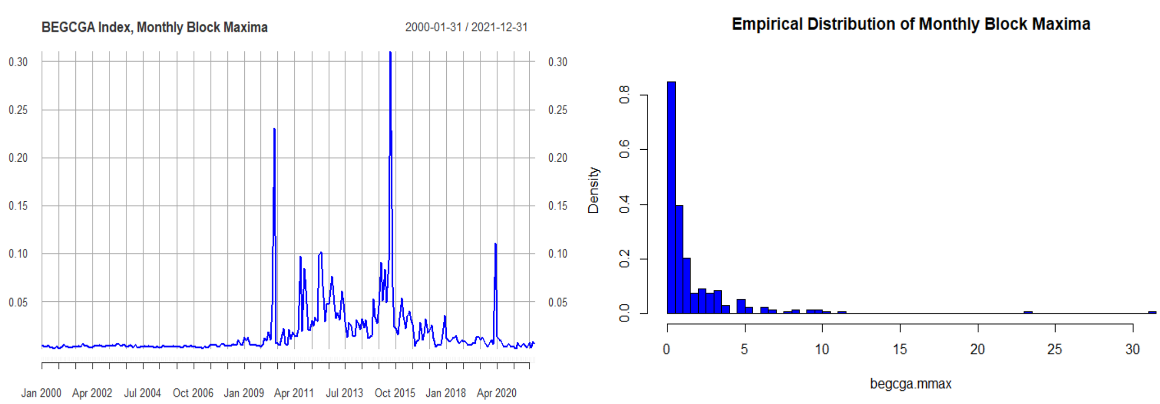

Figure 52. Left panel: monthly block maxima of the BEGCGA Index. Right Panel: distribution of monthly block maxima of the BEGCGA Index

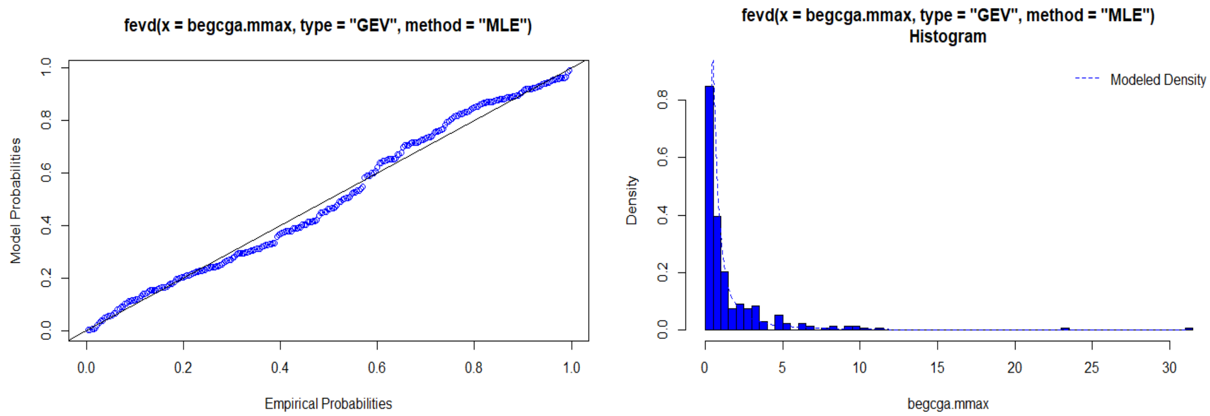

Figure 53. Left Panel: Fit model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

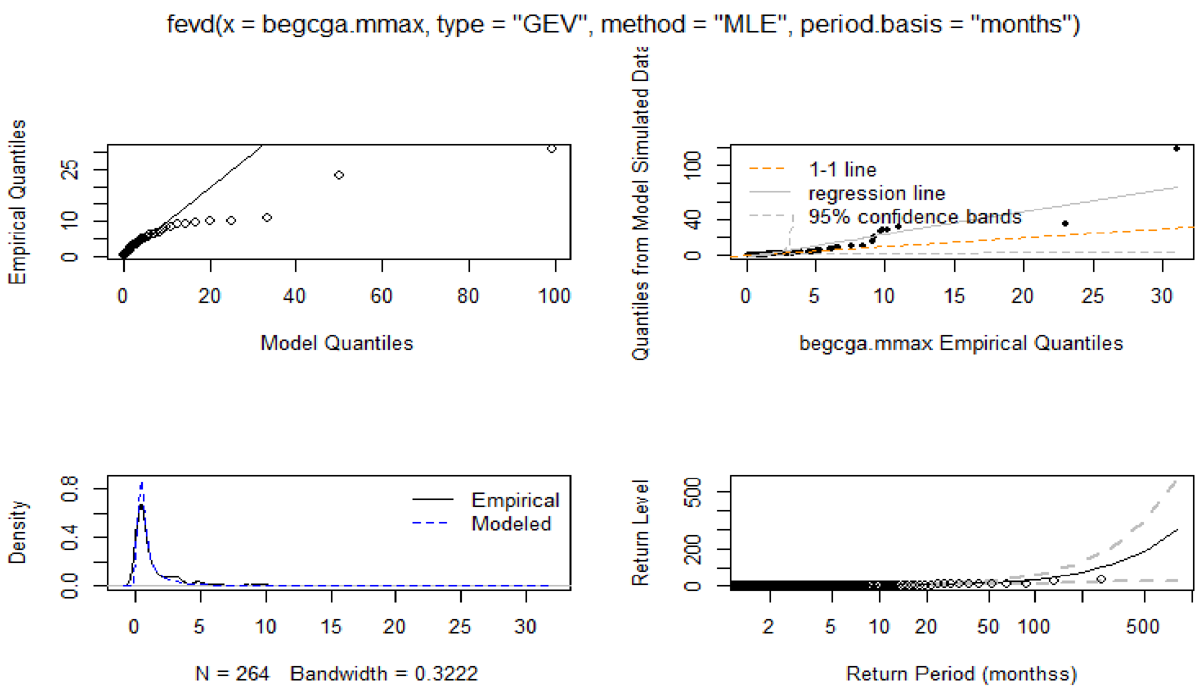

Figure 54. Diagnostic Plots from the GEV df fit to BEGCGA monthly maxima; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GEV df (top right); (c) density plots (bottom left); (d) return level plot (bottom right)

