SAPIENZA
UNIVERSITÀ DI ROMA

DOCTORAL SCHOOL OF STATISTICAL SCIENCES
PHD IN ACTUARIAL SCIENCE - XXIX CYCLE

Downside-risk averse investors and the use of Extreme Value Theory in optimal portfolio choice

Supervisor:
PROF. PAOLO ZAFFARONI

PhD Candidate:
JESSICA DONADIO
ID number: 1211674

A thesis submitted for the degree of
Doctor of Philosophy

27^{th} February 2017
Academic year A.A. 2015/2016

Sapienza Università di Roma

Downside-risk averse investors and the use of Extreme Value Theory in optimal portfolio choice

University of Rome "La Sapienza"
Faculty of Information Engineer, Informatics and Statistics
Department of Statistical Sciences

Doctoral School of Statistical Sciences
PhD Curriculum in Actuarial Science - XXIX Cycle

PhD Candidate
Jessica Donadio
ID number: 1211674

Supervisor:
Prof. Paolo Zaffaroni

Head of PhD School
Prof. Pier Luigi Conti

Head of PhD Curriculum
Prof. Fabio Grasso

Thesis defense on 27^{th} February 2017
Academic year A.A. 2015/2016

Abstract

The aim of this dissertation is to investigate the optimal portfolio selection problem for a risk-averse investor who wants to ameliorate the risk of extreme adverse events which cause unexpected large losses. Instead of referring to the modern portfolio theory (MPT), this thesis builds upon the foundations of Extreme Value Theory and focusses upon solving the optimisation problem minimizing quantile-based risk measures.

This choice is justified by the fact that the two main assumptions of the mean-variance approach, normally distributed returns and the use of variance as risk measure, could underestimate the extremes in presence of heavy-tailed distributions, consequently leading to an optimistic asset allocation. We take into account the potential heavy-tailedness in portfolio choice applying EVT for modelling the tails of the returns and focus on quantile-based risk measures (Value at risk and Expected shortfall), which are more appropriate for safety-first investors whose primary goal is to guard against the occurrence of unexpected large losses. We also believe that this approach better reflects the risk-based framework introduced by the new regulatory requirements for banking and insurance sectors.

The identified optimal portfolio is then compared with the one obtained under the classical MPT framework. Furthermore, an analysis to quantify and to understand how the extremal dependence between markets can impact the asset allocation problem is also performed. For this purpose two extremal dependence measures $\bar{\chi}$ and χ from Multivariate EVT are introduced. Intuitively, the fact that the securities can exhibit asymptotic independence or asymptotic dependence, means that tail diversification and a reduction of portfolio extreme risk can better be achieved by holding pairs of assets that are asymptotically independent.

The methodology proposed in this thesis is applied to the construction of portfolios of major international equity indices satisfying downside risk requirements of different tightness. Measures of extremal dependence are used to refine intuition about resulting portfolio allocations as well as separately analysing each chosen index (right and left) tails.

"Nothing great in the world has ever been accomplished without passion"
(Georg Wilhelm Friedrich Hegel)

To those that have consistently, truly believed in me.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Prof. Paolo Zaffaroni. Thank you for leading my first steps in the research world during my master studies, and then keep supporting me throughout the last three years. I constantly benefited from your guidance and encouragement all along my work. I am deeply grateful to you for allowing me to be your student in London, and introducing me to an extremely friendly and stimulating atmosphere both on a professional and personal level.

I would like to thank the Head of our Ph.D. School and all the Professors in our Department, especially in the Actuarial curriculum, for their precious scientific contributions and for giving me the opportunity to spend a long visiting period abroad, working on this very interesting theme. My whole stay has been a rewarding and most agreeable experience. I definitely enhanced my academic as well as personal knowledge.

During the last years I have been lucky enough to have had the possibility to meet many wonderful people with whom I truly enjoyed talking about my research. Thank you for all the fundamental insights provided. In particular, I would like to thank Alison Ahearn, unlimited source of precious ideas and advice, for always truly motivating and putting me into perspective. I could not have hoped for a better personal and professional support in a foreign college. Thank you to my colleague and friend Davide Benedetti, for always finding time to talk about my work with genuine enthusiasm and curiosity, and for all the Italian cooking we enjoyed. Thank you to Marco, Davide and Ivan, with whom I shared joy and pain over the last three years.

None of this would have been possible without the support of my friends outside the department. Thanks to Vittoria, Cecilia, Alessandra for always being there for me since high school. Thanks to Laura, Rachele and Roberta, for having been my friends throughout these years despite the distance.

Finally, I would like to thank my wonderful family for the support and unconditional love they have always given me. Thanks mum and dad for having accepted my decision to move abroad, and for having believed in me at all times. A special thanks to Dario, who has always been and always will be my strength.

Jessica Donadio

Rome, February 2017

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any that I have submitted, or is being concurrently submitted for a degree or diploma or other qualification at the University of Rome "La Sapienza" or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted for any such degree, diploma or other qualification at the University of Rome "La Sapienza" or any other University of similar institution.

Contents

List of Figures

List of Tables

Chapter 1

Motivation and background

The problem of portfolio optimisation has always been a topic of strong interest for investors and fund managers, and the need to create resilient portfolios, reducing the impact of extreme adverse events, is a growing concern given the financial crises suffered in recent decades. The new risk-based solvency requirements for the banking and insurance sectors further motivates the search for more appropriate methodologies to cope with rare events which cause heavy losses. This thesis presents a possible approach to tackling the problem of optimal portfolio choice: it is based on the theoretical foundations of Extreme Value Theory (see for example Embrechts et al., 1997; McNeil, 1999; McNeil et al., 2015) and the use of tail risk measures.

1.1 Introduction

In recent years, several crises have affected financial markets, causing significant global economic instabilities: for example the 1987 stock market crash was a major systemic shock, with the S&P 500 stock market index falling about 20 percent on a single day (October 19^{th}, known as Black Monday); then the Asian financial crisis triggered in July 1997, the hedge-fund crisis during 1997-1998 and the credit crisis beginning in 2007. This has led to numerous criticisms regarding the existing risk management systems and has also shown how inadequately the regulatory framework performed in times of crisis.
Thus, the problem of properly managing risks that includes forecasting in order to prevent negative events to severely impact against one's own investments or capital, has become a major concern for both financial institutions and regulators. In this context, the study of extreme events, like the above-mentioned episodes, clearly plays an important role and has received much attention in the last few decades being at the centre of investors' and risk managers' interest.

Extreme Value Theory[1] provides the solid mathematical and probabilistic fundamentals, essential in order to build statistical models describing such risky events.

1.2 Insurance industry: Risk management and CAT risks

Assessing the probability of rare events at the tails of the distributions, is an important issue not only in the risk management of financial portfolios. Indeed, EVT is well established and has been widely used in many fields of modern science: from engineering, hydrology, climate and weather, to computer science and telecommunications; moreover another natural application is in the insurance industry, where dealing with uncertainty and risks is the core business.

It is fundamental that the insurers comprehensively understand the risks of catastrophe, whether it is a man-made, a natural disaster or a pandemic[2], in order to better manage and consequently spreading these large-scale risks; this would bring an important benefit to the society as a whole as argued by Biffis (2013) as well. Comprehensive understanding means: first of all defining appropriate and advanced mathematical models to compute the risks of catastrophes, improving the ones currently in place, and EVT is an important tool to be considered; secondly, using the models as mysterious black boxes should be avoided if possible, being able to critically interpret their results.

This is not usually a trivial task for several reasons. By definition, large dataset on catastrophic losses are not available, consequently sophisticated estimation techniques are undermined by small sample sizes. In this situation, Bayesian techniques can help with reducing the statistical uncertainty about both parameter and capital estimates, in situations where observed data are insufficient to accurately estimate the tail behaviour of the loss distribution. However, this stream of literature does not fall within our current research scope.

A further complicating feature of extreme events, not promoting a deep and easy comprehension, is their complexity. Indeed, they are characterised by both complex dynamics and consequences, that systematically interweave throughout different insurance lines of business. Much attention has been focused on these aspects since the introduction of the new risk-based and market consistent regulatory framework known as *Solvency II* (SII).

[1]Henceforth often abbreviated as EVT.

[2]Later, we will examine which are the main perils considered in the modelling of catastrophe risk within the Solvency II framework.

1.2.1 Solvency II overview

"Solvency II is not just about Capital. It is a change of Behaviour."
(Thomas Steffen, Former Chairman of CEIOPS)

Solvency II is an EU legislative programme, implemented in all 28 Member States, including the UK, on 1^{st} January 2016. It introduces for the first time a harmonised, sound, robust, EU-wide insurance regulatory regime. The Solvency II Directive was adopted in November 2009, and amended by Directive 2014/51/EU of the European Parliament and of the Council of 16 April 2014 (the so-called Omnibus II Directive). The legislation replaces 14 EU insurance and reinsurance directives, commonly known as "Solvency I". Throughout its 40 years of existence, the "Solvency I" regime showed structural weaknesses, as already pointed out before.
It was not risk-sensitive, and a number of key risks, including market, credit and operational risks were either not captured at all in capital requirements or were not properly taken into account. This lack of risk sensitivity had the following consequences:

- Risks: Owing to its simplistic model, Solvency I did not lead to an accurate assessment of each risk an insurer were exposed to;
- Supervision: It did not ensure accurate and timely intervention by supervisors;
- Capital allocation: It did not entail an optimal allocation of capital, i.e. an allocation which is efficient in terms of risk and return for shareholders.

Example 1.1. Consider the Italian insurance market as an example. The capital requirement, known as solvency margin, for the life insurance business was simply determined as the sum of a percentage (a) of the mathematical reserves[3] plus a percentage (b) of the capital sum at risk (only when positive)[4], being each component conveniently amended to take into account reinsurance cessions. The formula changes based on the investment risk, mortality and expense risk affecting a particular type of product: for the traditional life insurance business a=4% and b=0.3% (except for term insurance where a is reduced to 0.1% when the contract duration is less than 3 years or to 0.15% when the contract length is less than 5 years); for linked life insurance business where the company entirely bears an investment risk a=4%, whereas when the company bears no investment risk must be either 1% if the contract duration is greater than 5 years or 0% otherwise (b always equals 0.3%). Without going too much into details, in the non-life insurance instead, the solvency margin was calculated taking the greater between a proportion of

[3] The mathematical reserves are defined as the provision made by an insurer to cover liabilities (excluding liabilities which have fallen due) arising under or in connection with life insurance business. In the calculation, no negative value for the mathematical reserve can be used under any policy.

[4] The capital sum at risk is defined as the benefit amounts payable as a consequence of the happening of the contingency covered by the policy contract, less the mathematical reserve in respect of the relevant contract. This value is positive only for life insurance policies including the death benefit. The percentage considered is b=0.30% of the gross capital sum at risk before deduction for reinsurance cessions.

written premiums and a proportion of the average claim paid over the previous 3 fiscal years (7 years for hail insurance), always conveniently considering the reinsurance cessions. Clearly in this setting: the P&C[5] risk capital calculation, took into account only the insurance sale or claims paid for the company as a whole, not differentiating between lines of business; moreover the rates applied were not related to the type of product and did not consequently reflect the actual riskiness of the business[6]; moreover, no market risk were considered; the life risk capital to be hold, being proportional to the mathematical reserves and to the capital sum at risk, was higher for larger companies, whereas it is well known that the size is an advantage in terms of ability of diversifying the business and holding more stable portfolios[7].

The Solvency II framework, like the Basel framework for banks, proposes to remedy these shortcomings. The new standards have been set up aiming at ensuring the safety and solvency of insurance companies. We can summarise the key objectives of Solvency II as follows:

- Greater risk awareness: Risk is central in both governance and operations;
- Improved consumer protections: It will ensure a uniform and enhanced level of policyholder protection across the EU. A more robust system will give policyholders greater confidence in the products of insurers;
- Modernised supervision: The "Supervisory Review Process" will shift supervisors' focus from compliance monitoring and capital to evaluating insurers' risk profiles and the quality of their risk management and governance systems;
- Deepened EU market integration: Through the harmonisation of supervisory regimes;
- Increased international competitiveness of EU insurers.

The introduction of the new solvency regime brought a big change in the everyday operating and mindset of insurers: important investments have been made in order to enhance the way of identification, measuring, reporting, managing and monitoring risks.

In general, we can talk about revolutionary change in the "risk management framework" or "risk management system". Such important investments in firms' IT systems, data quality[8], and methodologies to measure/aggregate Risk-Adjusted Capital clearly led

[5] Property and Casualty insurance.

[6] They were also based on studies referred to an insurance market scenario that has clearly become obsolete by now.

[7] Large companies are managed by teams of professionals in different fields, each one in charge of a specific area and presumably they have been well selected, despite the wide range of activities. Portfolio stability in terms of clients turnover as they more easily build customer loyalty; this means having the possibility to rely upon claim history per policyholder and so better managing the risks on an individual basis.

[8] The amount of data needed for Solvency II calculations is much greater than under the current regime. At a basic level, more data requires more validation and could increase the chance of data errors. Processes for collecting, checking and storing data will need to be more robust.

insurance business to become more expensive, regulated and volatile than it used to be. Market consistency increased firms' balance sheet volatility, depending on market conditions, thus reducing the predictability of profitability, dividends and, ultimately, shareholders' returns. Therefore, Solvency II is not just about capital. It is a comprehensive programme of regulatory requirements for insurers, covering authorisation, corporate governance, supervisory reporting, public disclosure and risk assessment and management, as well as solvency and reserving.
The SII programme is divided into three areas, known as pillars:

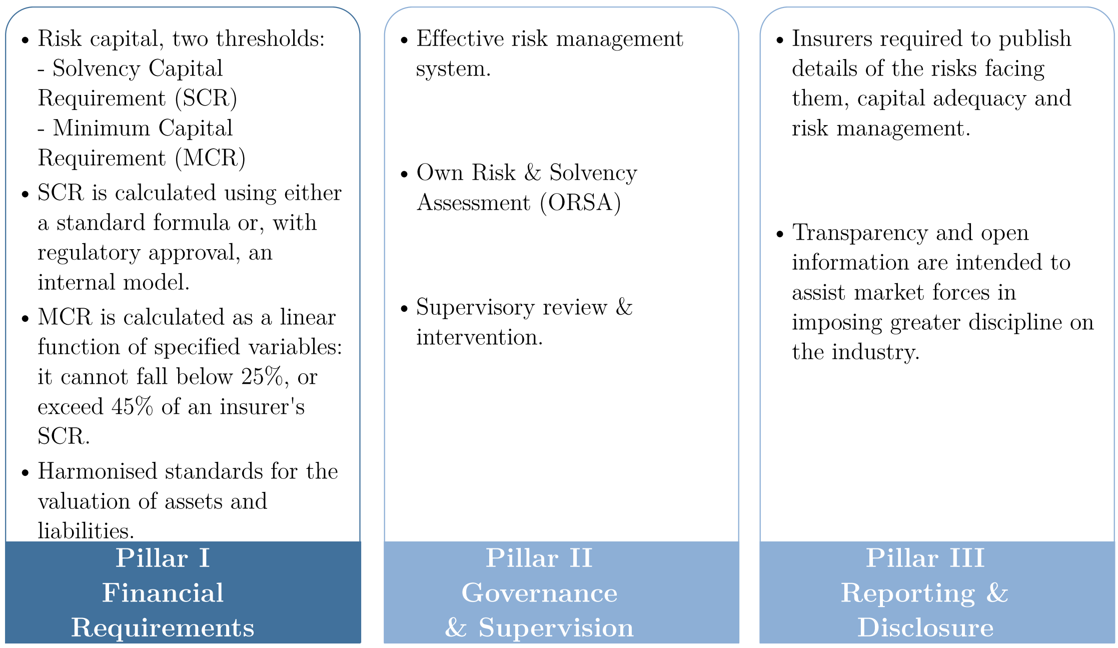


Figure 1.1: Solvency II pillars.

Pillar 1 sets out quantitative requirements, including the rules to value assets and liabilities (in particular, technical provisions), to calculate capital requirements and to identify eligible own funds to cover those requirements;

Pillar 2 sets out requirements for risk management, governance, as well as the details of the supervisory process with competent authorities; this will ensure that the regulatory framework is combined with each undertaking's own risk-management system and informs business decisions. An insurer must undertake an Own Risk and Solvency Assessment (ORSA), i.e. an internal risk assessment process that aims to ensure senior management have conducted a review of risks and that the insurer holds sufficient capital against those risks;

Pillar 3 addresses transparency, reporting to supervisory authorities and disclosure to the public, thereby enhancing market discipline and increasing comparability, leading to more competition.

Capital requirements under Solvency II are forward-looking and economic, i.e. they are tailored to the specific risks borne by each insurer, allowing an optimal allocation of capital across the EU. They are defined along a two-step ladder, including the solvency capital requirements (SCR) and the minimum capital requirements (MCR), in order to trigger proportionate and timely supervisory intervention.

1.2.2 Solvency Capital Requirement (SCR) calculation and catastrophe risks

The Solvency Capital Requirement (SCR) should reflect a level of eligible own funds that enables insurance and reinsurance undertakings to absorb significant losses and that gives reasonable assurance to policyholders and beneficiaries that payments will be made as they fall due. The SCR, according to the new regulatory regime, is determined as the economic capital to be held by insurance and reinsurance undertakings in order to ensure that ruin occurs no more often than once every 200 years[9]. It can be calculated using either a *standard formula*; a bespoke *internal model* that has been approved by the insurer's supervisor; or a *mixture* of both (partial internal model - IM -). The SCR standard formula, as the name suggests, is not "firm" specific and employs a formulaic factor based calculation[10]. It aims at achieving the right balance between risk-sensitivity and practicality. Moreover, both the use of undertaking-specific parameters (USPs), where appropriate, and standardised simplifications for small and medium-sized enterprises (SMEs) are allowed. The partial or full internal model, specific to individual firm and risk profile, requires the Regulator pre-approval and must be widely used and play an important part in firm's system of governance. A simulation-based IM aims to produce a distribution of the basic own funds over one year and then calculate the SCR from that.

The standard formula, characterised by a "modular" architecture, calculates capital requirements by risk type and then aggregate them with a diversification adjustment and using a correlation matrix provided by the supervisor. The modular structure is showed in figure 1.2. It is important to clarify that both SCR calculation methods (standard formula and internal model - whether full or partial -) take at least six specified risk modules into account, those highlighted in the red boxes.

The European Parliament defines *catastrophe risk* as "the risk of loss, or of adverse change in the value of insurance liabilities, resulting from significant uncertainty of pricing and provisioning assumptions related to extreme or exceptional events".

[9]The SCR shall correspond to the Value-at-Risk (VaR) of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of 99.5% over a one-year period. Essentially, the basic own funds are defined as the excess of assets over liabilities, under specific valuation rules (in particular market-consistency, meaning that for insurance liabilities, requires a best estimate - defined as the expected present value of future cash flows under Solvency II - plus a risk margin calculated using a cost-of-capital approach).

[10]The calculation is calibrated to seek to ensure a 99.5% confidence level over a 1 year period which is said to equate to a BBB rating.

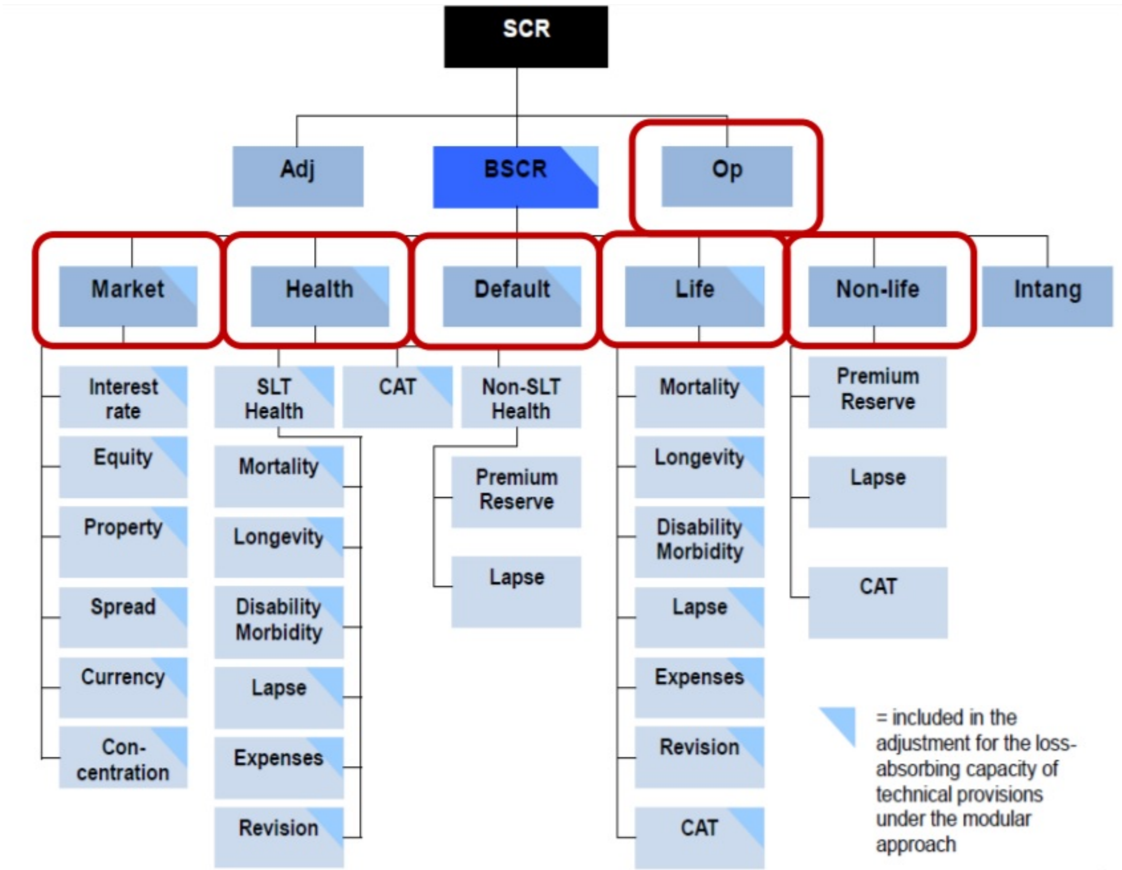


Figure 1.2: Standard formula modular structure.

As we can see in the image 1.2, the catastrophe risk is included in the life, non-life and health modules of the Standard formula. The life catastrophe risk sub-module captures the risk stemming from extreme death events that are not sufficiently captured by the mortality risk sub-module (e.g. a contagious disease process or a pandemic can affect many persons simultaneously, nullifying the usual assumption of independence among persons.). The non-life catastrophe risk sub-module is essentially split into three separate and independent sub-sub-modules that cover catastrophe risk related to natural perils (e.g. earthquake, flood, hail, subsidence, and wind storm -winter storm-), risk related to man-made events (e.g. fire, explosion and acts of terror) and other catastrophe events. Lastly, the health catastrophe risk sub-module is based on three standardized scenarios (mass accident, accident concentration and pandemic) to apply across all countries for medical expense plus accident and sickness products.

Considering the central role played by risk management and assessment, the extreme events and catastrophe risk within Solvency II, we believe that the methodology and tools presented in this thesis are versatile, naturally reflecting the mindset introduced with the new regulatory framework. These are essential for modelling (for instance in case an internal model approach is chosen) or for analyses purposes in order to report to the management on a continuous basis.

1.3 Portfolio selection

One of the classic questions of finance-portfolio theory is the following: how can we identify the optimal portfolio given a set of available assets? This is obviously an interesting problem, to investors in general (individuals, financial institutions and so on) as well as to academics. The concepts of portfolio optimisation and diversification have been instrumental in the development and understanding of financial markets and financial decision making.

The optimal portfolio choice problem was firstly addressed by Markowitz (1952) in his landmark paper, "Portfolio selection", that appeared in the March issue of the *Journal of finance*, in which he defined a mean-variance model, assuming normally distributed returns and using variance as a risk measure.

However, it is now almost universally accepted that the distribution of financial assets return is actually skewed and fat-tailed (following for example research by Mandelbrot, 1963). Therefore, although the assumption of normality makes calculations easier, it underweights extreme events, and might lead to an optimistic asset allocation. For this reason many studies were conducted exploring the non-normality of assets returns and suggesting alternative approaches.

1.3.1 Modern portfolio theory

When Markovitz published his paper on portfolio selection in 1952 he provided the foundation for modern portfolio theory (MPT) as a mathematical problem, being awarded the Alfred Nobel Memorial Prize in Economic Sciences in 1990.

The specific problem addressed by Markowitz was how to construct an efficient portfolio from a collection of risky assets. First, he quantified return and risk of a security, using the statistical measures of its expected return and standard deviation. Second, he suggested that investors should consider return and risk together, and determine the allocation of funds among investment alternatives on the basis of their return-risk trade-off. The idea was revolutionary at that time. First of all because it introduced the important principle of portfolio diversification: a portfolio's riskiness depends on the correlations of its constituents, not only on the average riskiness of its separate securities. Through diversification, risk can be reduced (but not generally eliminated) without changing the expected portfolio return. Secondly, it formulated the financial decision-making process as an optimisation problem: the so-called mean-variance optimisation (MVO) suggests that among the infinite number of portfolios that achieve a particular return objective, the investor should choose the portfolio that has the smallest variance. All other portfolios are "inefficient" because they have a higher variance and, therefore, higher risk[11].

Our intention in this section and Chapter 2, where we present the methodology in detail, is not to provide a complete survey of MVO, its extensions and related areas, but to give a brief overview of the problem, useful to understand the motivation behind our approach and which drawbacks in the Markowitz model we are able to overcome within our framework. For complete surveys please refer to Steinbach (2001), Rubinstein (2002), Fabozzi et al. (2007), and Markowitz (2014).

1.3.2 MPT's underlying assumptions

An important aspect of pareto-optimal (efficient) portfolios is that each determines a Von Neumann-Morgenstern utility function[12] (Von Neumann and Morgenstern, 1953) for which it maximizes the expected utility of the return on investment.
This allowed Markowitz (1959) to interpret his approach by the theory of rational behaviour under uncertainty: in chaps. 6 and 13, he assumes that the investor should act to maximize expected utility, and proposes mean-variance analysis as an approximation to the maximization of expected utility. Tobin (1958) notes that the mean variance framework is a special case of so-called expected utility maximization where investors are assumed to have quadratic utility or returns distributions are jointly normal.

a) Normally distributed returns.

As already mentioned, the seminal work of Markowitz made explicit the trade-off of risk and reward[13] in the context of a portfolio of financial assets and MPT argues that investors can benefit from diversification by investing in financial assets with lower correlations. Assuming that the financial returns are subject to joint normal distribution, the dependence between financial returns is fully described by the linear correlation coefficient and efficient portfolios are obtained via the traditional MVO program (we refer the reader to Section 2.2 of Chapter 2 for details on the model).

However, empirical studies find that the assumption of normality of financial returns

[11] MVO is used both for constructing portfolios of individual assets (asset level) and for asset allocation (asset class level).

[12] For introductions to utility theory see for example Ingersoll (1987) or Huang and Litzenberger (1988).

[13] Financial theory has long recognized the interaction of risk and reward. Others such as Sharpe (1964), Lintner (1965), and Ross (1976), have used equilibrium arguments to develop asset pricing models such as the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), relating the expected return of an asset to other risk factors. Specifically, risk based asset pricing models explain the expected excess return in terms of factors and the size of risk, measured by the covariance between the individual return and the factor return. A common theme of these models is the assumption of normally distributed returns. Even the classic Black and Scholes option pricing theory (Black and Scholes, 1973) assumes that the return distribution of the underlying asset is normal.

distribution is not hold, exhibiting fat tail and asymmetry that cannot be described by their mean-variances alone. Mandelbrot (1963) strongly rejected normality as a distributional model for asset returns, conjecturing that financial return processes behave like non-Gaussian stable processes. The latter are commonly referred to as “stable Paretian” distributions or “Levy stable” distributions[14]. While there have been several studies in the 1960s that have extended Mandelbrot’s investigation of financial return processes, the most notable is Fama (1963) and Fama (1965). Fama’s and others work led to a consolidation of the stable Paretian hypothesis. In the 1970s, however, closer empirical examination of the “stability” of fitted stable Paretian distributions also produced evidence that was not consistent with the stable Paretian hypothesis. Specifically, it was often reported that fitted characteristic exponents (or tail-indices) did not remain constant under temporal aggregation. Partly in response to these empirical “inconsistencies” various alternatives to the stable law were proposed in the literature, including the fat-tailed distributions in the domain of attraction of a stable Paretian law, finite mixtures of normal distributions, the Student t-distribution, and the hyperbolic distribution.

Stylized Facts. Although statistical properties of prices of stocks and commodities and market indexes have been studied using data from various markets and instruments for more than half a century, the availability of large data sets of high-frequency price series and the application of computer-intensive methods for analysing their properties have opened new horizons to researchers in empirical finance in the last decade and have contributed to the consolidation of a data-based approach in financial modelling.

The study of these new data sets has led to the settlement of some old disputes regarding the nature of the data: if one examines the properties of financial time series from a statistical point of view, the seemingly random variations of asset prices do share some quite non-trivial statistical properties.

Such properties, common across many instruments, markets and time periods, has been observed by independent studies and classified as *“stylized empirical facts”* (Cont, 2001); being obtained by taking a common denominator among the properties observed in studies of different markets and instruments, one gains in generality but, at the same time, tends to lose in precision of the statements that can be made about asset returns. Indeed, stylized facts are usually formulated in terms of qualitative properties[15]:

[14]Stable Paretian is used to emphasize that the tails of the non-Gaussian stable density have Pareto power-type decay. “Levy stable” is used in recognition of the seminal work of Paul Levy’s introduction and characterization of the class of non-Gaussian stable laws.

[15]We list only the stylized facts that are more related to our purpose: point out the insufficiency of the normal distribution for modelling the marginal distribution of asset returns and their heavy-tailed and asymmetric character. Other properties common to a wide set of financial asset are: Intermittency, Slow decay of autocorrelation in absolute or squared returns, Leverage effect, Volume/volatility correlation, Asymmetry in time scales (see Cont, 2001).

- Absence of autocorrelations: (linear) autocorrelations of asset returns are often insignificant, except for very small intraday time scales ($\simeq$ 20 minutes) for which microstructure effects come into play;
- Heavy tails: the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular this excludes stable laws with infinite variance and the normal distribution. However the precise form of the tails is difficult to determine;
- Gain/loss asymmetry: one observes large drawdowns in stock prices and stock index values but not equally large upward movements[16];
- Aggregational Gaussianity: as one increases the time scale Δ_t over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales;
- Volatility clustering: different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high-volatility events tend to cluster in time;
- Conditional heavy tails: even after correcting returns for volatility clustering (e.g. via GARCH-type models), the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution.

In many instances, the tails of the return distribution significantly affect portfolio performance (see, for example, Jobst and Zenios, 2001). Harvey and Siddique (2000) show that skew in stock returns can be relevant in portfolio selection. Others studies find evidence of time-varying skewness (Harvey and Siddique, 1999) and time-varying kurtosis (Jondeau and Rockinger, 2003; Brooks et al., 2005). According to the statistical properties characterising financial returns, as presented before, the departures from normality have been examined using the skewness (the third moment of distribution) and the kurtosis (the fourth moment of distribution).

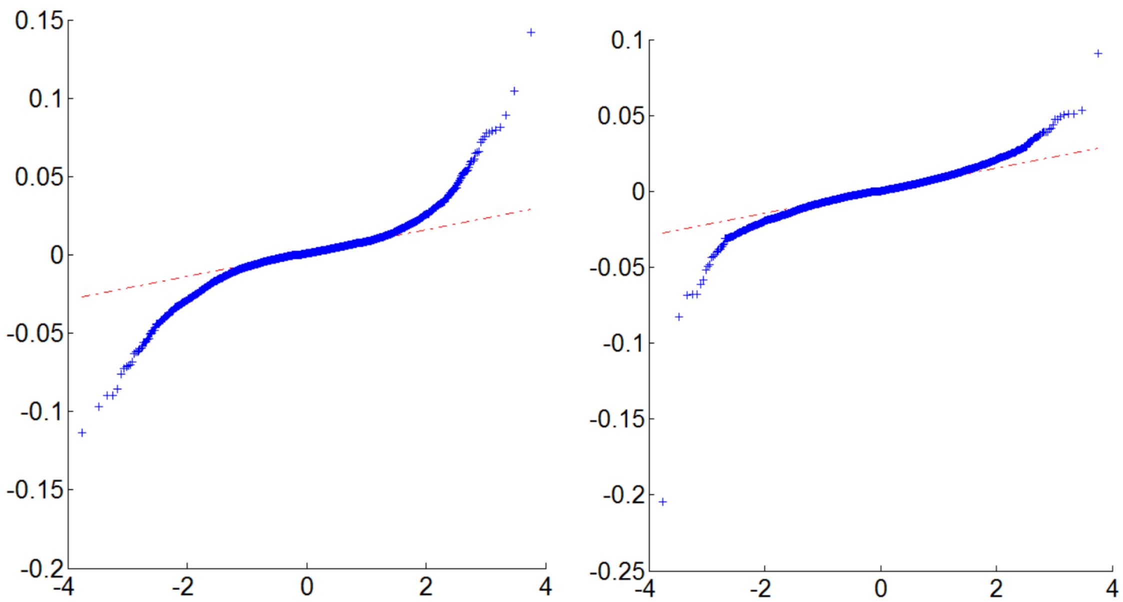
One can summarize the empirical results by saying that the distribution of financial returns tends to be asymmetric (usually negatively skewed), sharp peaked and heavy tailed. Specifically, negative skewness indicates a higher probability of negative returns, that is, the market gives higher probability to decreases than increases in asset pricing; the property of excess kurtosis of financial returns, implies that extreme market movements, in either directions, will occur with greater frequency in practice than would be predicted by the normal distribution.

For example, a 5% daily loss is observed to occur in equity markets approximately once every two years, while if returns were normally distributed, such a change would be

[16] This property is not true for exchange rates where there is a higher symmetry in up/down moves.

expected only once in every one thousand years, given the estimated return variances (Johansen and Sornette, 1999).

Figures 1.3 and 1.4 illustrate this concepts for the S&P500 and NASDAQ indices. Both the collection of images clearly show, via the quantile-quantile (QQ) plot[17] and the plot of the empirical histogram with the normal density function, that the distribution tails of the S&P500 and NASDAQ are heavier than the tails of the normal distribution. Moreover, in 1.4a and 1.4b the distributions are distinctly Leptokurtic[18].