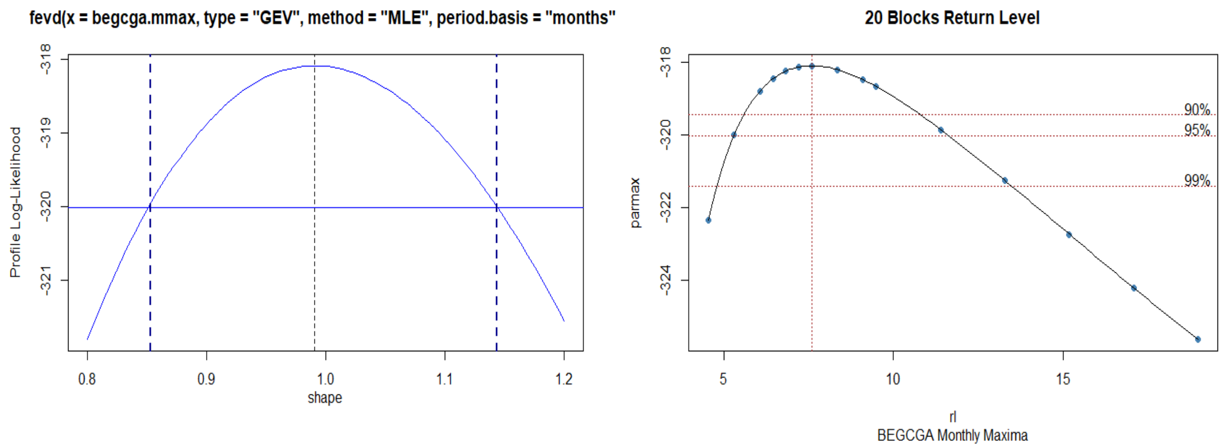

Figure 55. Left Panel: Profile-Likelihood for ξ in the BEGCGA monthly maxima. Right Panel: Profile-Likelihood of 20-period return level

The profile-likelihood method gives us an improved 95% confidence interval for the shape parameter of about (0.8531, 1.1432), slightly different to the delta method. We also get that approximate 95% confidence interval for the 20-week return level based on the profile-likelihood is about (5.64, 10.78).

Fitting a GP Distribution in BEGCGA Index Log Returns

Parametric Estimation

Running different approaches to find the appropriate threshold, we ended up to take the largest 15% from the empirical distribution, and then subtracting off the 85% quantile to get exceedances. This produces a threshold of 0.80%. We will use this threshold to fit a GP distribution to the left tail of our observations, as in the full data set. We use slightly different u = 0.85 to the right tail.

We can approximate threshold by visualizing our data.

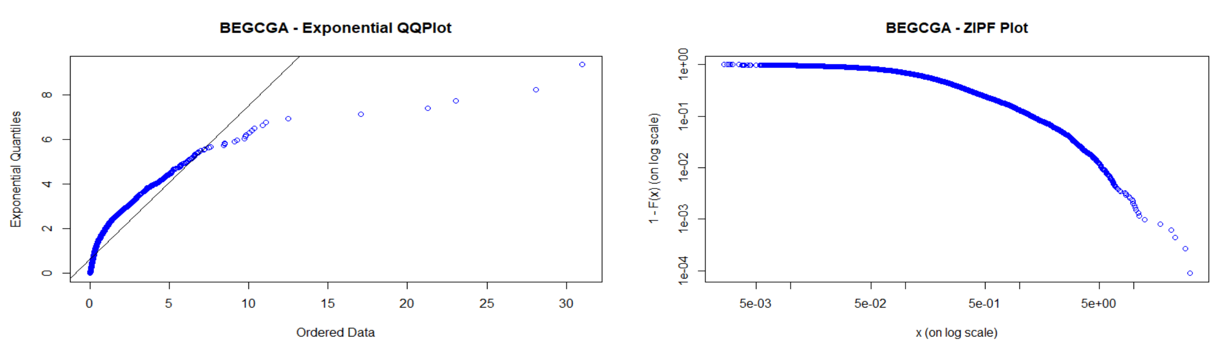

Figure 56. Left Panel: BEGCGA Exponential QQ-Plot. Right Panel: Log-Log Plot (Zipf Plot)

First, the clear concavity in the *exponential QQ-plot* (in the tail of the distribution) is a strong signal of the presence of heavy tails. The exponential is obtained by setting the parameter xi to 0. To look for the behavior of the survival function we use a *Zipf plot (Log-Log Plot)*. A negative (decreasing) linear slope is present. This is the first signal of the fat tailed nature of the data. But a Zipf plot verifies a necessary yet not sufficient condition.

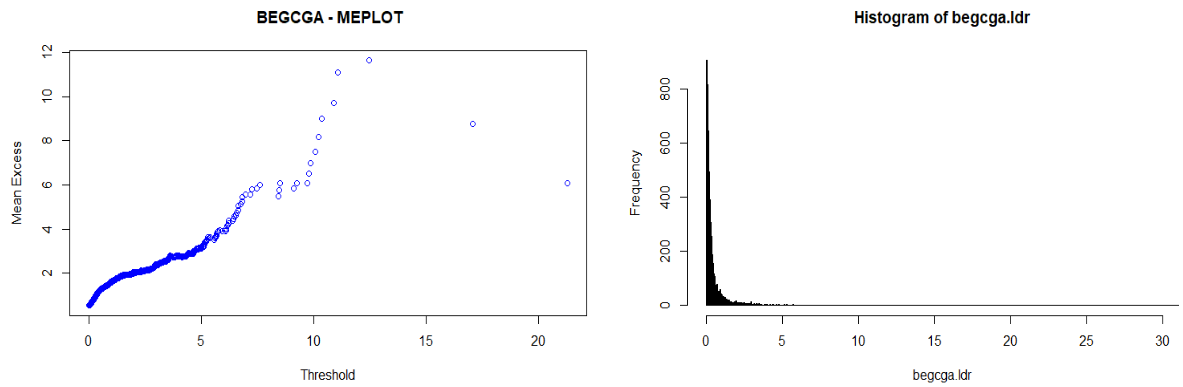

Figure 57. BEGCGA Index Mean Excess Plot and histogram of positive returns

The above left figure is another presence of a fat tail behavior. We observe the increasing linear behavior. The above *mean excess plot* appears to curve from $u = 0$ to $u \approx 1$, beyond which it is approximately linear. This leads to the conclusion that, for $u > uo$, $E(X - u \mid X > x)$ is a linear function of u. According to that, the threshold excesses are expected to change linearly with u at threshold levels for which the generalized Pareto model is suitable.

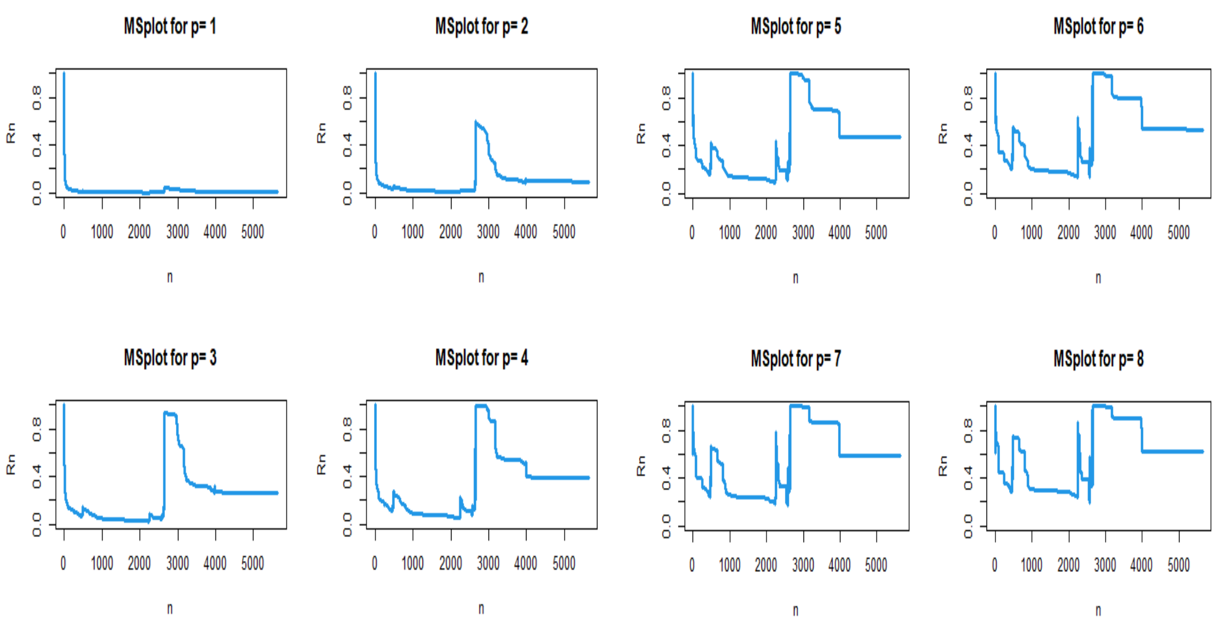

Figure 58. MS Plot: A simple code to check for the first 8 moments

The Maximum to Sum Plot (MS Plot) or the ratio of maximum and sum indicates the existence of at least the first moment. This means that the possible tail index varies between 1 to 2.

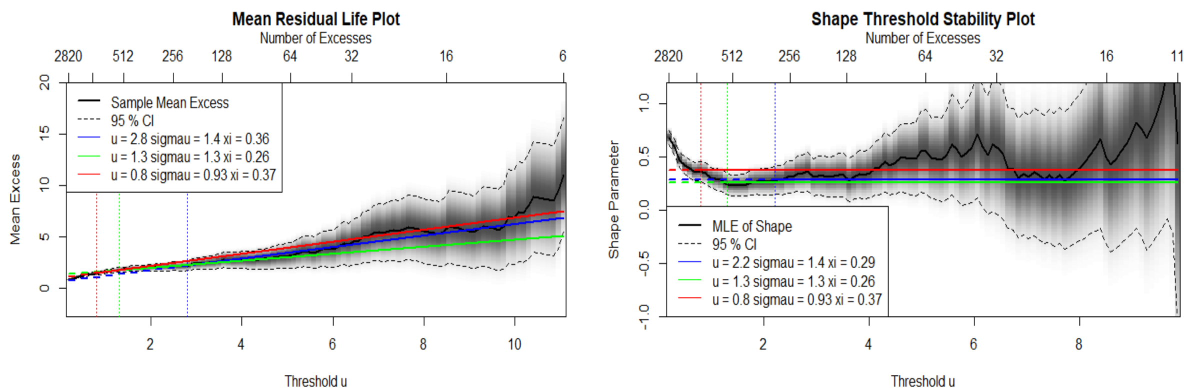

Figure 59. Left Panel: Mean Residual Life Plot. Right Panel: Shape Threshold Stability Plot