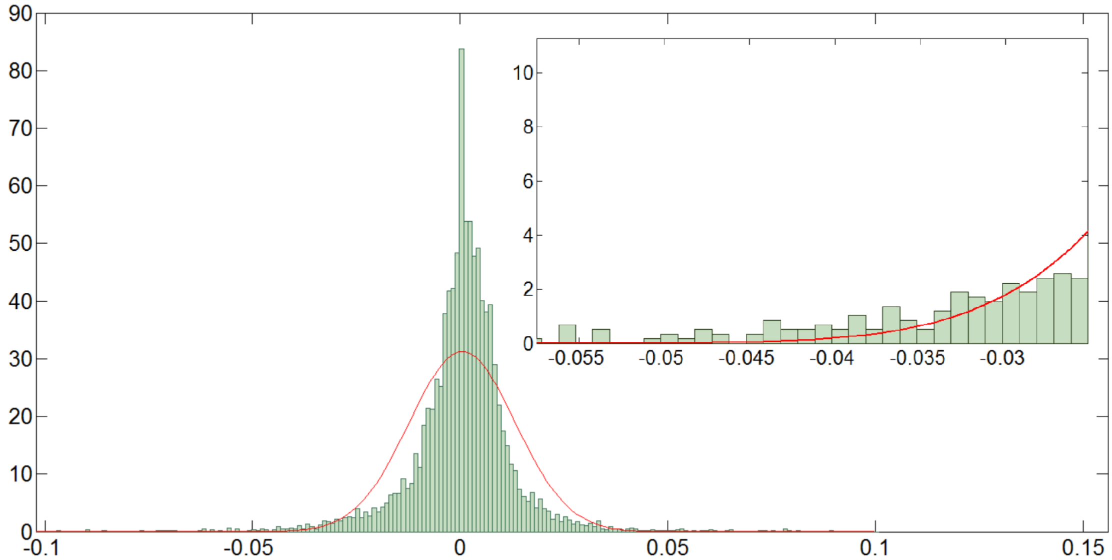
(a) NASDAQ

(b) S&P500

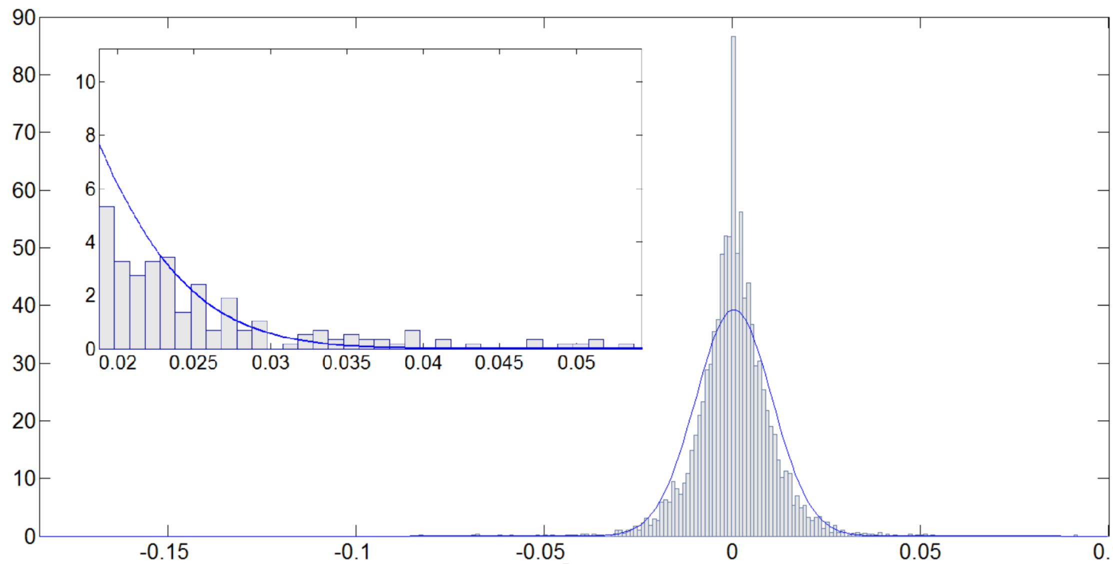
Figure 1.3: Quantile-quantile (QQ) plots for both the equity index returns: NASDAQ in (a) and S&P500 in (b). The figures display a quantile-quantile plot of the sample quantiles (abscissa) versus theoretical quantiles from a normal distribution (ordinate).

[17]A QQ plot is a graphical check to verify if two distributions are of the same type. Two random variables X and Y are said to be of the same type if their distributions are the same up to a change in location and scale, that is $X \stackrel{\text{def}}{=} aY + b$ for some a $a \in \mathbb{R}^+$, $b \in \mathbb{R}$. Since the QQ plot plots quantiles of two distributions, if they are of the same type, the plot should be linear. In this case we are checking whether the empirical distribution of the two equity index returns and the hypothesized normal distribution are of the same type.

[18]The distribution of the financial return series are characterized not only by heavy tails, but also by a high peakedness at the center.



(a) NASDAQ



(b) S&P500

Figure 1.4: Empirical distribution of the S&P500 and NASDAQ index returns both versus the normal density function (red line in (a) and blue line in (b)), plus a focus on the left tail in (a) and a focus on the right tail in (b).

	Min	**Max**	**Mean**	**St Dev**	**Skewness**	**Kurtosis**
S&P 500	-20.47	9.10	0.045	1.02	-1.51	33.22
NASDAQ	-11.35	14.17	0.050	1.28	-0.10	11.88

Table 1.1: Descriptive statistics for simple daily returns (as %) of the financial assets: the S&P 500 index and the NASDAQ composite index. Price indices on the period: 2/1/1980-14/5/2002.

The time series considered starts on the 2^{nd} January 1980 and ends on 14^{th} May 2002 (data source: *Datastream*) and the corresponding empirical statistics of the two equity index returns can be found in Table 1.1. We refer to the definition of kurtosis (often referred to as "excess kurtosis") that implies a kurtosis of zero for the standard normal distribution. Therefore, positive kurtosis indicates a "heavy-tailed" distribution and negative kurtosis indicates a "light tailed" distribution.

Consequently, the normality assumption for the financial returns is no longer suitable and can largely lead to misleading results. The implication that returns of financial assets have a heavy-tailed distribution may be profound to a risk manager in a financial institution who still relies on the classical normality assumption for the modelling of financial returns distribution.

Indeed, in presence of fat-tailed asset returns, this assumption will lead to a systematic underestimate of the true riskiness of a portfolio, where risk is measured as the likelihood of achieving a loss greater than some threshold. Therefore, it is clearly vital to specify the distribution of asset returns for risk management, financial forecasting and optimal asset allocation purposes as we have seen that a failure may lead to biased results and, consequently, to wrong decisions.

The identification of the most suitable and accurate distribution to fit financial returns data is still topic of ongoing research. There are dozens of parametric models proposed in the literature, but from the empirical features described above, one can conclude that, in order for a parametric model to successfully reproduce all the above properties of the marginal distributions it must have at least four parameters: a location parameter, a scale (volatility) parameter, a parameter describing the decay of the tails and eventually an asymmetry parameter allowing the left and right tails to have different behaviours.

Student t distributions have been used to explain the leptokurtic behaviour but have failed to explain the time varying volatility and the skewness (Kon, 1984; Blattberg and Gonedes, 1974). Among the well-known candidates we can also find the generalized hyperbolic distributions (Prause, 1999; Eberlein et al., 1998; Bibby and Srensen, 2003), and the stable Paretian distributions (Rachev and Mittnik, 2000; Rachev et al., 1998). The Autoregressive Conditional Heteroskedastic (ARCH) model proposed by Engle (1982) provides a simple framework to explain the time varying (conditional) volatility. Afterwards, Bollerslev (1986) proposed a GARCH model, which is an extension to the ARCH model. Many other extensions of the ARCH model have been suggested to explain other features such as leverage effect. Even though these models typically assume a normal distribution for the innovation, they are able to generate data with unconditionally fat tails, heavier than what an i.i.d. normal distribution would suggest. The kurtosis implied by these models using a conditional normal distribution, however, does not match the large kurtosis found in the return series. The observation that GARCH models with normal disturbances cannot generate sufficient leptokurtosis to replicate that observed in actual data was in part the motivation for the study of Bollerslev (1987). He proposed a more general model that allowed the innovation to have a standardized t-distribution, so that extreme values, occurring more often than under the normality assumption, may be accommodated. This model has the appeal of a finite variance with fat tails.

Recently, it has been proposed that the classes of generalized extreme value distributions and generalized Pareto distributions (see for example Diebold et al., 2000; Bali, 2003; Rocco, 2014) provide a more robust modelling of financial returns distribution. We refer to the latter stream of literature to provide the foundations of our project. For a detailed overview of the related work, please see Section 2.1 in Chapter 2.

b) Portfolio Riskiness: variance as proxy for risk.

The second central moment of returns, has been the subject of a large literature in finance: from risk based asset pricing models, explaining the expected excess return in terms of factors and the size of risk, to derivative pricing, for example (Black and Scholes, 1973) introduced the well known formula for option pricing, where the volatility is the key input variable.

As described earlier, variance of returns has been widely used as a proxy for risk in financial returns: in particular, it is the key risk measure used in the mean variance approach to identify the optimal portfolio an investor would hold according to his risk aversion. However, it is well known that variance has been criticized because it is a symmetric measure and treats *downside risk* and *upside risk* in the same way, while the investors generally dislike deviations below the mean and like deviations above the mean (Post and Vliet, 2004).

Downside risk is the financial risk of experiencing unexpected losses associated with an investment project and the magnitude of these losses. Risk includes the chance that a security increases or decreases in value unexpectedly. However, most people are concerned with the possibility of an unexpected decline - which is indeed known as downside risk.

As early as Roy (1952), who assumed that an investor's goal was to minimize his/her risk, economists have recognized that investors care differently about downside losses than they care about upside gains. Recently, Ang et al. (2006) in their article show that the cross-section of stock returns reflects a premium for bearing downside risk. Investors who are sensitive to downside losses, relative to upside gains, require a premium for holding assets that covary strongly with the market when the market declines. Hence, in an economy with agents placing greater emphasis on downside risk than upside gains, assets with high sensitivities to downside market movements have high average returns. The general, and quite intuitive, idea behind downside risk is that the left hand side of a return distribution involves risk while the right hand side contains the better investment opportunities.

> *"It was understood that risk relates to an unfortunate event occurring, so for an investment this corresponds to a low, or even negative, return. Thus getting returns in the lower tail of the return distribution constitutes this "downside risk". However, it is not easy to get a simple measure of this risk."*
>
> (Quoted from Granger, 2009)

In addition, the variance measures the spread of the distribution around the mean and so, by definition, it concerns the central part of the distribution rather than focussing on the more extreme values that represent the financial risk hidden in the tails (in particular we care more about the left one). Clearly, considering all the evidence in support of the importance of including the modelling of extremal events we discussed earlier, this is not a desirable feature.

All the arguments presented are powerful reasons for replacing variance. Therefore, some authors introduced several downside risk measures in their optimization program. First Markowitz (1959) advocates substituting the variance with *semi-variance* instead, that is a measure of the dispersion of the data that fall below the mean (or a target value) always calculated as an average of the squared deviation of values lower than the mean[19]. Other examples are Roy (1952) and Arzac and Bawa (1977) who consider Value at Risk and refer to the so called *safety first principle*[20].

Incorporating skewness, heavy-tailedness and tail-risk measures.

We rely on the foundation introduced by the safety first criterion to develop our model. Moreover, following the reasoning presented in the whole section, we go beyond the clas-

[19] In order to put the accent on returns below the mean, semi-variance is defined as

$$\tilde{\sigma_X} = \mathbb{E}[(X - \mathbb{E}X)\mathbb{1}_{[X<\mathbb{E}X]}]^2$$

[20] Please, refer to Section 2.1 in Chapter 2 for further details.

sical MVO and its assumptions using extreme value theory to reproduce the fact that the returns are heavy-tailed (instead of assuming normality) and employing quantile-based risk measures (instead of the "symmetric" variance).
The methodology implemented will allow us to embody the risk preferences of an investor representing the asymmetric treatment of upside and downside risk. This behavioural profile has been observed in several studies. Furthermore, the application of tail risk measures will reflect the impelling need of properly managing and considering also the extreme events. This characterises risk managers' mindset within the new solvency regime, thinking beyond what is expected on average.

1.4 Measuring Extreme Risks

The main task of risk management is to try to assess and quantify risks. There is a considerable interest, both from a theoretical and a practical point of view, in a quantitative assessment of the risk involved in a financial position. Ideally, the best and most informative risk measure for the returns of some financial security would be looking at the whole tail of their distribution. However, the return distribution generally is:

(a) unknown: what we observe in the market are the return time series and so the realisations of the random variable that produced those values; however, we do not know the exact distribution the data were drawn from;
(b) difficult to handle: considering the whole distribution contains, in a certain sense, too much information. A tractable risk measure should be a synthesis of the data that characterise the precise distribution.

If the financial position is described by the resulting discounted net worth at the end of a given period $X(\omega)$ when the scenario $\omega \in \Omega$ is realised, where $X : \Omega \to \mathbb{R}$ is a real-valued function on some set Ω of possible scenarios, then a quantitative measure of risk is given by a mapping ρ from a certain space $\mathcal{X}$ of functions $X \in \mathcal{X}$ on Ω to the real line. The discounted net worth corresponds to the profit and losses of the position (P&L).

Definition 1.1 (Risk measure). Let $(\Omega; \mathcal{F}; \mathbb{P})$ be an atomless probability space. Throughout, all random variables are defined on $(\Omega; \mathcal{F}; \mathbb{P})$ and all probability measures are defined on $(\Omega; \mathcal{F})$. Let $\mathcal{L}^0$ be the space of all random variables, $\mathcal{L}^1 = \{X \in \mathcal{L}^0 \mid \|X\|_1 = E(|X|) < +\infty\}$ the subspace of all integrable random variables and $\mathcal{L}^\infty$ the set of all (essentially) bounded random variables[21].
A risk measure ρ is a real-valued function that associates to a risk in a set $\mathcal{X}$ a real number $\rho : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subseteq \mathcal{L}^0$ [22].

[21] For further details please refer to Embrechts et al. (2015b); Pesenti et al. (2016) and to Chapter 6 of McNeil et al. (2005).
[22] $\mathcal{X}$ typically contains $\mathcal{L}^\infty$, and is closed under addition and positive scaler multiplication.

Throughout this thesis we assume that the argument of ρ represents a financial loss. We list below some risk measure's properties for $X, Y \in \mathcal{X}$:

(i) Monotonicity: $\rho(X) \leq \rho(Y)$ if $X \leq Y$;
(ii) Translation-invariance: $\rho(X + c) = \rho(X) + c$, $\forall c \in \mathbb{R}$;
(iii) Positive homogeneity: $\rho(\lambda X) = \lambda\rho(X)$, $\forall \lambda > 0$;
(iv) Subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$;
(v) Convexity: $\rho((1 - \omega)X + \omega Y) \leq (1 - \omega)\rho(X) + \omega\rho(Y), \omega \in [0, 1]$;
(vi) Law-invariance: $\rho(X) = \rho(Y)$ if X and Y have the same distribution $F_X = F_Y$;

Definition 1.2 (Monetary Risk measure)**.** A *monetary risk measure* is a risk measure satisfying (i) and (ii).

In order to tackle the second point (b), several risk measures have been adopted during the last decades. First of all the variance (or alternatively standard deviation) that we have widely reviewed in the previous sections.
Later, in the early 1990s, a number of financial institutions (J.P. Morgan, Bankers Trust,...) proposed a new risk measure to quantify by a single number the firms aggregate exposure to market risk. This is commonly known as *value at risk* (VaR).

1.4.1 Value at Risk

The value at risk has become the benchmark risk measure in today's financial world being used not only for market risk calculation but, for example, also for credit, operational, liquidity, and legal risk estimate.

VaR (at the confidence level α) is defined as the loss of a financial position over a fixed time horizon Δ that would be exceeded with small probability $(1 - \alpha)$.

Definition 1.3 (Value-at-Risk)**.** VaR at level $\alpha \in (0, 1)$ of a loss variable L, $\mathrm{VaR}_\alpha : \mathcal{L}^0 \to \mathbb{R}$, is defined as the α-quantile of the loss distribution:

$$\mathrm{VaR}_\alpha(L) = q_\alpha(L) = inf\{l \in \mathbb{R} : P(L \leq l) \geq \alpha\},\ L \in \mathcal{L}^0. \tag{1.1}$$

Looking at the rigorous mathematical definition, we can intuitively understand that the VaR, from a statistical point of view, is nothing more than a quantile (usually a high quantile of level α, typically above the 95^{th} percentile) of the *loss distribution* ($L_{[t;t+\Delta]}$[23]) related to the financial position held.

Practitioners in risk management are often concerned with the so called *profit and loss* (P&L) *distribution*, that is the distribution of the change in value of the financial position over a given holding period (i.e. $-L_{[t;t+\Delta]}$ the loss random variable). However, in

[23] We denote the value of a portfolio at time t by $V(t)$ and assume that the rv $V(t)$ is observable at time t. For a given time horizon Δ, such as 1 or 10 days, the loss of the portfolio over the period $[t, t + \Delta]$ is given by $L_{[t;t+\Delta]} = -[V(t + \Delta) - V(t)]$.

risk management we are mainly concerned with the probability of observing large losses and hence with the upper tail of the loss distribution. Thus, we often drop the P from P&L, both in notation and language, and we refer to the loss distribution instead. It is a standard convention in statistics to present results on tail estimation for the upper tail of distributions. Moreover, actuarial risk theory is a theory of positive rvs. Hence our focus on loss distributions facilitates the application of techniques from these fields as well.

Because of its intuitive appeal and simplicity, it comes as no surprise that VaR has become the standard risk measure used around the world today: it is at the core of international regulation in both banking (BCBS, 2012, 2013) and insurance (European Parliament , 2009b) sectors. The two regulatory frameworks respectively go under the names of Basel (I, II, III) and Solvency II. An excellent historic overview of Basel Committee's[24] working can be found in Tarullo et al. (2008). Whereas, for a broader overview on Solvency II see Sandström (2010) and refer to SCOR (2008) for a presentation of the main issues both from a methodological and a more practical point of view.

Despite its simplicity and good attitude towards the tail of the distribution rather than focussing on the central part, (Artzner et al., 1997, 1999) have criticized VaR as a measure of risk on two grounds. Firstly, they showed that VaR is not necessarily *subadditive* so that, in their terminology, is not a *coherent* risk measure.

Definition 1.4 (Coherent Risk measure)**.** A risk measure satisfying the four axioms of monotonicity (i), translation invariance (ii), positive homogeneity (iii), and subadditivity (iv) is called *coherent*.

There are cases where a portfolio can be split into sub-portfolios such that the sum of the VaR corresponding to the sub-portfolios is smaller than the VaR of the total portfolio. This does not properly reflect the notion of diversification and may cause problems if the risk-management system of a financial institution is based on VaR-limits for individual books. Moreover, another critical aspect is that VaR does not give any information about the potential size of the loss that exceeds it. Two distributions may have the same VaR yet be dramatically different in the tail.