For the graphical analysis, the empirical estimates of the sample mean excess are plotted against a length of thresholds. The threshold level is chosen to be the lowest point where the sample mean excess is approximately lying on a straight line. The inference of a threshold level is not always a simple task of visual interpretation of the MRL plot, and it may require an extra procedure, which is the parameter estimation. The left panel of the Figure 59 represents the MRL plot for the BEGCGA Index returns. Solid jagged line is the empirical MRL with approximate 95% confidence interval (dashed lines). The MRL suggested by maximum likelihood parameter estimates for the threshold u = 0.80, 1.3 and 2.8 are upper, middle, and lower colorful solid straight line. Vertical dashed lines indicate the thresholds selected. The sampling density for the MRL is shown by a greyscale image, where lighter greys indicate low density. The threshold level choice *u* = 0.80 gives 938 exceedances and is sustained by the empirical MRL being close to linear above this level.

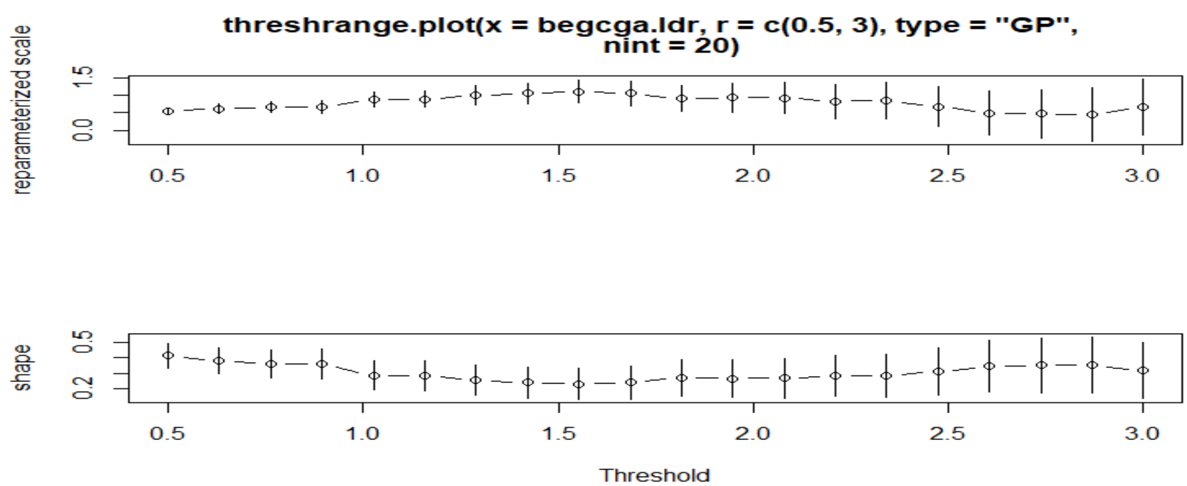

Figure 60. Threshold Sensitivity Plot

A complementary approach to the MRL plot diagnostic is to fit a GP distribution at a range of thresholds and observe the stability of parameter estimates.

Once the threshold is set, parameters can be estimated through the Maximum Likelihood Method (MLE)

Table 11. Fitting GPD to BEGCGA Index – Parametric Estimation

	Method: MLE		
	All Dataset (u = 0.8)	*Left Tail (u = 0.80)*	*Right Tail (u = 0.85)*
Negative Log-Likelihood Value	1,213.935	588.2595	602.251
	Estimated parameters		
scale (ses)	0.9262 (0.058)	0.9121 (0.0747)	0.9600 (0.0785)
scale CIs	[0.8228-1.0297]	[0.7656-1.0586]	[0.8060-1.1139]
shape (ses)	0.3708 (0.0482)	0.3736 (0.0696)	0.3674 (0.0692)
shape CIs	[0.2763-0.4653]	[0.2372-0.5099]	[0.2317-0.5030]
AIC	2,43.871	1,180.519	1,208.502
BIC	2,441,558	1,188.777	1,216.738
	Return Level for Period (units)		
5-units [CIs]	16.43 [12.17-20.68]	19.64 [11.35-27.92]	12.80 [8.93-16.66]
10-units [CIs]	21.74 [14.98-28.50]	25.93 [12.98-38.88]	24.54 [12.97-36.10]
20-units [CIs]	28.61 [18.16-39.06]	34.08 [14.31-53.84]	32.16 [14.46-49.86]
Notes:	*Sources: r-packages ("evd", "evir", "extRemes", "fExtremes", "mev", "lmom", "extremeStat", "ismev", "eva")* *The sign of the left tail has been changed so that positive values correspond to losses*		

All confidence intervals do not include zero or negative value and so are statistically compatible with thick tailed distributed extreme returns. Point estimate also indicate fat tailed distributions.