1.4.2 Expected shortfall

Acerbi and Tasche (2002) proposed the use of expected shortfall (ES) instead of VaR. Expected shortfall is the expected size of a loss given that it exceeds VaR, and it so takes into account the whole tail of the distribution and, moreover, it fulfils the properties required by a coherent risk measure according to Artzner et al. definition. In particular, see the paper of Embrechts et al. (2015b), where the authors offer seven proofs of the

[24] The complete name is Basel Committee on Banking Supervision (BCBS).

subadditivity of ES, each revealing an important property of ES (such as its dual representation, optimization properties, continuity, consistency with convex order, and natural estimators).

Definition 1.5 (Expected Shortfall, Acerbi and Tasche 2002)**.** The Expected Shortfall (ES) at level $\alpha \in (0,1)$ of a loss variable L, $\mathrm{ES}_\alpha : \mathcal{L}^0 \to \mathbb{R} \cup \{+\infty\}$, is defined as:

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 q_u(L)\, du = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_u(L)\, du,\ L \in \mathcal{L}^0. \tag{1.2}$$

Theorem 1.1. *For $p \in (0,1)$, $ES_\alpha(L)$ is subadditive on $\mathcal{L}^0$. That is,*

$$ES_\alpha(X+Y) \leq ES_\alpha(X) + ES_\alpha(Y) \tag{1.3}$$

for all $X, Y \in \mathcal{L}^0$ and $p \in (0,1)$

Remark 1.1. In practice $\mathrm{ES}_\alpha(L)$ is only used on $\mathcal{L}^1$, but (1.3) is trivially true if the right-hand side is infinite; this is reflected in the $\mathbb{R} \cup \{+\infty\}$ above in the definition 1.5 of $\mathrm{ES}_\alpha(L)$[25].

In the expression 1.2, instead of fixing a particular confidence level α, we average VaR_α over all levels $u \geq \alpha$ and thus "look further into the tail" of the loss distribution. ES_α depends only on the distribution of L and $\mathrm{ES}_\alpha \geq \mathrm{VaR}_\alpha$.

Remark 1.2. For continuous loss distributions an even more intuitive expression can be derived which shows that expected shortfall can be interpreted as the expected loss that is incurred in the event that VaR is exceeded. For F_L continuous, the definition of Expected Shortfall is equivalent to

$$\mathrm{ES}_\alpha(L) = E[L \mid L > \mathrm{VaR}_\alpha], \tag{1.4}$$

hence its name.

This measure has been independently studied by Rockafellar and Uryasev (2000, 2002) under the name conditional Value-at-Risk (CVaR); in particular, in these papers they showed that expected shortfall can be obtained as the value of a convex optimization problem.

Throughout the literature there is no unanimity when it comes to definitions and notation for specific risk measures. Indeed, there are actually a number of variants on the expected shortfall risk measure with multiple names: for example tail conditional expectation (TCE), worst conditional expectation (WCE) and the already mentioned conditional

[25] An important part of the early literature considers risk measures defined on $\mathcal{L}^\infty$; however, insurance and financial portfolios are primarily exposed to unbounded risks. Embrechts et al. (2015b) shows that subadditivity of $\mathrm{ES}_\alpha(L)$ on $\mathcal{L}^\infty$ directly implies the case of $\mathcal{L}^0$.

VaR (CVaR)[26]. However, for continuous loss distributions they all coincide. A wide discussion regarding different existing notions of expected shortfall and their relationship are presented in the paper of Acerbi and Tasche (2002). One problem with tail risk measures like VaR, TCE, and WCE, when applied to discontinuous distributions, may be their sensitivity to small changes in the confidence level α. In other words, they are not in general continuous with respect to the confidence level α. In contrast, ES_α is continuous with respect to α. Hence, regardless of the underlying distributions, one can be sure that the risk measured by ES_α will not change dramatically when there is a switch in the confidence level by, say, some base points.

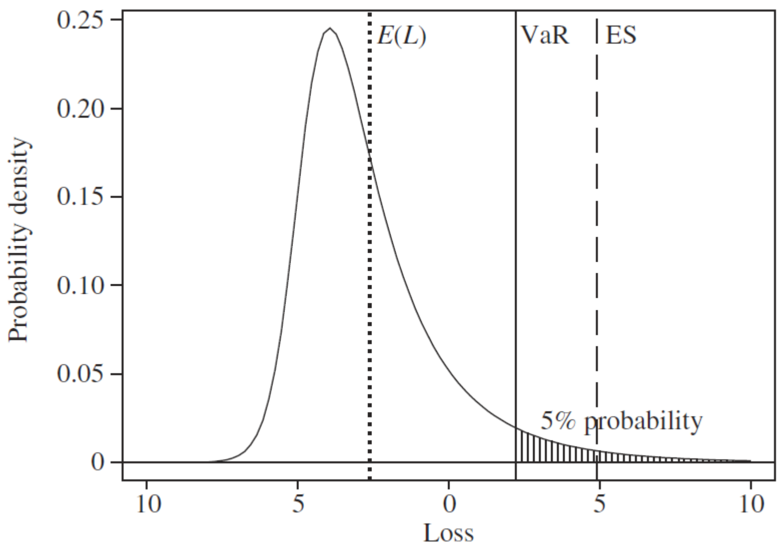


Figure 1.5: An example of a loss distribution with the 95% VaR marked as a vertical line; the mean of the loss rv is shown with a dotted line and the alternative risk measure ES at the 95% confidence level is marked with a dashed line.

1.4.3 Risk measure's properties

Considering the obvious central role played by tail risk measures within the solvency regime and as an integral part of risk management, the study of their properties have been the focus of both academic and practitioners communities.

The notion of *convex* measure of risk is an extension of the concept of coherent risk measure, and its analysis in the context of risk management and mathematical finance began with Föllmer and Schied (2002) (see also Frittelli and Gianin, 2002).

[26] As showed in (Artzner et al., 1999), proposition 5.1, we have the inequality $TCE_\alpha \leq WCE_\alpha$. The identification of TCE and WCE is to a certain degree a temptation though the authors actually did their best to warn the reader. WCE is in fact coherent but very useful only in a theoretical setting since it requires the knowledge of the whole underlying probability space while TCE lends itself naturally to practical applications but it is not in general coherent.

Definition 1.6 (Convex Risk measure)**.** A *convex risk measure* is a risk measure fulfilling the axioms of monotonicity (i), translation invariance (ii) and convexity (v).

Afterwards, another relevant concept subject of study was the relation between the risk measure's property of ranking risks and stochastic dominance. Although a financial risk, often modelled by a probability distribution, cannot be characterized by a single number, sometimes one needs to assign a number to a risk position: the ranking of risks is an example. Risk measures are suitable for such purpose being risk quantitative tools that map risks to numbers. We are aware that there is a long tradition of decision theory, which uses stochastic orderings to formulate some favourable properties of risks. In this context, Bäuerle and Müller (2006) studied in general the problem of consistency of risk measures with stochastic orders. In particular, they say that a risk measure ρ is consistent with a stochastic order $\preceq$, if considering two risks such that $X \preceq Y$, this implies $\rho(X) \leqslant \rho(Y)$. They found out that monotone[27] law-invariant risk measures are consistent with *first order stochastic dominance* ($\preceq_{st}$ or *usual stochastic order*), while convexity[28] of a law-invariant risk measure implies consistency with respect to *convex order* ($\preceq_{cx}$, the converse statement is not true). Moreover, they also proved that law-invariant risk measures, having both the monotonicity and convexity property, are consistent with respect to *increasing convex order* ($\preceq_{icx}$, also knowns as *stop-loss order* in actuarial science). It is easy to translate the last result into one for the *increasing concave ordering* ($\preceq_{icv}$, also known as *second order stochastic dominance* (SSD)) because of the the following equivalence: $X \preceq_{icv} Y$ holds if and only if $-X \preceq_{icx} -Y$. The finding regarding the convex ordering is particularly relevant because it is often used in the actuarial sciences to compare risks, and it is related to notions of risk aversion. It has the intuitively appealing meaning that any risk averse decision maker prefers the risk which is smaller in this ordering. Therefore, this consistency is an important property a risk measure should have. We refer the reader to the section 2 of their paper to recall some basic definitions and results from the theory of stochastic orders.

1.4.4 Estimation and Robustness

The issue regarding the lack of knowledge concerning the distribution of returns, we identified at the beginning of the current section 1.4 (a), leaves us with the problem of how actually computing the risk measures described so far. Therefore, always within the context of discussing desirable properties risk measures should satisfy, a wide literature acknowledged the need of their estimation from historical and/or simulated data via reliable

[27] That is satisfying the property of monotonicity.

[28] We would like to highlight to the reader that this means satisfying the property of convexity; it is clearly different from referring to a "convex risk measure" because, as defined by Föllmer and Schied (2002), it would require to satisfy also the properties of monotonicity and translation invariance.

estimators.

A whole industry of papers on estimating VaR and/or ES has emerged, and this under a wide variety of model assumptions on the underlying P&L process $(L_t)_t$. McNeil et al. (2015, 2005) contains a summary of results stressing the more statistical issues. Another excellent companion text, that has more an econometric point of view, is Danielsson (2011). Extreme Value Theory (EVT) methodology is for instance presented in McNeil and Frey (2000). The recent paper of Chavez-Demoulin et al. (2015) use EVT in combination with generalized additive models in order to model VaR for Operational Risk based on covariates, while Chavez-Demoulin et al. (2014) propose a nonparametric extension of the classical Peaks-Over-Threshold method from Extreme Value Theory to fit the time varying volatility in situations where the stationarity assumption may be violated by erratic changes of regime.

Another important issue when proceeding with risk measures estimation is robustness, namely whether risk measure estimates remain relatively insensitive to small perturbation in the underlying model and/or data changes '(Hampel, 1971; Huber and Ronchetti, 2009). Without robustness (defined in an appropriate sense), results may not not meaningful, since then small measurement errors in the loss distribution can have a huge impact on the estimate of the risk measure. The importance of robustness properties of risk measures has only fairly recently become a focal point of regulatory attention. By now, numerous academic as well as applied papers address the topic. Conflicting views typically result from different notions of robustness. Huber and Ronchetti (2009) have a pure statistical approach and robustness mainly concerns so-called *distributional* (or *Hampel-Huber*) *robustness*: what are the consequences when the shape of the actual underlying distribution deviates slightly from the assumed model? Hansen and Sargent (2008) is characterised by an economics decision making approach, the emphasis lies more on robust control, in particular, how should agents cope with fear of model misspecification. Finally, an area of research that also uses the term robustness is the field of Robust Optimization as, for instance, summarised in Ben-Tal et al. (2009). The main point is that, in some form or another, robustness lies at the core of financial and insurance risk management. Since law-invariant risk measures are a specific type of statistical functionals, their robustness properties are already extensively studied in the statistical literature. Robustness of risk measures is the focus of a growing academic literature Cont et al. (2010), Bellini et al. (2014), Kiesel et al. (2016) Kou et al. (2013), Krätschmer et al. (2014) and Embrechts et al. (2015a).

It is often argued in the literature that quantile-based risk measures such as VaR are more robust when compared to mean-based risk measures such as ES when the notion of robustness used is Hampel's (Hampel, 1971; Huber and Ronchetti, 2009). These type of arguments can be found for instance in Cont et al. (2010), Kou et al. (2013), Kou and

Peng (2014) and Emmer et al. (2015). In particular, the former authors showed that, in general, quantile estimators are robust with respect to the weak topology, and coherent distortion estimators are not robust in the same sense; this is consistent with Hampel's notion of robustness for L-statistics, as discussed in Huber and Ronchetti (2009)[29]. Kou et al. (2013) state that "Coherent risk measures are not robust", they showed that VaR is more robust compared to ES, and this was with respect to a small change in the data by using tools, such as influence functions and breakdown points; see, also Kou and Peng (2014), for similar results with respect to Hampel's notion of robustness.

Therefore, with reference to the weak topology, most of the common risk measures (variance, expected shortfall, expectiles, and mean) are discontinuous (Kiesel et al., 2012). One possible approach to overcome this conflict between robustness and risk ranking consistency, is to use other metrics for the analysis of robustness that lead to less restrictive definitions. In this context, for instance Cambou and Filipović (2015) find out that "ES is robust, and VaR is non-robust based on the notion of ϕ-divergence". Krätschmer et al. (2012, 2014), instead, argue that Hampel's classical notion of quantitative robustness is not suitable for risk measurement, and they propose and analyse a refined notion of robustness that applies to tail-dependent law-invariant convex risk measures on Orliz spaces. Using a different definition, they introduce a continuous scale of robustness and thus introducing an index of quantitative robustness. Moreover, mean, VaR, and Expected Shortfall are robust with respect to a stronger notion of convergence (metric) that is *the Wasserstein distance* (see for example Emmer et al., 2015; Kiesel et al., 2012) and Bellini et al. (2014) show that Expectiles are Lipschitz-continuous with respect to the Wasserstein distance with constant $K = max\{\alpha/(1-\alpha); (1-\alpha)/\alpha\}$, which implies continuity with respect to the Wasserstein distance.

Another relevant approach is the one followed in Pesenti et al. (2016) that consists in relaxing the requirement that risk measures has to be robust on the whole space of integrable random variables. Conducting an analysis of regions on which (low-invariant convex) risk measures are robust, is reasonable because, as Cont et al. (2010) observed, "this case is not generally interesting in econometric or financial applications since requiring robustness against all perturbations of the model is quite restrictive".

With reference to research related to risk aggregation, Embrechts et al. (2015a) show that VaR is less robust compared to ES with respect to dependence uncertainty in aggre-

[29]As highlighted in Embrechts et al. (2015a), the qualitative robustness of a statistical estimator, as in Hampel (1971), is equivalent to the continuity of the corresponding risk measure at the true distribution, with respect to the weak topology. In general, to analyse statistical robustness of a given risk measure, one typically studies continuity properties of that risk measure in a given metric d, say. Hence we say that a law-invariant risk measure ρ is *d-robust* at a distribution F if $d(F_n, F) \to 0$ implies that $\rho(X_n) \to \rho(X)$, where $X_n \sim F_n, n = 1, 2, \ldots$ and $X \sim F$. For example, the Lévy distance is used in Cont et al. (2010) and note that the Lévy distance metrizes the weak topology on the set of distributions. A law-invariant risk measure is said to be robust in Hampel's sense if it is continuous with respect to convergence in distribution.

gation. The authors introduce a notion of aggregation-robustness, under which ES and other spectral risk measures are robust. They also show that VaR generally exhibits a larger dependence-uncertainty spread compared to ES. Robustness is studied in the context of risk aggregation also in Pesenti et al. (2016), where a risk measure is applied on an *aggregation function* of a random vector (risk factors), this composition is defined as *aggregation measure*. Their findings generalise the results on aggregation robustness in Embrechts et al. (2015a) to the class of law-invariant convex risk measures with uncertainty in the marginal distributions, including, as a special case, the aggregation via the ordinary sum. In particular, they show that in order to have robustness of aggregation measures, it is sufficient that the margins of the vector of risk factors belong to uniformly integrable sets and that the aggregation function is Lipschitz continuous in the tail. It is relevant that there are no constraints required on the risk factors' dependence structure. In contrast, Value at Risk, that is a non-convex risk measure, requires restrictions on the possible dependence structures of the input vectors, in order to be robust in a risk aggregation context .

One more important contribution is the introduction of a concept of robustness into the problem of risk sharing among agents, this insight is presented in the academic paper of Embrechts et al. (2016).

From a more practical point of view, Emmer et al. (2015) point out that the concept of robustness used in Cont et al. (2010) is potentially more intuitive because it takes the estimation procedure into account. Robustness is meant as the sensitivity of the risk measure estimate to the addition of a new data point to the data set used for the estimation procedure. It turns out that for the same risk measure the estimation method can have a significant impact on the sensitivity. For instance, the risk measure estimate can react in a completely different way on an additional data point if we fit a parametric model instead of using the empirical loss distribution[30].

Thus, robustness in the sense of Cont et al. (2010) relates more to sensitivity to outliers in the data sample than to mere measurement errors. However, in finance and insurance, large values do occur and are not outliers or measurement errors, extreme values often occur as part of the data-generating process. In particular, in (re)insurance, one could argue that large claims are actually more accurately monitored than small ones, and their values better estimated. Hence the question of robustness in the sense of Cont et al. (2010) may not be so relevant in this context.

Finally, note that in practice, the estimation of ES will often be based on larger sub samples than the estimation of VaR. For instance, when using simulation iterations, ES is estimated over the whole observations simulated for the tail, while the VaR estimate is

[30]The authors conclude that "...historical Value-at-Risk, while failing to be subadditive, leads to a more robust procedure than alternatives such as Expected Shortfall".

based on a small neighbourhood of the order statistic.

> *"This new look at robustness will then help us to bring the argument against coherent risk measures back into perspective: robustness is not lost entirely, but only to some degree when VaR is replaced by a coherent risk measure, such as ES."*
>
> (Krätschmer et al. (2014))

Given the overview of the recent literature on robustness of risk measure we can understand that the discussion is not yet closed: robustness can be interpreted in several ways and it is not clear what kind of robustness the Basel Committee has in mind (Embrechts and Hofert, 2014).

1.4.5 Robust Backtesting of Risk Measures: Elicitability

The question 8 of BCBS (2012) concerns possible regulatory regime switch for market risk management from VaR to ES and the issue of "robust backtesting". Replacing VaR by ES in Basel is under current discussion, and in particular Under the Fundamental Review of the Trading Book (FRTB) capital charges for the trading book are based on the coherent expected shortfall (ES) risk measure, which show greater sensitivity to tail risk[31].

An interesting criterion when estimating a risk measure is *elicitability*, introduced by Osband (1985) and Lambert et al. (2008), then by Gneiting (2011). In the latter paper, the author offers a theory to assess the quality of statistical forecasts introducing the notion of elicitability. For ease of discussion, we will refer to a forecast as *elicitable* if it is "properly" backtestable. We are not presenting its rigorous definition here, thus for the precise mathematical meaning of "properly" we refer to the latter paper and the references therein. An elicitable risk measure is a statistic of a P&L distribution that can be represented as the solution of a forecasting-error minimization problem. Intuitively, elicitability of a functional (risk measure) of probability distributions may be interpreted as the property that the functional can be estimated by generalised regression. Another feature, that makes elicitability a relevant property for a risk measure, is that we can use consistent scoring functions to compare series of forecasts obtained by different modelling approaches and obtain objective guidance on the approach that gives the best forecasting performance (see Gneiting, 2011, for a detailed discussion).

It is shown that in general, VaR is elicitable whereas ES is not, leading to less straightforward backtesting methods (note that in Gneiting, 2011, ES is actually referred to as Conditional Value-at-Risk, CVaR; see also Bellini and Bignozzi, 2015; Ziegel, 2016). In Section 3.4 the author states that: "This negative result may challenge the use of the

[31] As a result of the Fundamental Review of the Trading Book (BCBS, 2013) a 10-day ES at the 97.5% level will be the main measure of risk for setting trading book capital under Basel III (BCBS, 2016).

CVaR functional as a predictive measure of risk, and may provide a partial explanation for the lack of literature on the evaluation of CVaR forecasts, as opposed to quantile or VaR forecasts". These findings gave rise to a rich debate on the subject.

Although Gneiting (2011) suggested that the lack of elicitability of ES called into question our ability to backtest ES and its use in risk management, a consensus is now emerging that the problem of comparing the forecasting performance of different estimation methods is distinct from the problem of addressing whether a single series of ex ante ES estimates is consistent with a series of ex post realizations of P&L, and that there are reasonable approaches to the latter (Kratz et al., 2016). For instance, Acerbi and Székely (2014) recently pointed out that elicitability (or lack of elicitability) is not relevant for backtesting of risk measures but rather for comparing the forecast performance of different estimation methods.

However, it should be noted that ES satisfies more general notions of elicitability, namely *conditional elicitability* or *joint elicitability.* Alternatively, ES is also said to be "2^{nd} order" elicitable. This concept was introduced by Emmer et al. (2015)[32] and offers a way of splitting a forecasting method into two component methods involving elicitable statistics and separately backtesting and comparing their forecast performances. Since ES is the expected loss conditional on exceedance of VaR, we can first backtest VaR using an appropriate consistent scoring function and then, treating VaR as a fixed constant, we can backtest ES using the squared error scoring function for an elicitable mean.

Other recent contributions are: Fissler and Ziegel (2016) show that VaR and ES are jointly elicitable in the sense that they jointly minimize an appropriate bi-dimensional scoring function, allowing the comparison of different forecasting methods that give estimates of both VaR and ES; Acerbi and Székely (2016) introduce a new concept of backtestability satisfied in particular by expected shortfall; another recent paper on the subject is also Davis (2016).

With a view on the feasibility of backtesting, in recent studies Bellini and Bignozzi (2015) and Ziegel (2016) suggest Expectiles as law-invariant, coherent (i.e. subadditive) and elicitable alternatives to Expected Shortfall[33]. In particular in Bellini and Bignozzi (2015), it is shown that, under a slight modification of elicitability, the only elicitable and coherent risk measures are the expectiles. However, while Expectiles indeed have a number of attractive features, their underlying concept is less intuitive than the concepts for VaR or ES. In addition, Expectiles are not comonotonically additive which implies that in applications they may fail to detect risk concentrations due to non-linear dependencies (Emmer et al., 2015).

[32]In particular, they showed that for continuous distributions with finite means, ES is conditionally elicitable; moreover, for distributions with finite second moments, the variance is conditionally elicitable.

[33]See Gneiting (2011): for distributions with finite means, Expectiles are elicitable.

1.4.6 Comparison: What is the best risk measure in practice?

The extensive debate on desirable properties of regulatory risk measures, in particular VaR and ES, is summarized in the academic papers Embrechts et al. (2014) and Emmer et al. (2015)[34]. Both risk measures are used in industry and regulation; VaR more in banking, ES more in insurance, the latter especially under the Swiss Solvency Test[35].

Summarising what we have presented so far, we compare and contrast both risk measures on the basis of four broadly defined criteria: **(C1)** Existence; **(C2)** Ease of accurate statistical estimation; **(C3)** Subadditivity, and **(C4)** Robust forecasting/backtesting and risk measure's robustness.
Clearly, concerning **(C1)**, $VaR_\alpha(L)$, when properly defined, always exists. For $ES_\alpha(L)$ to exist, one needs $\mathbb{E}(L) < \infty$. The latter condition may seem reasonable, see however the discussion on Operational Risk in McNeil et al. (2005) Section 10.1.4 and see also Chavez-Demoulin et al. (2015).

Regarding **(C2)**, and this especially for α close to 1 corresponding to values typical for financial and insurance regulation, both measures are difficult to estimate with great statistical accuracy. Of course, confidence intervals around ES are often much larger than corresponding ones around VaR. One needs an excellent fit for $1 - F_L(x)$, x large, that is the tail of the distribution. EVT, as summarized in McNeil et al. (2005) Chapter 7, comes in as a useful tool. EVT cannot be the absolute cure for solving such high-quantile estimation problems, but it offers, more importantly, guidance on the kind of questions from industry and regulation which are beyond the boundaries of sufficiently precise estimation.

With regard to **(C3)**, we already discussed the fact that in general VaR is not subadditive whereas ES is the convex risk measure used most widely in the practice of risk management. The debate for the lack of proper aggregation has been ongoing within academia since VaR was introduced around 1994. This very important property tips the balance clearly in favour of ES. More importantly, thinking in ES-terms makes risk managers concentrate more on the "what-if" question, whereas the VaR has been heavily criticized as it only looks at the frequency of extremal losses but not at the severity.

With reference to **(C4)**, backtesting models to data and sensitivity of risk measures to underlying model deviations and/or data changes, remain crucial features throughout finance. The notions of elicitability and robustness add new aspects to this discussion. We have seen that the literature discussing these two characteristics is still developing.

[34] Refer to BCBS (2014) and IAIS (2014) instead for contributions from the regulators of the banking and insurance, respectively. See also BCBS (2016) for a recent discussion concerning market risk under Basel III and Sandström (2010) for an overview in the context of Solvency II.

[35] ES at the 99% level for annual losses is the primary risk measure in the Swiss Solvency Test (SST).

The major findings are the followings. The Value at Risk is elicitable and robust. The fact that VaR does not cover tail risks "beyond" VaR is a more serious deficiency although ironically it makes VaR a risk measure that is more robust than the others. ES makes good for the lack of subadditivity and sensitivity for tail risk of VaR but has recently be found to be not elicitable and not robust in the sense of Cont et al. (2010).

However, all is not lost. Concerning its weaknesses in backtesting procedures, there is a recent stream of literature that pointed out that lack of elicitability is not relevant for backtesting of risk measures but rather for comparing the forecast performance of different estimation methods. Moreover, ES satisfies more general notions of elicitability, namely conditional elicitability and there are a number of feasible approaches available to the backtesting of ES (e.g. based on distribution forecasts, linear approximation of ES with VaR at different confidence levels, or directly with Monte-Carlo tests). It must be conceded that to reach the same level of certainty more validation data is required for ES than for VaR. Regarding the question of robustness there are conflicting views that typically arise from different notions robustness itself. When robustness is to be interpreted in a precise statistical sense, like in Huber and Ronchetti (2009) or Hampel (1971), there is a conflict between robustness of a risk measurement procedure and the subadditivity of the risk measure and Expected Shortfall is not robust. This conclusion seems in line with the results replacing elicitability by robustness. However, Expected Shortfall is robust with respect to a stronger notion metric namely the Wasserstein distance. In addition, researches related to risk aggregation, show that VaR is less robust compared to ES with respect to dependence uncertainty in aggregation.

The above publications and discussions show that there is no widely accepted concept of risk measure that satisfies the various stakeholders involved. Clearly, when looking at the discussion in order to make the international banking and insurance world a safer place for all involved, a whole new field of research is opening up. At this moment, it is difficult to identify the right approach.

ES seems the best for use in practice, despite some caveats with regard to its estimation and backtesting, which can be carefully mitigated. As argued also by Emmer et al. (2015) there is not currently sufficient evidence to justify an all-inclusive replacement of ES by its recent competitor Expectile. Nonetheless, it is certainly worthwhile to keep in mind Expectiles as an alternatives to ES and VaR in specific applications. Whereas there is a tendency to move from VaR to ES, for a while both risk measures will coexist for regulatory purposes.

Building on these reasons, for the purpose of this thesis we adopt both VaR and ES as risk measures in our model. This will allow us to compare their performance within our framework.

1.5 Organization of the thesis

The thesis is divided into two main parts: the first part (Chapter 1) widely discuss the purpose of this research, providing a complete overview of the motivation and key aspects of the theme. The first chapter offers an extensive discussion of the importance of the topic in the context of financial and insurance applications, including the current capital modelling framework, as well as an introduction to risk measures and a recap on standard portfolio selection theory. The second part, represents the core of our work, which is structured in 3 chapters (Chapters 2, 3, and 4) devoted to providing details regarding the methodology of the empirical study and ultimately giving a presentation of the empirical results. Thus, the rest of the dissertation is structured as follows:

Chapter 2 is dedicated to presenting the core methodology of our work. We begin by reviewing the fundamentals of Markowitz's MVO approach; useful to introduce our optimal portfolio allocation problem and to understand how we overcame the two main MPT's assumptions. We then introduce extreme value theory (EVT), providing an overview of its role in risk management (RM), as a method for modelling and measuring extreme risks. The main theoretical results of the so called "threshold exceedances approach", upon which we build our model, are presented in detail. All of the preliminaries are employed to build and formally define the optimal portfolio selection problem. Finally, we discuss several practical but crucial aspects characterising the EVT modelling implementation;

Chapter 3: is devoted to the presentation of a multivariate framework for studying the dependence among rare events in finance, suitable when asset returns consist of more than one component, as in the case of a portfolio. We aim at performing an analysis to identify the correct extremal dependence structure of the portfolio's constituents and to understand how this may impact the optimal asset allocation problem. For this purpose two measures from multivariate EVT are introduced $\bar{\chi}$ and χ, and their estimation and statistical inference via the Hill estimator is discussed;

Chapter 4: This final chapter outlines the application of the methodology presented in Chapters 2 and 3 to a portfolio of international equity market indices, leading up to the concluding section 4.6 . In particular, in the empirical application we consider the equity indices of twelve large capitalization international markets. *Thomson Reuters Datastream* is the data source and the time series of the total market indices cover the period from 02-Jan-1980 to 30-Aug-2016. The portfolio optimisation is performed considering the market equity indices of the Group of Five G-5 countries: France, Germany, Japan, the United Kingdom, and the United States.

Conference presentations

The material discussed in this thesis, with slight variations, has been presented throughout the last academic year at the following international Conferences:

- Astin Colloquium 2016 of the International Actuarial Association (IAA), held in Lisbon (Portugal) between 31^{th} May 2016 and 3^{rd} June 2016. Organisation: the Portuguese Institute of Actuaries (IAP) and the ASTIN Section of the IAA;
- Invited speaker at the 5^{th}Iberian Congress of Actuaries 2016, Lisbon (Portugal), $6^{th} - 7^{th}$ June 2016. Organisation: the Portuguese Institute of Actuaries (IAP) and the the Spanish Institute of Actuaries (IAE);
- 3^{rd} European Actuarial Journal Conference 2016, held in Lyon (France) in September $5^{th} - 8^{th}$, 2016. Organisation: Institut de Science Financière et d'Assurance (ISFA, Universitò Lyon 1) and Institut des Actuaires (IA).

Chapter 2

Portfolio optimisation using EVT

This chapter is dedicated to presenting the core methodology of our work. In a world where the new regulatory framework of financial sectors (banking and insurance) require a proper understanding of managing risks, and give importance to extreme adverse events, this thesis focusses upon solving the optimal selection problem in a consistent manner.

2.1 Introduction

The question of properly modelling key aspects of real-world optimum portfolio choice problems of individuals and institutions, has always been an important focus for the academic research as well as practitioners' interest, whose earning depends on the outcome of their investment decisions.

The modern portfolio selection theory can be traced back to the seminal work by Markowitz, and his mean-variance formulation clearly had far-reaching impact on finance. It changed the focus of investment analysis away from individual security selection toward the concept of diversification and the impact of individual securities on a portfolio's risk-return characteristics. Markowitz claims that an investor who wants to decide how to allocate funds among various assets should act maximising expected portfolio return while minimising portfolio risk, measured using the variance operator. Moreover, for investors willing to assume normally distributed returns, he showed that solving the asset allocation problem was as straightforward as solving a system of linear equations (see Section 2.2).

Markowitz's famous paper is usually the only contribution many of us might recognise. Actually, many years later, comparing Roy's paper (Roy, 1952) to his own, Markowitz (1999) writes "On the basis of Markowitz (1952), I am often called the father of modern portfolio theory (MPT), but Roy (1952) can claim an equal share of this honour".

Roy claims that most firms are primarily concerned with avoiding possible disaster, and thus define the so called *principle of Safety First* which asserts that "it is reasonable, and probable in practice, that an individual will seek to reduce as far as is possible the chance of a catastrophe occurring[1]". This concept was introduced by Roy (1952) and then further developed by Arzac and Bawa (1977).

To optimally allocate wealth among assets, a safety-first firm defines a threshold d, below which the portfolio wealth is considered to be a disaster. In this framework, while Markowitz is concerned about the variability of a portfolio, Roy is rather interested in avoiding extreme losses through minimizing the disaster probability.

The safety-first criterion was originally stated in terms of securing a minimal return level with a high probability. The impractical original problem, was made operational by Roy himself through the use of Bienaymé-Tchebycheff inequality to derive an upper bound for the disaster probability, and then solves a surrogate problem of the original safety-first formulation by minimizing this upper bound under a single-period setting. However, the problem is that while this bound is robust, it is also a highly inaccurate estimate of the downside risk.

Contrary to the Markowitz's approach, concentrating on the first and second order central moments, Roy's safety-first principle is concerned more about the extreme events with significant magnitude that clearly lies in the tail of the probability distribution. When tackling the portfolio selection problems under the safety-first principle, almost all the works in the literature deviate from the original conceptual framework set by Roy that consist in minimizing the disaster probability, while subject to a budget constraint and a mean constraint.

The first approximation approach is to replace the disaster probability by its upper bound as Roy does himself. The second approximation approach in the literature of the safety-first principle is to remove the mean constraint, while keeping the disaster probability as the objective to be minimized. A possible reason to ignore the mean constraint could be due to the fact that Roy does not mention it explicitly in his paper; many authors argue, however, that Roy does it implicitly in the article[2]. Roy is, without doubt, the father of using a probability objective or probability constraint in asset allocation.

In this thesis we develop a methodology that is consistent with the safety-first investment strategy just presented. We have several motives for this research approach. First, we believe portfolio selection with limited downside risk to be a practical problem. Even if agents are endowed with standard concave utility functions such that to a first order approximation they would be mean-variance optimisers, practical circumstances often impose constraints that elicit asymmetric treatment of upside potential and downside risk.

[1] Such a principle has been applied to the theory of risk in insurance companies. See Cramér (1930).

[2] In Roy (1952) on p.434, the author states that "If in default of minimizing $P(\xi \leq d)$, we operate on $\sigma^2/(m-d)^2$, this is equivalent to maximizing $(m-d)/\sigma \dots$", where, in terms of Roy's notations, ξ is the final wealth, σ^2 is its variance and m its expected value, while d is the preselected disaster level. Mathematically, the two expressions are equivalent if and only if $m > d$. Thus, the mean constraint $(m > d)$ is indeed included in Roy's original safety-first principle.

For instance, prudential regulation of banks and insurance companies request to those institutions to calculate their capital requirements using quantile-based or mean-based risk measures, like VaR and ES respectively. Likewise, pension funds are often required by law to structure their investment portfolio such that the risk of underfunding is kept very low, e.g. equity investment may be capped. Second, there is a wealth of experimental evidence for loss aversion (see for instance Knetsch and Thaler, 1990). Other evidence is provided through consumption behaviour: for example Deaton (1991, 1992) shows that consumption responds asymmetrically to good and bad states.

In general, the inclusion of downside risk measures in portfolio optimisation problem represent the more natural way of characterising the mindset of very risk averse investors who act primarily to guard against catastrophic losses.

Although variance is still universally recognised as a possible risk measure, it is often not completely adequate for the type of risk that regulators and practitioners are trying to evaluate and manage. In light of the evidence in favour of investors' sensitivity to downside losses, we replace the symmetric variance with the two risk measures commonly used in finance: Value at Risk and Expected Shortfall. Assessing risks using these quantile or mean based measures is consistent with the safety-first approach[3] of Roy (1952) and Arzac and Bawa (1977)[4].

Choosing this framework we do not mean to claim that this is the best choice one could made in every situation and we should abandon the MVO. Indeed, as argued by Bensalah et al. (2002):

"There is no final answer or optimal measure of risk"

The optimality of a given allocation is conditional on the actual correspondence of its underlying assumptions to the risk preference of the investor. In our case, a plausible profile of risk preference, alternative to that based on variance, that matches the idea of an investor very concerned with extreme events (that cause large losses), is perfectly represented and justified by the safety-first criterion.

The past few decades have witnessed numerous extensions and generalizations of Markowitz's groundbreaking classic model and typically mainstream finance rests on the assumption of normality. For normally distributed returns, using VaR_α or ES_α poses no difficulty, due to the following feature of the normal distribution, or in general elliptical distributions: a linear combination of the marginal components is again normally distributed. If we assume that asset returns are multivariate normally distributed, then any

[3]Notice that risk minimisation with respect to VaR or ES and with reference to the failure probability are, with minor variations, the same problem.