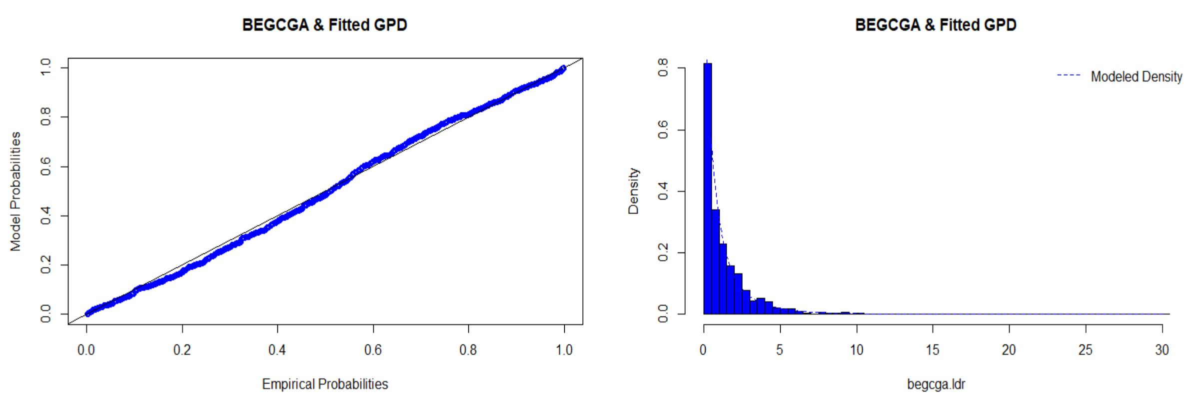

Figure 61. Left Panel: Fit model against Empirical Probabilities Plot. Right Panel: Data Histogram against Modeled Density

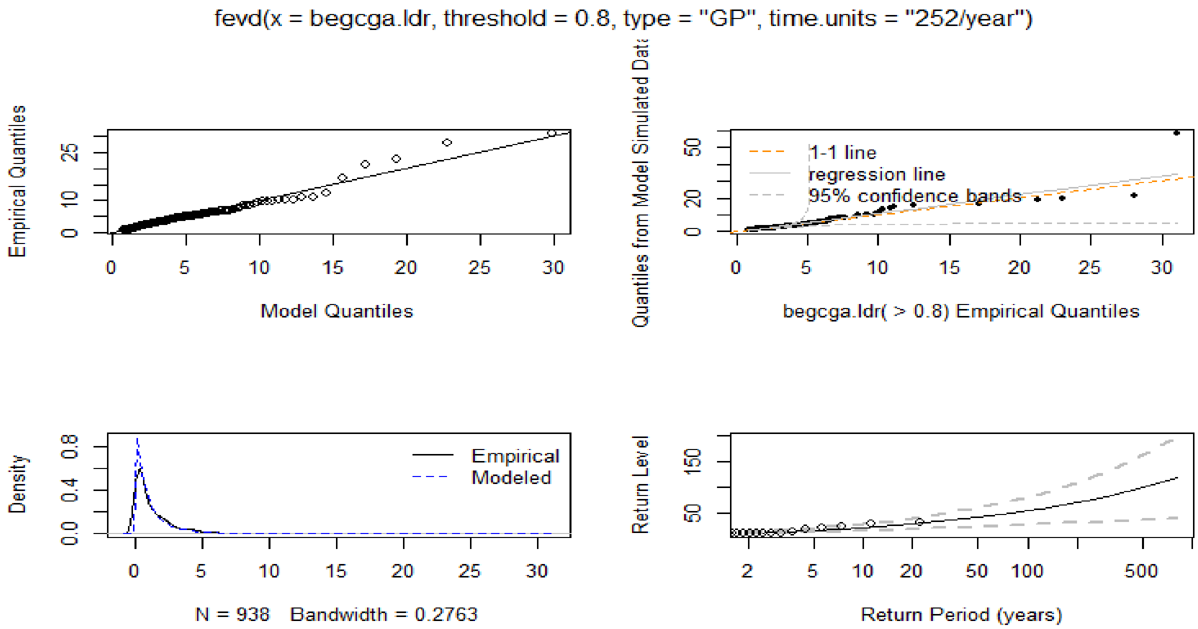

Figure 62. Diagnostic Plots from the GP df fit to BEGCGA sample data; (a) quantile-quantile plot (top left); (b) quantiles from a sample drawn from the fitted GP df (top right); (c) density plots (bottom left); (d) return level plot (bottom right)

Therefore, all the four diagnostic plots support the GPD model.

The point estimate of GP distribution implies the existence of a second moment as opposed to the GEV one.

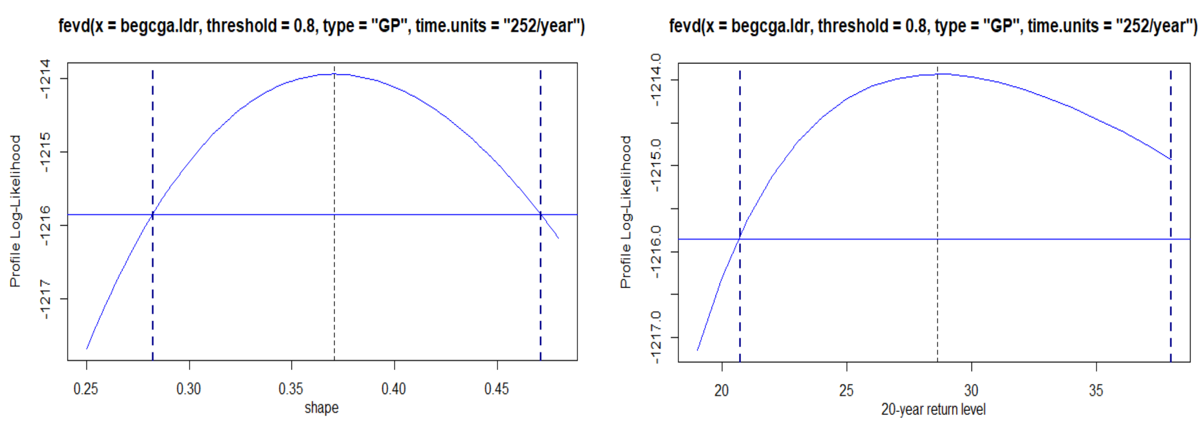

Figure 63. Left Panel: Profile-Likelihood for ξ in the BEGCGA sample size. Right Panel: Profile-Likelihood of 20-period return level