[4]They noted that the original criterion fails to order risky assets, which are unambiguously ordered by the principle of absolute preference. By considering a lexicographic form of the safety-first principle, to the extent that the investor is only concerned about safety when the failure probability is above a critical level and otherwise maximizes expected return, and by allowing for borrowing and lending, the authors were able to resolve the theoretical shortcomings of the original criterion.

linear combination of those assets (a linear portfolio of the assets) is normally distributed. If $X \sim N(0, \sigma_X^2)$ represents the losses of any such portfolio, then the VaR at confidence level α is $VaR_\alpha(X) = \sigma_X \Phi^{-1}(\alpha)$, where $\Phi(z) = P(Z \leq z)$ and $Z \sim N(0,1)$. This implies that the VaR of the portfolio is a simple linear transformation of the portfolio's variance. Therefore, choosing weights which minimize the portfolio's variance is equivalent to choosing weights which minimize the value at risk, or for that matter, the expected shortfall. These weights will not depend on the confidence level α.

However, several empirical studies found out that the distribution of financial returns actually exhibits heavy tails and asymmetry, features that cannot be described only looking at its mean and variance[5]. Such departure from normality is crucial for safety-first investors. Ignoring this evidence could lead to systematically underestimate the probability of observing very large losses on the portfolio, and might result in an optimistic asset allocation.

The question, then, is whether it is possible to improve the estimate of the two risk measures taking into account the heavy tailed feature of asset returns, generating failure probabilities that are considerably different from normal probabilities. Since in our setting we are concerned with extreme market movements, and moreover VaR/ES calculations, by definition, deal with the tails of a probability distribution, techniques from Extreme Value Theory (EVT) may be particularly effective.

In this work we examine the optimal portfolio selection problem for risk averse investors who are concerned with extreme market movements and aim to reduce the risk of suffering from unexpected large losses. We build our model relying on Roy's safety first approach to investment strategy and writing downside risk in terms of portfolio Value-at-Risk (VaR) and Expected Shortfall (ES). These are currently the most used risk measures in the financial world. Time series of financial asset values exhibit well known statistical features, such as heavy tails. Therefore, in order to take into account this aspect, we relax the normality assumption that characterised the MPV, and apply the tools of univariate extreme value theory to the asset allocation problem. Classical Extreme Value Theory (EVT), through the *threshold exceedances approach*, yields a methodology for modelling the tails of the distribution and estimating the risk measures.

This enables a much more generalised framework to be developed, employing the distributional assumption most appropriate to the type of financial assets. The identified optimal portfolio is then compared with the one obtained under the classical MPT framework. For problems involving downside risks, that are far in the tails of the distribution (high confidence level α for the risk measures), we expect our technique to outperform the MVO; as it is built upon knowledge of the tail behaviour that must be followed by

[5] Widely discussed in Section 1.3.2 *b* Chapter 1.

any unbounded fat-tailed distribution.

In addition, we perform an analysis to quantify and to understand how the extremal dependence between markets may impact the asset allocation problem. EVT provides useful dependence measures for this purpose (Coles et al., 2001), and we find that about half of the international equity markets examined are asymptotically independent. This asymptotic independence should induce a greater benefit to diversification as we move deeper into the tail (at the highest quantiles α). The methodology is tested on real data and applied to portfolios made up of international equity market indices.

2.1.1 Overview of related work

The safety-first approach to portfolio selection calls for accurate estimates of the failure probability. In this sense, as far as failure can be considered a rare event, EVT provides suitable tools for accurate calculations. Managing extreme market risk is a goal of any financial institution or individual investor and is nowadays also the focus of financial regulators and their effort to guarantee solvency of the whole system.

The approach to EVT in this thesis follows most closely Embrechts et al. (1997); McNeil et al. (2005, 2015) that is a priceless resource for the application of univariate extreme value theory in the Insurance and Finance fields. Other texts on EVT include for example Beirlant et al. (1996, 2004); De Haan and Ferreira (2006); Falk et al. (2011) and Reiss et al. (2007). All of these texts emphasize applications of the theory in insurance and finance although much of the original impetus for the development of the methods came from hydrology. The application of EVT to financial modelling is a growing field and the literature is wide; the theory for multivariate extreme value theory is fully developed for instance in Resnick (1987) while the book by Coles et al. (2001) provides an excellent introduction to the application of both univariate and multivariate extremes.

The first influential studies applying EVT to VaR computation are the following: Embrechts et al. (1998, 1999) Pownall and Koedijk (1999), Danielsson and de Vries (1997b, 2000); Danielsson et al. (1998), Longin (2000, 2001), McNeil (1997, 1998); McNeil and Frey (2000) and Neftci (2000).

Embrechts et al. (1999) advocate the use of a parametric estimation technique which is based on a limit result for the excess-distribution over high thresholds. This approach will be adopted in this thesis to estimate return distributions of financial time series via EVT[6]. Longin (2000) and McNeil (1998) use estimation techniques based on limit theorems for block maxima. Longin ignores the stochastic volatility exhibited by most financial return series and simply applies estimators for the i.i.d.-case. McNeil, instead, uses a similar approach but shows how to correct for the clustering of extremal events

[6]Explained in detail in Section 2.3.

caused by stochastic volatility. Danielsson and de Vries (1997a,b) use a semiparametric approach based on the Hill-estimator of the tail index. In Pownall and Koedijk (1999), the authors study the crisis of Asian markets, and provide a conditional approach to VaR calculation via the use EVT. They find that the methodology yields an improvement in VaR estimation, compared to the technique employed by RiskMetrics. Neftci (2000) compares the EVT approach to VaR calculation to the standard one based on the normal distribution, dealing with several interest rates and exchange rates. Much more research comparing EVT-driven estimates of VaR to other approaches (both standard and new ones) has been done since these seminal papers appeared. The main overall conclusions one can draw from analysing this literature is that: EVT based estimates of VaR outperform estimates obtained with other methodologies for very high quantiles, namely for $\alpha \geq 0.99$; moving further into the tails, the EVT estimates improve; when $\alpha < 0.99$, evidence is mixed and EVT can perform worse than other techniques.

The following literature, instead, extends the use of extreme value theory for statistically estimating extremal risk measures based on time series observations, to the problem of portfolio choice.

Jansen et al. (2000); Jansen (2001) use univariate semi-parametric EVT approach in the context of portfolio selection with limited downside risk to construct optimal safety-first portfolios. In their framework, they face the problem of choosing between investing in a mutual fund of bonds or a mutual fund of stocks, and find that an assumption of tail fatness is plausible and employ EVT (precisely, the Hill estimator) to calculate the risk associated to each possible portfolio. Along similar lines, Susmel (2001) apply EVT on the tails of the unconditional distribution of Latin American emerging markets stock returns and studies their implications for portfolio diversification of a US investor according to the safety first criterion. The authors find that the Latin American emerging markets have significantly fatter tails than industrial markets, especially, the lower tail of the distribution.

Hyung and De Vries (2007) consider the portfolio problem of choosing the mix between stocks and bonds under a downside risk constraint. They remarks that downside risk criteria for portfolio selection, like the safety-first approach, often yield corner solutions, namely portfolio combinations in which only the least risky assets are chosen. The authors argue that this is due to a focus on a first-order asymptotic expansion of the tail distribution of stock returns, while a second order expansion of the downside risk provides more balanced solutions, in which some proportion of risky assets appears in the optimal portfolio mix as well. Empirical evidence of this theoretical insight is presented by computing optimal portfolios for the problems considered in Jansen et al. (2000) and Susmel (2001) with their new approach, and showing that in this way corner solutions appear less frequently.

Consigli (2002), solves the asset allocation problem in a mean-VaR setting with data including the Argentinean Eurobond crisis of July 2001, and evaluating downside risk by means of both EVT and a jump-diffusion model. Both methods yield accurate estimates of the tail risk in the cases analysed; the latter seems to be more accurate, while the former provides more stable estimates, thus being more appropriate for the determination of capital adequacy in which a regulator is interested.

The mean-VaR portfolio selection problem is analysed also by Bensalah et al. (2002) who considers both the direct problem of maximizing return, subject to a constraint on the VaR, and its dual, that is the problem of minimizing VaR, subject to a lower bound constraint to the expected return. Bensalah makes a comparison between different ways of calculating VaR: duration, historical simulation, normal VaR with normality assumption and EVT. Considering a portfolio of two fixed-income securities (a 1-year treasury-bill and a 5-year zero-coupon bond) and several highly conservative values for the confidence levels, he finds that historical simulation and normal VaR lead to the same allocation irrespectively of the given confidence level: investing the whole capital in the short-term security, both in the maximum return and in the minimum risk problem. This is not the case when they consider EVT based calculations of VaR. Indeed, two important features appear: first the portfolio composition changes as the confidence level grows; second, the riskier (in terms of duration) asset is given non-zero weight (in particular, is an increasing function of the confidence level). In addition, they show that its weight is greater in the minimum risk than in the maximum return problem.

Therefore, Bensalah concludes that an EVT approach, being tailored on extreme events and differently responding to various confidence levels, seems to be better suited to regulatory purposes and to cope with periods of market stress.

Finally, Bensalah et al. (2002) proposes an algorithm to solve the problem when dealing with multiple assets. The reason for its introduction is the relevance of the dimensionality issue in portfolio selection. An improvement over this algorithm is provided by Bradley and Taqqu (2004b); this article, together with Bradley and Taqqu (2004a), are our selected benchmark papers.

2.1.2 Organisation of the chapter

This chapter outlines the following sections: Section 2.2 review the fundamentals of Markowitz's MVO approach in order to introducing our optimal portfolio allocation problem and to understand how we overcame the two main MPT's assumptions.

In Section 2.3, a brief introduction to the broad variety techniques that EVT offers is given. We then present the main theoretical results of the so called "*threshold exceedances approach*", upon which we build our model.

The next Section 2.4 is dedicated to formalising the key assumption, that allow us to model the excess distribution above a fixed threshold of a generic risk, via a Generalised Pareto distribution (GPD). Further, the estimation method to perform the GPD fit is discussed. These are then used in Section 2.5 to define estimators for the tail probabilities and for the two risk measures VaR_α and ES_α.

All of these preliminaries are employed in Section 2.6 to build and formally define our optimal portfolio selection problem for risk averse investors who are concerned with possible extreme market movements. Firstly, this involves the need of introducing some preliminary notation and defining our setting. The rv *loss of the portfolio* $\mathcal{P}$ is derived in order to then properly formalise the asset allocation problem. Secondly, its specific modelling via the univariate EVT approach *structure variable method* (SVM) and all the expressions for the variables of interest are presented.

Finally, in Section 2.7, we discuss several practical aspects characterising the EVT modelling implementation. First, we deal with the estimation procedure on a data sample of multivariate observations. Second, we examine the problem of choosing a proper threshold over which fitting our GPD model. Lastly, a two stage methodology is presented to perform the optimisation needed considering multiple markets. This is necessary as the method usually applied to the bivariate asset problem (called *brute force method*) becomes numerically intractable. In this section we will also perform a simulation study in order to better understand the sensitivity of the EVT threshold exceedances approach to the choice of the threshold. Both the non parametric (Hill estimator) and fully parametric (GPD model) are considered. We find that, in this case, the GPD-based tail estimator is more robust to threshold changes then the Hill method.

2.2 Mean variance optimisation

Markowitz's seminal paper on portfolio selection undoubtedly has had a major impact not only on academic research but on the financial industry as a whole. It contains the first mathematical formalisation of the idea of diversification of investments: the financial version of "the whole is greater than the sum of its parts". In the MVO framework, efficient portfolios are formed by choosing an asset based upon its interaction with other assets in the portfolio as well as on its contribution to the overall portfolio, and not on the basis of its standalone performance. He postulates that an investor should maximize expected portfolio return while minimising portfolio risk, measured as variance of portfolio return. Let us present more details on the mathematical model.

Consider an investment universe of with d different assets $S_1, S_2, \dots, S_d$ with uncertain future returns $R_1, R_2, \dots, R_d$. We denote by $\mathbf{R} = [R_1, \dots, R_d]^\intercal$ the vector of these returns. A portfolio is represented by the n-dimensional vector $\boldsymbol{\omega} = [\omega_1, \dots, \omega_n]^\intercal$ where ω_i denotes the proportion of the total funds invested in security i. The R_i are considered to be

random variables, while the ω_i, $i = 1, \dots, d$ are fixed by the investor.
The (uncertain) return of the portfolio, R_P, depends linearly on the weights $\boldsymbol{\omega}$ and is expressed as

$$R_P(\boldsymbol{\omega}) = \omega_1 R_1 + \omega_2 R_2 + \cdots + \omega_n R_d = \boldsymbol{\omega}^\mathsf{T}\mathbf{R}$$

The yield R_P on the portfolio as a whole is a random variable as well, being a weighted sum of the R_i random variables. The return $R_P(t)$ of a portfolio at time t can be defined to be the total value P_t of the portfolio divided by the total value at an earlier time $t-1$, that is

$$R_P(t) = \frac{P_t}{P_{t-1}} - 1$$

hence its simply the percentage change in the value from one time to another. Markowitz portfolio theory provides a method to analyse how good a given portfolio is only looking at the mean and the variance of the returns of the assets contained in the portfolio.

Notation. Defining Ω, a subset of $\mathbb{R}^d$, as the set of permissible portfolios, $\boldsymbol{\omega} \in \Omega$ means that the portfolio weights have to satisfy the constraints we impose upon our portfolio. We represent the expected returns of the securities by $\boldsymbol{\mu} = [\mu_1, ..., \mu_d]^\mathsf{T}$, where $\mu_i = E(R_i)$, $i = 1, \dots, d$. Therefore, since the expectation operator is linear, the expected return from the portfolio, with weights $\boldsymbol{\omega}$, as a whole is simply the weighted sum of the expected return from the individual assets

$$E(R_P) = \sum_{i=1}^{d} \omega_i \mu_i = \boldsymbol{\mu}^\mathsf{T}\boldsymbol{\omega} \tag{2.1}$$

We denote by σ_i the standard deviation of R_i, σ_{ij}; let ρ_{ij} be the correlation coefficient of the returns of assets S_i and S_j (for $i \neq j$), and Σ is the (symmetric) $[d\text{x}d]$ covariance matrix of the returns of all the assets:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \dots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \dots & \sigma_{nn} \end{pmatrix}$$

where $\sigma_{ii} = \sigma_i^2$ and $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ (for $i \neq j$). All valid covariance matrices are positive semidefinite matrices (namely that $\boldsymbol{\omega}^\mathsf{T}\Sigma\,\boldsymbol{\omega} \geqslant 0$, $\forall\boldsymbol{\omega}$), or equivalently, all of their eigenvalues are non-negative. However, we assume that Σ satisfies the stronger property of positive definiteness (that is $\boldsymbol{\omega}^\mathsf{T}\Sigma\,\boldsymbol{\omega} > 0$, $\forall\boldsymbol{\omega} \neq 0$.), and this is equivalent to say that none of the assets can be perfectly replicated by a combination of the remaining securities. This

assumption ensures that Σ is an invertible matrix.
For a given portfolio with weights $\boldsymbol{\omega}$, we can compute the variance and the standard deviation of the portfolio return as

$$Var_{R_P}(\boldsymbol{\omega}) = \boldsymbol{\omega}^{\mathsf{T}} \Sigma\, \boldsymbol{\omega} = \sum_{i=1}^{d} \sum_{j=1}^{d} \omega_i \omega_j \sigma_{i,j} \qquad \text{and} \qquad \sigma_{R_P}(\boldsymbol{\omega}) = \sqrt{\boldsymbol{\omega}^{\mathsf{T}} \Sigma\, \boldsymbol{\omega}} \tag{2.2}$$

The standard deviation of the portfolio return, also referred to as portfolio volatility, is frequently used as a risk measure of the portfolio with weights $\boldsymbol{\omega}$. The variance operator, is not linear and so the risk of a portfolio, as measured by the variance, is not equal to the weighted sum of risks of the individual assets.

This provides a way to quantify the benefit of diversification and shows that the relevant risk of an asset is not its stand-alone volatility, but rather its contribution to portfolio risk, and that this is primarily a question of its covariance with all the other securities in the portfolio. Henceforth, the concepts of risk and correlation became inseparable.

An investor is supposed to be risk-averse, hence he/she wishes for a low variance on return (i.e. small risk) and a high expected rate of return is also desired. The mean-variance optimization problem is the following:

$$\max_{\boldsymbol{\omega} \in \Omega} \; \boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{\omega} - \lambda\, \boldsymbol{\omega}^{\mathsf{T}} \Sigma\, \boldsymbol{\omega}$$

where λ is an investor specific risk aversion parameter that determines the trade-off between expected portfolio return and portfolio risk. The case $\lambda = 0$ corresponds to the investor only caring about getting a large expected return on the portfolio no matter what the risk is. An increasing λ corresponds to the investor becoming more willing to sacrifice part of the expected return to bear a lower risk (measured as lower portfolio variance). Starting with the first paper of Markowitz on the subject, most MVO formulations featured a variety of constraints.
One formulation we can see very often is the following:

$$\max_{\boldsymbol{\omega} \in \Omega} \; \boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{\omega} - \lambda\, \boldsymbol{\omega}^{\mathsf{T}} \Sigma\, \boldsymbol{\omega} \tag{2.3}$$

$$\textit{subject to} : \sum_{i=1}^{d} \omega_i = 1 \tag{2.4}$$

$$\omega_i \geqslant 0, i = 1, \ldots, d \tag{2.5}$$

The objective models the actual goal of the investor, a tradeoff between risk and reward. The constraint (2.4) means that the portfolio is fully invested and the condition (2.5), also known as the *long-only investing constraint*, is the same as saying that only long positions are allowed. Therefore, if one wants to include short sales in the model, it will be necessary to omit this condition.

Alternative formulations of the MVO problem are obtained by either maximizing the expected return subject to an upper limit on the portfolio variance, or by minimizing the portfolio variance subject to a lower limit on the expected return, that is keeping the constraints but substituting (2.3) with one of the following

$$\begin{aligned} &\max_{\boldsymbol{\omega}\in\Omega} \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\omega} && \min_{\boldsymbol{\omega}\in\Omega} \boldsymbol{\omega}^{\mathsf{T}}\Sigma\boldsymbol{\omega} \\ &\boldsymbol{\omega}^{\mathsf{T}}\Sigma\boldsymbol{\omega} \leq \sigma^2_{max} && \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\omega} \geqslant R_{min} \end{aligned} \tag{2.6}$$

Here the investor's goal is *split* between objective and reward condition. For different choices of $\boldsymbol{\omega}$ the investor will get different combinations of $\mu_P = E(R_P)$ (2.1) and $\sigma^2_P = Var_{R_P}$ (2.2). The set of all possible combinations (μ_P, σ^2_P) is called the *attainable set*. The set of optimal portfolios with minimum σ^2_P for a given or more R_{min} and maximum μ_P for a given σ^2_{max} or less, are called the efficient set (or efficient frontier); those are optimal in the sense that an investor cannot achieve a greater expected return, μ_P, without increasing his risk, σ_P. Any rational investor making decisions based only on the mean and variance of the distribution of returns of a portfolio, wants a high profit and a small risk, and would only choose to own portfolios on this efficient frontier. The specific portfolio he chooses depends on his level of risk aversion.

If the constraint set only includes linear equality and inequality constraints then the MVO is a quadratic program (QP) and can be solved by using standard numerical optimization software. Modern portfolio optimization software can also deal with non-linear constraints, such as risk limits or risk contribution limits on groups of securities, as well as constraints with discrete elements such as number of holdings and/or trades constraints. Such formulations are typically solved by using software that has conic optimization and integer optimization capabilities[7]. One of the appeals of MVO in quantitative portfolio construction is the flexibility this approach offers for incorporating multiple constraints reflecting client guidelines, regulatory restrictions, as well as the discretionary views of the portfolio manager.

2.3 Extreme value theory

The statistical analysis of extremes is key to many of the risk management problems related to insurance, reinsurance, and finance. EVT is a powerful and yet fairly robust framework for studying the tail behaviour of a distribution, and it is a canonical tool in the field of rare event estimation.

[7]See, for example, Fabozzi et al. (2007) for a discussion of numerical optimization software for solving portfolio optimization problems.

EVT provides two principal kinds of model for describing the extreme values of a succession of random variables. Specifically, there are two families of extreme value distributions corresponding to the two possible approaches for sampling extreme events. The first is *generalized extreme value* (GEV) *distribution* that is used in *Block Maxima approach* (which models the maxima of a set of blocks dividing the series) and the second is the *generalized Pareto distribution* (GPD) employed in the more modern group of *models for threshold exceedances* (considering as extreme any observation that exceeds some high level). The latter type of models are generally considered to be the most useful for practical applications. The block maxima method has the major defect that it is very wasteful of data; to perform the analyses one only the maximum losses in large blocks are retained. For this reason it has been largely superseded in practice by methods based on *threshold exceedances.*

The history of EVT dates back to early work by Fisher and Tippett (1928), Gnedenko (1943), who proved that the limiting distribution of normalized maxima of i.i.d. random variable (with cumulative distribution function F), belongs to one of the following non-degenerate distribution families: Weibull, Gumbel, Fréche. When modelling the maxima of a random variable, extreme value theory plays the same fundamental role as the central limit theorem plays when modelling sums of random variables. In both cases, the theory tells us what the limiting distributions are.

We consider a sequence of i.i.d. rvs $(X_i)_{i\in\mathbb{N}}$ with unknown underlying distribution function F. Suppose that block maxima $M_n = \max(X_1, \ldots, X_n)$ of i.i.d. rvs converge in distribution under an appropriate normalization, that is there exist sequences of real constants (d_n) and (c_n), where $c_n > 0$ for all n, such that

$$\lim_{n\to\infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = \lim_{n\to\infty} F^n(c_n x + d_n) = H(x) \tag{2.7}$$

for some non-degenerate cdf $H(x)$ and recalling that $P(M_n \leq x) = F^n(x)$. The role of the GEV in the study of maxima is formalized by the following definition and theorem.

Definition 2.1 (maximum domain of attraction). If 2.7 holds for some non degenerate cdf H, then F is said to be in the maximum domain of attraction of H, written $F \in MDA(H)$.

Theorem 2.1 (Fisher-Tippett, Gnedenko). *If $F \in MDA(H)$ for some non-degenerate cdf H then H must be a distribution of type H_ξ , i.e. a GEV distribution.*

Definition 2.2 (GEV). The cdf of the (standard) GEV distribution is given by

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}) & ,\xi \neq 0, \\ \exp(-e^{-x}) & ,\xi = 0, \end{cases} \tag{2.8}$$

where $1 + \xi x > 0$. A three-parameter family is obtained by defining $H_{\xi,\mu,\sigma}(x) = H_\xi((x - \mu/\sigma))$ for a *location* parameter $\mu \in \mathbb{R}$ and a *scale* parameter $\sigma > 0$.

The parameter ξ is known as the shape parameter of the GEV distribution and H_ξ defines a type of distribution, meaning a family of distributions specified up to location and scaling. The extreme value distribution in Definition 2.2 is generalized in the sense that the parametric form subsumes three types of distribution which are known by other names according to the value of ξ : when $\xi > 0$ the distribution is a Fréchet distribution; when $\xi = 0$ it is a Gumbel distribution; when $\xi < 0$ it is a Weibull distribution.

Considering the broad variety techniques that EVT offers, in this section we will focus on the *threshold exceedances* approach, as the method we employ to tackle our portfolio allocation problem. In particular, we present in detail the parametric GPD approach. However, this is not the only way available to estimate the tail of a distribution. Indeed, within the *threshold exceedances* class of models, one may further distinguish two styles of analysis. There are the semi-parametric models built around the Hill estimator and its relatives (Beirlant et al., 1996; Danielsson and de Vries, 1997a; Danielsson et al., 1998) and the fully parametric models based on the generalized Pareto distribution or GPD (Embrechts et al., 1998, 1999). Both are theoretically justified and empirically useful when used correctly. We favour the latter style of analysis for reasons of simplicity, both of exposition and implementation. One obtains simple parametric formulae for measures of extreme risk for which it is relatively easy to give estimates of statistical error using the techniques of maximum likelihood inference[8].

EVT-based methods have two features which make them attractive for tail estimation: they are based on a sound statistical theory and they offer a parametric form for the tail of a distribution. Hence, these methods allow for some extrapolation beyond the range of the data, even if care is required at this point.

From a mathematical point of view, the *threshold exceedances* method is based on Balkema and De Haan (1974) and Pickands (1975). The statistical theory was initially worked out by Davison (1984) and Davison and Smith (1990).

2.3.1 General Theory: Threshold Exceedances

Let $X_1, X_2, \ldots$ be identically distributed random variables with unknown underlying distribution function $F(x) = P(X_i \leq x)$ (we work with distribution functions and not densities) representing financial losses. These may have a variety of interpretations, such as operational losses, insurance losses and losses on a credit portfolio over fixed time intervals. Moreover, they might represent risks which we can directly observe or they might

[8] A list of pros of using the GPD approach instead of the Hill estimator is provided by McNeil and Frey (2000), including greater stability with respect to the choice of the cut-off and the possibility of modelling any kind of extreme value distribution, not only the heavy-tailed ones.

also represent risks which we are forced to simulate in some Monte Carlo procedure, because of the impracticality of obtaining data.

Generalized Pareto Distribution. The main distributional model for exceedances over thresholds is the generalized Pareto distribution (GPD).

Definition 2.3 (GPD). The GPD is a two parameter distribution with distribution function

$$G_{\xi,\beta}(x) = \begin{cases} 1-(1+\xi x/\beta)^{-1/\xi} & ,\xi \neq 0, \\ 1-exp(-x/\beta) & ,\xi = 0, \end{cases} \tag{2.9}$$

where $\beta > 0$, and $x \geqslant 0$ when $\xi \geqslant 0$ and $0 \leqslant x \leqslant -\beta/\xi$ when $\xi < 0$. The parameters ξ and β are referred to, respectively, as the *shape* and *scale* parameters.

This distribution is generalized in the sense that it subsumes certain other distributions under a common parametric form. ξ is the important shape parameter of the distribution and β is an additional scaling parameter.

If $\xi > 0$ then $G_{\xi,\beta}(x)$ is a reparametrized version of the ordinary Pareto distribution, which has a long history in actuarial mathematics as a model for large losses; $\xi = 0$ corresponds to the exponential distribution and $\xi < 0$ is known as a Pareto type II distribution (short tailed). The first case is the most relevant for risk management purposes since the GPD is heavy-tailed when $\xi > 0$. The cdf and pdf of the GPD for various values of ξ and $\beta = 1$ are shown in Figure 2.1.

Whereas the normal distribution has moments of all orders, a heavy-tailed distribution does not possess a complete set of moments. In the heavy-tailed case, GPD with $\xi > 0$, one can find that $E[X^k] = \infty$ for $k \geqslant 1/\xi$[9]. For example, when $\xi = 1/2$, the GPD is an infinite variance (second moment) distribution; when $\xi = 1/4$, the GPD has an infinite fourth moment. The mean of the GPD is defined provided $\xi < 1$ and is

$$E(X) = \beta/(1-\xi). \tag{2.10}$$

Excess Distributions. The role of the GPD in EVT is as a natural model for the excess distribution over a high threshold. We define this concept along with the mean excess function, which will also play an important role in the theory.

Definition 2.4 (excess distribution over threshold u). Let X be an rv with cdf F. The excess distribution (also known as distribution of *excesses losses*) over the threshold u has cdf

$$F_u(x) = P(X-u \leq x \mid X > u) = \frac{F(x+u)-F(u)}{1-F(u)} \tag{2.11}$$

for $0 \leqslant x < x_F - u$, where $x_F \leqslant \infty$ is the right endpoint of F.

[9]This is also discussed in Section 3.3.2 Chapter 3 as a consequence of Theorem 3.1 by Gnedenko (1943).

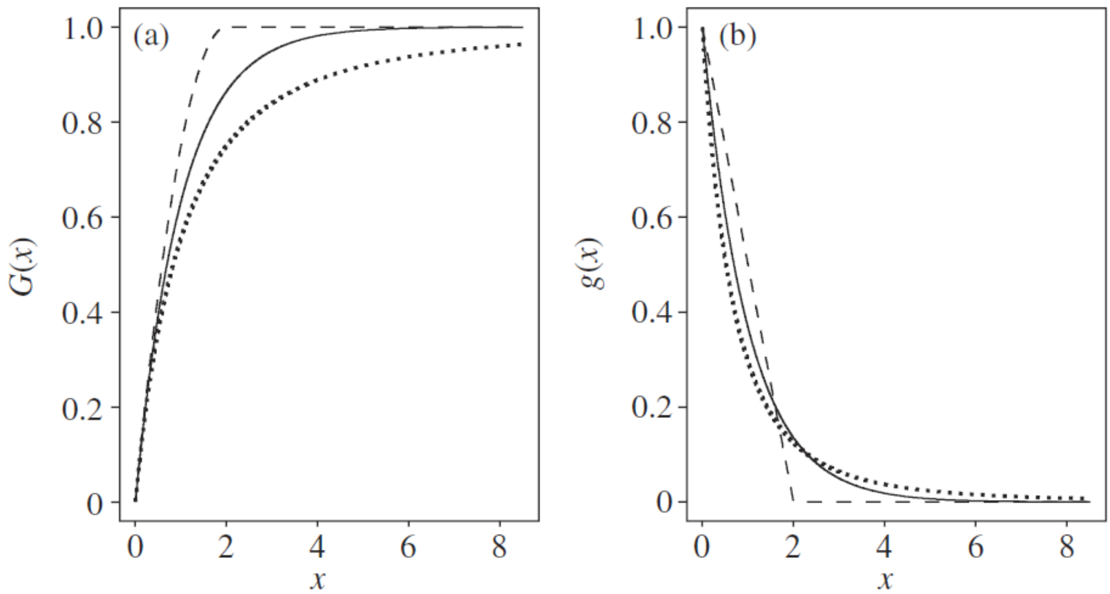


Figure 2.1: (a) Distribution function of GPD in three cases: the solid line corresponds to $\xi = 0$ (exponential); the dotted line to $\xi = 0.5$ (a Pareto distribution); and the dashed line to $\xi = -0.5$ (Pareto type II). The scale parameter β is equal to 1 in all cases. (b) Corresponding densities.

The excess cdf F_u describes the distribution of the *excess loss* over the threshold u, given that u is exceeded. It is very useful to observe that it can be written in terms of the underlying F. Mostly we would assume our underlying F is a distribution with an infinite right endpoint, i.e. it allows the possibility of arbitrarily large losses, even if it attributes negligible probability to unreasonably large outcomes, e.g. the normal or t-distributions. But it is also conceivable, in certain applications, that F could have a finite right endpoint. An example is the beta distribution on the interval $[0; 1]$ which attributes zero probability to outcomes larger than 1 and which might be used, for example, as the distribution of credit losses expressed as a proportion of exposure.
In survival analysis the excess cdf is more commonly known as the *residual life cdf*; it expresses the probability that, say, an electrical component which has functioned for u units of time fails in the time period $(u, u + x]$.

Definition 2.5 (mean excess function). The mean excess function of an rv X with finite mean is given by

$$e(u) = E(X - u \mid X > u). \tag{2.12}$$

The mean excess function $e(u)$ expresses the mean of F_u as a function of u. The mean excess function is known as the *mean residual life function* and gives the expected residual lifetime for components with different ages. For the special case of the GPD, the excess cdf and mean excess function are easily calculated.

Example 2.1 (excess distribution of exponential and GPD). If F is the cdf of an exponential rv, then it is easily verified that $F_u(x) = F(x)$ for all x, which is the famous *lack-of-memory* property of the exponential distribution, namely the residual lifetime of the aforementioned electrical component would be independent of the amount of time that component has already survived. More generally, if X has cdf $F = G_{\xi,\beta}$, then, using (2.11), the excess cdf is easily calculated to be

$$F_u(x) = G_{\xi,\beta(u)}(x), \ \beta(u) = \beta + \xi u, \tag{2.13}$$

where $0 \leqslant x < \infty$ if $\xi \geqslant 0$ and $0 \leqslant x \leqslant -(\beta/\xi) - u$ if $\xi < 0$. The excess distribution remains a GPD with the same shape parameter ξ but with a scaling that grows linearly with the threshold u. The mean excess function of the GPD is easily calculated from (2.13) and (2.10) to be

$$e(u) = \frac{\beta(u)}{1-\xi} = \frac{\beta + \xi u}{1-\xi}, \tag{2.14}$$

where $0 \leqslant u < \infty$ if $0 \leqslant \xi < 1$ and $0 \leqslant u \leqslant -\beta/\xi$ if $\xi < 0$. It may be observed that the mean excess function is *linear in the threshold* u, which is a characterizing property of the GPD.

Example 2.1 shows that the GPD has a kind of stability property under the operation of calculating excess distributions. We now give a mathematical result that shows that the GPD is, in fact, a natural limiting excess distribution for many underlying loss distributions. The following limit theorem is a key result in EVT and explains the importance of the GPD.

Theorem 2.2 (Pickands (1975)-Balkema and De Haan (1974)). *We can find a (positive-measurable) function $\beta(u)$ such that*

$$\lim_{u \to x_F} \sup_{0 \leqslant x < x_F - u} \mid F_u(x) - G_{\xi,\beta(u)}(x) \mid = 0, \tag{2.15}$$

if and only if $F \in MDA(H_\xi)$[10], $\xi \in \mathbb{R}$.

That is, for a large class of underlying distributions F (in particular those F for which normalized maxima converge to a GEV distribution[11]), as the threshold u is progressively raised, the excess distribution F_u converges to a generalized Pareto; moreover, the shape parameter of the limiting GPD for the excesses is the same as the shape parameter of the limiting GEV distribution for the maxima. Essentially all the commonly used continuous distributions of statistics are in $MDA(H_\xi)$ for some ξ, so Theorem 2.2 proves to be a very

[10] This notation means that F is in the *maximum domain of attraction* of H_ξ.

[11] The class contains all the common continuous distributions of statistics and actuarial science (normal, lognormal, χ, t, F, gamma, exponential, uniform, beta, etc.).

widely applicable result that essentially says that the GPD is the *canonical distribution* for modelling excess losses over high thresholds.

2.4 Modelling Excess Losses

Our model for a generic risk X_i having distribution F assumes that, for a certain u, the excess distribution above this threshold may be taken to be exactly GPD for some ξ and β. We exploit Theorem 2.2 by assuming that we are dealing with a loss distribution $F \in MDA(H_\xi)$ so that, for some suitably chosen high threshold u, we can model F_u by a generalized Pareto distribution. We formalise this with the following assumption.

Assumption 2.1. *Let F be a loss distribution with right endpoint x_F and assume that for some high threshold u we have*

$$F_u(x) = G_{\xi,\beta}(x)$$

for $0 \leqslant x < x_F - u$ and some $\xi \in \mathbb{R}$ and $\beta > 0$.

This is clearly an approximation based on the asymptotic EVT results, since in practice the excess distribution will generally not be exactly GPD.

2.4.1 The method: GPD fit

Assuming we have realisations of our loss rvs we use statistics to make the model more precise by choosing a sensible u and estimating ξ and β. Given loss data $X_1, \ldots, X_n$ from F, a random number N_u will exceed our threshold u; it will be convenient to relabel these data $\tilde{X}_1, \ldots, \tilde{X}_{N_u}$. For each of these exceedances we calculate the amount $Y_j = \tilde{X}_j - u$ of the excess loss. We wish to estimate the parameters of a GPD model by fitting this distribution to the N_u excess losses. There are various ways of fitting the GPD including maximum likelihood (ML) and probability-weighted moments (PWM)[12]. The former method is more commonly used and is easy to implement if the excess data can be assumed to be realizations of independent rvs, since the joint density will then be a product of marginal GPD densities.

Writing $g_{\xi,\beta}$ for the density of the GPD, the log-likelihood may be easily calculated to be

$$\ln L(\xi, \beta; Y_1, ..., Y_{N_u}) = \sum_{j=1}^{N_u} \ln g_{\xi,\beta}(Y_j) = -N_u \ln \beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \ln \left(1 + \xi \frac{Y_j}{\beta}\right), \quad (2.16)$$

[12] A method proposed by Hosking et al. (1985).

which must be maximized subject to the parameter constraints that $\beta > 0$ and $1+\xi\frac{Y_j}{\beta} > 0$ for all j . Solving the maximization problem yields a GPD model $G_{\hat{\xi},\hat{\beta}}(x)$ for the excess distribution F_u.