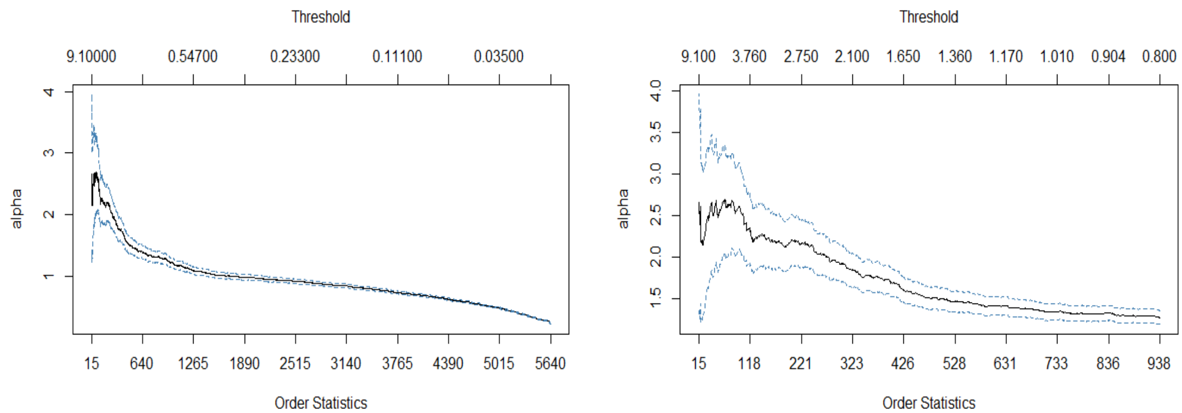

Figure 64. Estimates of the Hill ξ for the Daily BEGCGA Returns. Left Panel: Full data series; Right Panel: Data above threshold u = 0.8

We perform a non-parametric Hill estimation, expecting to find a stable region. The results related to the estimation of the parameter stabilize up to some order. As the statistical order increases, Hill tail-index becomes unstable. According to the results, it is observed that the estimation of the tail parameter stabilizes in such an area which indicates an alpha above 1.

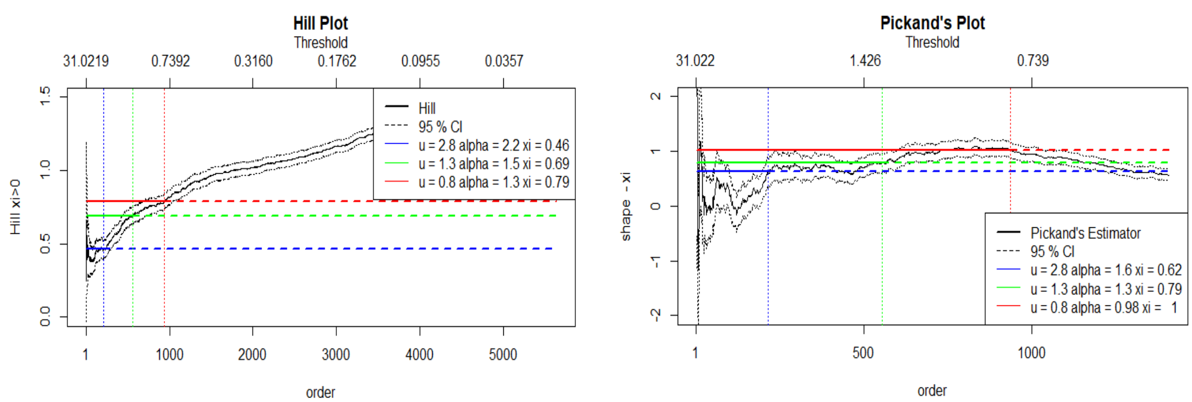

Figure 65. Left Panel: Hill Plot; Right Panel: Pickands Plot; Note: "evmix" package in r

By prearranging threshold levels, both estimators, Hill and Pickands, show a remarkable agreement in the high level of ξ, as in Figure above. In this application, both estimators exhibit an α-stable behavior, in different levels of thresholds. In both cases the shape parameter is high, and the tail index infers infinite second moment asymptotically.

Conclusions

EVT provides the necessary asymptotic theorems for explicitly modelling the tails of the distribution of returns. The probability framework starts with the famous Fisher-Tippett-Gnedenko theorem which characterizes the three types of max-stable distributions.

This thesis employs extreme value theory to examine the limiting distribution of extreme returns in Greek financial market, focusing on weather on not returns are governed by a heavy tailed stable distribution. The data for this exercise consists of the daily returns of an equity index, the Athens Stock Exchange Index, and a bond index, BEGCGA Index. In general, our results support this tail behavior.

Two alternative techniques are applied, the Blocks Maxima Method and the POT approach. Interestingly, when investigating what distribution is most appropriate for describing the underlying price returns, the Block Maxima approach and the POT model give rise to almost the same outcomes.

Each method has its advantages and issues. If the data set is large enough, GEV may prove useful as it can avoid dealing with data clustering issues, provided that blocks are sufficiently large. Furthermore, the estimation is simplified as the selection of a threshold u is not needed. In academic examples, using 20-30 years of daily returns is common, see for example, McNeil et al. (2005). However, from a practical viewpoint, it is arguable that observations so far back in the past have any relevance to the present market conditions. The POT method has its advantages in modelling the available data more efficiently than the BMM as it uses excesses over a threshold. However, the definition of a suitable high threshold for defining the generalized Pareto distribution is not an easy task. Frequently, there is more than one suitable threshold for modelling the tail fraction of extreme values. This was clearly observed in the graphical diagnostics for threshold choice.

After fitting the GEV distribution to all our maximum and minimum series, we conclude that the return series are well represented by the Fréchet distribution. This was already expected due to the same conclusion reached by other studies in financial research related to maximum and minimum returns, as financial series normally follow this distribution.

After applying the GPD distribution to our data, using the POT approach with a random threshold level for the positive (maximum returns) and negative (minimum returns) tail, another conclusion was the fact that we can assume that the maximum and minimum return series are well represented by this distribution. In using GPD, one is always faced with the classical tail-estimation tradeoff problem. The threshold issue may seem easy to resolve by resorting to statistical methods that would indicate where the tail begins. However, no reliable methods of this type exist. Typically, high threshold is chosen subjectively by looking at certain plots, such as the Hill plot or the mean excess plot, which are standard in EVT. Therefore, identifying the threshold between the body and the tail is a matter of subjective choice based on visual inspection, which cannot be achieved on a large scale.

The empirical results followed by this thesis allow us to determine the type of extreme value distribution. The asymptotic distributions belong to the domain of attraction of the Fréchet type. The thin tailed Gumbel and exponential distributions with rapidly decreasing tails are strongly rejected against the fat tailed Fréchet and Pareto distributions with slowly decreasing tails. In all cases, the shape index ξ is estimated to be positive and statistically different from zero.

Log-returns exhibit a complicated dependence structure. Therefore, the direct use of standard tail index estimates, as those of Hill or Pickands estimators, may become questionable. These estimators are very sensitive with respect to dependence in the data, even more when data are aggregated over weeks or months.

Our conclusion is that EVT can be useful for assessing the size of extreme events. From a practical point of view this problem can be approached in different ways, depending on data availability and frequency, the desired time horizon, and the level of complexity one is willing to introduce in the model. One can choose to use the BMM or the POT method, and finally rely on point or interval estimates. Generally, it is very difficult to make a final decision about the value of a tail index or the finiteness of a certain power moment. However, the applications considered in the present thesis show that log-returns have certain infinite power moments. This is seen by using simple diagnostic tools such as QQ or mean excess plots, and by more precise means as tail index estimators. Finally, we find it is worthwhile computing interval estimates as they provide additional information about the quality of the model fit.

References

Anderson, C., W., (1971), “Contributions to the Asymptotic Theory of Extreme Values”, PhD Thesis, University of London, London, UK

Bali, T. G. (2003a). “An extreme value approach to estimating volatility and Value-at-Risk”. Journal of Business 76(1):83–108.

Bali, T. G. (2003b). “The generalized extreme value distribution”. Economics Letters 79:423–427.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J., (2004b), Statistics of Extremes: Theory and Applications, Probability and Statistics, Wiley.

Beirlant, J., P. Vynckier and J. L. Teugels (1996). “Tail index estimation, Pareto quantile plots and regression diagnostics”. Journal of the American Statistical Association 91:1659– 1667.

Bekiros, S. D., and D. A. Georgoutsos (2005). “Estimation of Value-at-Risk by extreme value and conventional methods: a comparative evaluation of their predictive performance”. Int. Fin. Markets, Inst. and Money 15:209–228

Bekiros, S. D., and D. A. Georgoutsos (2008a). “Extreme returns and the contagion effect between the foreign exchange and the stock market: evidence from Cyprus”. Applied Financial Economics 18:239–254.

Billingsley, P., (1995), Probability and Measure, Wiley, 3ème edition.

BUCHER, A., AND J., SEGERS, (2017), ON THE MAXIMUM LIKELIHOOD ESTIMATOR FOR THE GENERALIZED EXTREME-VALUE DISTRIBUTION, arXiv:1601.05702v3 [math.ST] 15 Mar 2017,

Coles, S., (2004), An Introduction to Statistical Modeling of Extreme Values, Springer Series in Statistics, Springer,

Embrechts, P., C. Klüppelberg and T. Mikosch, (1997), Modelling Extremal Events for Insurance and Finance, Berlin: Springer

Fallahgoul, H., and G., Loeper, (2018), “Modelling Tail Risk with Tempered Stable Distributions: An Overview”, Monash CQFIS working paper

Fama, F. (1965), “The Behavior of Stock Market Prices”, Journal of Business, 38, 34-105

Fama, F. and R., Roll, (1968), “Some Properties of Symmetric Stable Distributions”, Journal of the American Statistical Association, Sep. 1968, Vol. 63, No. 323 (Sep. 1968), pp. 817-836

Feller, W., An Introduction to Probability Theory and Its Applications, VOLUME II, SECOND EDITION, JOHN WILEY & SONS, INC., N.Y.

Fisher, R.A., Tippett, L.H.C., (1928), “Limiting forms of the frequency distribution of the largest or smallest member of a sample”, Proceedings of the Cambridge Philosophical Society 24, 180–190

Galambos, J., J. Lechner and E. Simiu (eds.), (1994), Extreme value theory and Applications. 3 v. Dordrecht (The Netherlands): Kluwer.

Ghosh, S., and S., Resnick, (2010), “ A discussion on mean excess plots”, Stochastic Processes and their Applications 120 (2010) 1492–1517

Gilleland, E., and R., W., Katz, (2016), “extRemes 2.0: An Extreme Value Analysis Package in R”, Journal of Statistical Software, August 2016, Volume 72, Issue 8.

Gilleland, E., and M. Ribatet & A.G. Stephenson, (2013), “A software review for extreme value analysis, Extremes” 16:103–119.

Gnedenko, B. V., (1943), “Sur la distribution limite du terme maximum d’une serie aleatoire”. Annals of Mathematics 44:423–53.

Gnedenko, B., and Kolmogorov. 1954. Limit distributions for sums of independent. Am. J. Math. 105.

Granger, C. W. J., and D. Orr, (1972), "Infinite Variance" and Research Strategy in Time Series Analysis, Journal of the American Statistical Association, Jun. 1972, Vol. 67, No. 338 (Jun. 1972), pp. 275-285

Gumbel, E. J., (1941), "The Return Period of Flood Flows", Ann, Math, Vol. 12, pp. 163-190

Gumbel, E., J., (1954), “Statistical Theory of Extreme Values and Some Practical Applications”, National Bureau of Standards Applied Mathematics Series 33, (U. S. Government Printing Office, Washington, D.C.)

Gumbel, E. J. (1958). Statistics of Extremes. New York: Columbia University Press.

Haan, L. de, and A. Ferreira, (2006), Extreme value theory. An Introduction. New York: Springer

Hill, B. M. (1975). “A simple general approach to inference about the tail of a distribution”. The Annals of Statistics 3:1163–l174.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F., (1985), “Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments”, Technometrics 27 (3), 251-261.

Hu, Y., and C., Scarrott, (2018), “evmix: An R package for Extreme Value Mixture Modeling, Threshold Estimation and Boundary Corrected Kernel Density Estimation”, Journal of Statistical Software, April 2018, Volume 84, Issue 5.

Jenkinson, A.F. (1955) “The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements.” Quart. J.R. Met. Soc., 81,158-171.

Karamata, J., (1933), “Sur un mode de croissance régulière. Théorèmes fondamentaux”, Bull. Soc. Math. France, 61:55–62.

Kim, Y. S., S. T. Rachev, M. L. Bianchi, I. Mitov, and F. J. Fabozzi, (2011), “Time series analysis for financial market meltdowns”, Journal of Banking & Finance 35:1879–1891.

Kinnison, Robert R., (1985), Applied Extreme Value Statistics, Macmillan Publishing Company, New York.

Kotz, S., and S. Nadarajah (2000). Extreme Value Distributions: Theory and Applications. London: Imperial College Press.

Leadbetter MR, Lindgren G. and Rootzen H., (1983), “Extremes and Related Properties of Random Sequences and Processes”, Springer-Verlag

Leadbetter, M., R., & Rootzen, H., (1988), “Extremal theory for stochastic processes”, Annals of Probability 16, 431-478

Leadbetter, M.R., (1991), “On a basis for Peaks over Threshold modeling”, Statist. Probab. Lett. 12, 357–362

Longin, F., M., (1999), “Optimal margin level in futures markets: extreme price movements”, The Journal of Futures Markets 19(2):127–152

Mandelbrot, B. (1963), “The Variation of Certain Speculative Prices”, Journal of Business, 36, 394-419

McNeil, A., (1998b), “Calculating Quantile Risk Measures for Financial Return Series using Extreme Value Theory”, unpublished manuscript, Departement Mathematik, ETH Zurich

McNeil, A. J., R. Frey and P. Embrechts (2005), Quantitative Risk Management, Princeton (New Jersey): Princeton University Press.

Nolan, P., J., (2020), Univariate Stable Distributions, Models for Heavy Tailed Data, Springer

Panaretos, J., and Z. Tsourti, (2003), “Extreme value index estimators and smoothing alternatives: A critical review”, MPRA Paper 6390, University Library of Munich, German

PICKANDS, J. (1975), “Statistical inference using extreme order statistics”, Annals of Statistics 3, 119-131.

Rachev, S. T., S. Mittnik and J. R., Kim, (1997), “Time Series with Unit Roots and Infinite-Variance Disturbances”, Appl. Math. Left. Vol. 11, No. 5, pp. 69-74, 1998

Resnick, S., (1987), Extreme Values, Regular Variation, and Point Processes, Springer Verlag

Samorodnitsky, G. and M. Grabchak, (2010), “Do financial returns have finite or infinite variance? a paradox and an explanation”, Quantitative Finance

Scarrott, C., and A., MacDonald, (2012), “A REVIEW OF EXTREME VALUE THRESHOLD ESTIMATION AND UNCERTAINTY QUANTIFICATION”, REVSTAT – Statistical Journal, Volume 10, Number 1, 33-60.

Sheikh, A. Z., and H. Qiao, (2009), “Non-normality of market returns: A framework for asset allocation decision making”, Whitepaper, J.P. Morgan Asset Management, JPMorgan Chase & Co.

Simkova, T., (2021), “Confidence intervals based on L-moments for quantiles of the GP and GEV distributions with application to market-opening asset prices data”, JOURNAL OF APPLIED STATISTICS, 2021, VOL. 48, NO. 7, 1199–1226

Smith, R. L., (1985), “Maximum likelihood estimation in a class of nonregular cases”, Biometrika 72 (1), 67–90.

Smith, R. L., (1987), ESTIMATING TAILS OF PROBABILITY DISTRIBUTIONS, The Annals of Statistics, 1987, Vol. 15, No. 3, 1174-1207

Tolikas, K., and R., Brown, (2006), “The Distribution of the Extreme Daily Share Returns in the Athens Stock Exchange”, The European Journal of Finance, Vol. 12, No. 1, 1-22, January 2006

Tsay, R. S., (2005), Analysis of Financial Time Series. Hoboken (New Jersey), Wiley-Interscience.

Vilasuso, J., and D. Katz (2000), “Estimates of the likelihood of extreme returns in international stock markets”, Journal of Applied Statistics 27(1):119–130.