Remark 2.1 (Non-i.i.d. data.). For insurance or operational risk data the i.i.d. assumption is often unproblematic, but this is clearly not true for time series of financial returns. If the data are serially dependent but show no tendency to give clusters of extreme values, then this might suggest that the underlying process has extremal index $\theta = 1$. In this case, asymptotic theory (see Section 7.4 in McNeil et al. (2005)) suggests a limiting model for high-level *threshold exceedances*, in which exceedances occur according to a Poisson process and the excess loss amounts are i.i.d. generalized Pareto distributed. If extremal clustering is present, suggesting an extremal index $\theta < 1$ (as would be consistent with an underlying GARCH process), the assumption of independent excess losses is less satisfactory. The easiest approach is to neglect this problem and to consider the ML method to be a *quasi-maximum likelihood (QML) method*, where the likelihood is misspecified with respect to the serial dependence structure of the data. We follow this course in this thesis. First of all because the i.i.d. assumption is a convenient starting point in order to develop our methodology. Second, it may not affect our conclusions since, the point estimates will still be reasonable (unbiased estimators); although standard errors may be too small Kearns and Pagan (1997). However, the latter do not directly impact the portfolio allocation problem as we will see in detail later on.

2.5 Modelling Tails and Measures of Tail Risk

In this section we describe how the GPD model for the excess losses is used to model the tail of the underlying loss distribution F and to estimate associated risk measures. Wherever the tail of a loss distribution is of interest, whether for market, credit, operational or insurance risks, the *threshold exceedances* method provides a simple tool for modelling and estimating measures of tail risk.

2.5.1 Tail distribution

To make the necessary theoretical calculations we again make Assumption 2.1.
In view of (2.11) we have for $0 \leqslant x < x_F - u$, that $F(x+u) = F(u) + (1 - F(u))F_u(x)$.
By setting $y = u + x$ and being $\bar{F}(y) = 1 - F(y) = P(Y > y), \forall y \geqslant u$ we get

$$\begin{aligned} F(y) &= F(u) + (1 - F(u))F_u(y - u) \\ &= 1 + \bar{F}(u)[F_u(y - u) - 1],\ u \leqslant y < x_F \end{aligned}$$

This formula shows that we may move easily to an interpretation of the model in terms of the tail of the underlying distribution $F(y)$ for $y \geqslant u$. Indeed, continuing from the previous expression and using (2.9), we obtain

$$\begin{aligned}\bar{F}(y) &= \bar{F}(u)[1 - F_u(y-u)] \\ &= \bar{F}(u)\bar{F}_u(y-u) = \bar{F}(u)[1 - G_{\xi,\beta}(y-u)] \qquad (2.17) \\ &= \bar{F}(u)\left(1 + \xi\frac{y-u}{\beta}\right)^{-1/\xi}, \; u \leqslant y < x_F \qquad (2.18)\end{aligned}$$

The expression (2.18), knowing $\bar{F}(u)$, gives us a formula for the tail probabilities under the parametric GPD method. We would have obtained the tail decomposition (2.17) also observing that, for $x \geqslant u$[13]

$$\begin{aligned}P(X > x) &= P(X > u)P(X > x \mid X > u) \\ &= P(X > u)P(X - u > x - u \mid X > u) \\ &= \bar{F}(u)\bar{F}_u(x-u).\end{aligned}$$

2.5.2 Value at Risk

Mathematically we define our measures of extreme risk in terms of the loss distribution F. Let the confidence level of our risk measures be $0.95 \leqslant \alpha < 1$. Value-at-Risk (VaR) is the α^{th}quantile of the distribution F, that is

$$VaR_\alpha = F^{-1}(\alpha)$$

where F^{-1} is the inverse of F. For a given probability $\alpha > F(u)$ the VaR estimate is calculated by conveniently inverting the tail formula (2.18). Solving the following

$$\alpha = F(VaR_\alpha) = 1 - \bar{F}(VaR_\alpha) = 1 - \bar{F}(u)\left(1 + \xi\frac{VaR_\alpha - u}{\beta}\right)^{-1/\xi}$$

we get a parametric expression for the quantile based risk measure, that is

$$VaR_\alpha = u + \frac{\beta}{\xi}\left(\left(\frac{1}{\bar{F}(u)}(1-\alpha)\right)^{-\xi} - 1\right) \qquad (2.19)$$

2.5.3 Expected shortfall

Expected shortfall at a confidence level α is the expected loss size, given that VaR_α is exceeded.

[13]Kolmogorov defines conditional probabilities as $P(A|B) = P(A \cap B)/P(B), P(B) > 0$. Therefore, $P(X > x \mid X > u) = P(X > x)/P(X > u), x \geqslant u$.

$$ES_\alpha = \mathbb{E}[X \mid X > VaR_\alpha]$$

It may also be obtained by using the model of the tail of the distribution given in Eq. (2.18). Indeed, for the linearity of expectation and recalling the mean excess function expression presented in Definition 2.5, we have

$$\begin{aligned} ES_\alpha &= \mathbb{E}[X - VaR_\alpha + VaR_\alpha \mid X > VaR_\alpha] \\ &= \mathbb{E}[X - VaR_\alpha \mid X > VaR_\alpha] + \mathbb{E}[VaR_\alpha \mid X > VaR_\alpha] \\ &= \mathbb{E}[X - VaR_\alpha \mid X > VaR_\alpha] + VaR_\alpha \\ &= e(VaR_\alpha) + VaR_\alpha \end{aligned} \tag{2.20}$$

Suppose the random variable Y has a generalized Pareto distribution with shape parameter $\xi < 1$, this condition is necessary for the expected value to exist, and scale parameter β. Then for $s < x_F$ the mean excess function is linear in the threshold s, as we showed in the Example 2.1 Eq. (2.14) and is given by $e(u) = \dfrac{\beta + \xi s}{1 - \xi}, \beta + \xi s > 0$. Therefore, using this expression, the Definition of mean excess function 2.5, and the following relation

$$e_Y(s) = e_{Y-u}(s - u)$$[14]

we obtain a parametric expression for the Expected Shotfall starting from (2.20)

$$\begin{aligned} ES_\alpha &= e_{Y-u}(VaR_\alpha - u) + VaR_\alpha = \frac{\beta + \xi(VaR_\alpha - u)}{1 - \xi} + VaR_\alpha \\ &= \frac{VaR_\alpha}{1 - \xi} + \frac{\beta - u\xi}{1 - \xi} \end{aligned} \tag{2.21}$$

2.5.4 Estimation in practice

We note that, under Assumption 2.1, tail probabilities, VaRs and expected shortfalls are all given by formulas of the form $g(\xi, \beta, \bar{F}(u))$. Assuming that we have fitted a GPD to excess losses over a threshold u, as described in Section 2.4.1, we estimate these quantities by first replacing ξ and β in formulas (2.18),(2.19),(2.21) by their estimates $\hat{\xi}$, $\hat{\beta}$. Of course, we also require an estimate of $\bar{F}(u)$. For this purpose we take the obvious *empirical estimator*

$$\widehat{\bar{F}}(u) = \frac{N_u}{n}, \text{ where } N_u = \textstyle\sum_{i=1}^{n} \mathbb{1}_{\{X_i > u\}}$$

and n is the total number of observations. That is, we use the method of *historical simulation* (HS). In doing this, we are implicitly assuming that there is a sufficient proportion of sample values above the threshold u to estimate $\bar{F}(u)$ reliably. An immediate question is, why do we not use the HS method to estimate the whole tail of $F(x)$ (i.e. for all $x \geqslant u$)?

[14] $e_{Y-u}(s-u) = \mathbb{E}(Y - u - (s - u) \mid Y - u > (s - u)) = \mathbb{E}(Y - s \mid Y > s) = e_Y(s)$.

This is because historical simulation is a poor method in the tail of the distribution where data become sparse. In setting a threshold at u we are judging that we have sufficient observations exceeding u to enable a reasonable HS estimate of $F(u)$, but for higher levels the historical method would be too unreliable. We hope to gain over the empirical method by using a kind of extrapolation based on the GPD for more extreme tail probabilities and risk measures. For tail probabilities we obtain an estimator, first proposed by Smith (1987), of the form

$$\widehat{\bar{F}}(y) = \frac{N_u}{n}\left(1 + \hat{\xi}\frac{y-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}, \; y \geqslant u \tag{2.22}$$

It is important to observe that this estimator is only valid for $y \geqslant u$. This estimate can be viewed as a kind of HS estimate augmented by EVT and it can be constructed whenever we believe data come from a common distribution, although its statistical properties are best understood in the situation when the data may also be assumed independent or only weakly dependent. For confidence levels $\alpha \geqslant 1 - N_u/n$ we obtain analogous point estimators of VaR_α and ES_α

$$\widehat{VaR}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-\alpha)\right)^{-\hat{\xi}} - 1\right) \tag{2.23}$$

$$\widehat{ES}_\alpha = \frac{\widehat{VaR}_\alpha}{1-\hat{\xi}} + \frac{\hat{\beta} - u\hat{\xi}}{1-\hat{\xi}} \tag{2.24}$$

2.6 Optimisation problem using EVT

We study the optimal portfolio selection problem for risk averse investors concerned with extreme market movements and aiming to ameliorate the risk of suffering from large losses caused by those rare events.

In this setting, we find natural building our model relying on Roy's safety first approach (Roy, 1952) instead of the classical MVO reviewed in Section 2.2. Indeed, while Markowitz's typical investor is concerned about maximizing the portfolio expected return controlling the portfolio variability, safety first investors are rather interested in avoiding extreme losses through minimizing the disaster probability.

We refer to the stream of literature that remove the mean constraint, while keeping the disaster probability as the objective to be minimized. In particular, our model proceeds to the risk minimisation expressing portfolio downside risk in terms of the two risk measures: Value-at-Risk (VaR) and Expected Shortfall (ES). We relax the normality assumption that characterised the MPV, and apply the tools of univariate extreme value theory to the asset allocation problem.

2.6.1 Portfolio Loss distribution

In order to precisely formalise our model, we need to introduce first some preliminary notation and define the random variable representing the loss of the portfolio.

Indeed, this is the core loss distribution on which we will be able to apply the GPD model for the excess losses to describe its tails and to estimate associated risk measures.

We represent the uncertainty about future states of the world by a probability space $(\Omega, \mathcal{F}, P)$, which is the domain of all random variables (rvs) we introduce below. Consider an investment universe composed of a number d of available assets, $S_1, S_2, \ldots, S_d$, each with a well defined value[15] $S_{i,t}, i = 1, \ldots, d$ at any time t .
Viewed from today, the value of each asset one time period in the future is a random variable $S_{i,t+1}, , i \in \{1, \ldots, d\}$[16]. The price process of asset i is denoted by $(S_{i,t})_{t\in\mathbb{N}}$.
In (market) risk management one is interested in the (positive) tail of the so-called Profit-and-Loss (P&L) distribution function (cdf) of the random variable (rv) of interest[17].

Therefore, let the rvs $X_{1,t+1}, X_{2,t+1}, \ldots, X_{d,t+1}$ represent financial losses arising from those financial positions over the period $[t, t+1]$, i.e. the *negative (simple) returns* on the investment in each of the single securities in our universe over the next unitary period. $\boldsymbol{X}_{t+1} \in \mathbb{R}^d$ is the correspondent vector, where each component is defined as

$$X_{i,t+1} = -(S_{i,t+1} - S_{i,t})/S_{i,t}, i = 1, \ldots, d$$

Following standard practice in finance and risk management we use *logarithmic prices* as *risk factors*[18]. $\boldsymbol{Z}_t = \log \boldsymbol{S}_t \in \mathbb{R}^d$ is a random vector of *risk factors*, where

$$\boldsymbol{Z}_t = (Z_{1,t}, \ldots, Z_{d,t})^\mathsf{T},\ Z_{i,t} = \log S_{i,t}\ i \in \{1, \ldots, d\} \tag{2.25}$$

Risk factors are observable, because we already assumed that asset prices are observable as well, thus $\boldsymbol{Z}_t$ is known at time t[19]. It will be convenient to define the series of *risk-factor changes* $(\boldsymbol{Y}_t)_{t\in\mathbb{N}}$ by $\boldsymbol{Y}_t := \boldsymbol{Z}_t - \boldsymbol{Z}_{t-1}$. These are the objects of interest in most statistical studies of financial time series. Assuming (2.25), we find that, the *log-returns* of the securities in the universe corresponds to the *risk-factor changes*

$$Y_{i,t+1} = Z_{i,t+1} - Z_{i,t} = \log S_{i,t+1} - \log S_{i,t} = \log(S_{i,t+1}/S_{i,t}), i = 1, \ldots, d$$

[15] We will not enter into a more detailed economic and/or actuarial discussion of the precise meaning of value in a more general case: it suffices to mention possibilities like "fair value", "market value" (mark-to-market), "model value" (mark-to-model), "accounting value" (often depending on the jurisdiction under which industry is regulated) and indeed various combinations of the above.

[16] The rvs $S_{i,t}, i = 1, \ldots, d$ are assumed to be observable so that $\boldsymbol{S}_t$ is known at time t. In the same way, $\boldsymbol{S}_{t+1}$ is known at time $t+1$ but it is random from the viewpoint of time t.

[17] Clearly in risk management we are concerned with possible high losses. Thus, the P&L distribution is actually intended as "loss distribution"; its upper (right) tail represents large positive losses.

[18] Frequently used risk factors are logarithmic prices of financial assets, yields and logarithmic exchange rates.

[19] Risk factors are usually assumed to be observable so that $\boldsymbol{Z}_t$ is known at time t.

Now, consider a linear portfolio $\mathcal{P}$ of the d securities and denote by λ_i the number of shares of asset $S_{i,t}, i \in \{1,\ldots,d\}$ in the portfolio $\mathcal{P}$ at time t[20]. We denote the *value* of this portfolio at time t by V_t and assume that the rv is observable at time t.
Following standard risk-management practice, the value of the portfolio $\mathcal{P}$ is modelled as a function of time and the d-dimensional random vector $\boldsymbol{Z}_t$ of risk factors, i.e. we have the representation

$$V_t = f(t, \boldsymbol{Z}_t) \tag{2.26}$$

for some measurable function $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. A representation of the portfolio value in this form (2.26) is termed a *mapping of risks*.
Considering logarithmic prices of financial assets as risk factors as we did in (2.25) and in view of (2.26), the value of the linear portfolio $\mathcal{P}$ in $t+1$ will be a linear combination of the individual asset prices with weights λ_i; i.e we have the following representation

$$V_{t+1} = f(t+1, \boldsymbol{Z}_{t+1}) = \sum_{i=1}^{d} \lambda_i S_{i,t+1} = \sum_{i=1}^{d} \lambda_i \exp(Z_{i,t+1}) \tag{2.27}$$

The *loss* of the portfolio over the period $[t, t+1]$ is given by

$$L_{[t,t+1]} = L_{t+1} = -(V_{t+1} - V_t)$$

Using the mapping (2.27) the portfolio loss can be written as

$$\begin{aligned} L_{[t,t+1]} = L_{t+1} &= -(f(t+1, \boldsymbol{Z}_{t+1}) - f(t, \boldsymbol{Z}_t)) \\ &= -(f(t+1, \boldsymbol{Z}_t + \boldsymbol{Y}_{t+1}) - f(t, \boldsymbol{Z}_t)) \end{aligned}$$

The distribution of $L_{[t,t+1]}$ is termed the *loss distribution*. We distinguish between the *conditional loss distribution*, i.e. the distribution of L_{t+1} given all available information up to and including time t, and the *unconditional loss distribution*; this issue is taken up in more detail in Remark 2.5. Since $\boldsymbol{Z}_t$ is known at time t, the loss distribution is determined by the distribution of the risk-factor change $\boldsymbol{Y}_{t+1}$. We therefore introduce another piece of notation, namely the *loss operator* $l_{[t]} : \mathbb{R}^d \to \mathbb{R}$, which maps risk-factor changes into losses. It is defined by

$$l_{[t]}(\boldsymbol{x}) := -(f(t+1, \boldsymbol{Z}_t + \boldsymbol{x}) - f(t, \boldsymbol{Z}_t)), \boldsymbol{x} \in \mathbb{R}^d, \tag{2.28}$$

and we obviously have $L_{t+1} = l_{[t]}(\boldsymbol{Y}_{t+1})$. We are interested in finding the expression for

[20] For instance, if we were to consider portfolios of default-free zerocoupon bonds with maturity T_i and price $p(t, T_i)$, λ_i would have denoted the number of bonds with maturity T_i in the portfolio. In a detailed analysis of the change in value of a bond portfolio one takes all yields $y(t, T_i)$, as risk factors. See Section 2.1.3, Example 2.6 in McNeil et al. (2005).

the *loss* of the linear portfolio $\mathcal{P}$, that is in view of (2.27)

$$
\begin{aligned}
L_{t+1} &= -(f(t+1, \boldsymbol{Z}_{t+1}) - f(t, \boldsymbol{Z}_t)) && (2.29)\\
&= -\sum_{i=1}^{d} \lambda_i(\exp(Z_{i,t+1}) - \exp(Z_{i,t}))\\
&= -\sum_{i=1}^{d} \lambda_i(\exp(Z_{i,t} + Y_{i,t+1}) - \exp(Z_{i,t}))\\
&= -\sum_{i=1}^{d} \tilde{\omega}_i(\exp(Y_{i,t+1}) - 1) = -\sum_{i=1}^{d} \tilde{\omega}_i R_{i,t+1}\\
&= \sum_{i=1}^{d} \tilde{\omega}_i X_{i,t+1} && (2.30)
\end{aligned}
$$

where $\tilde{\omega}_i = \lambda_i \exp(Z_{i,t}) = \lambda_i S_{i,t}$ and $R_{i,t+1} = \exp(Y_{i,t+1}) - 1$ is the *simple (or arithmetic) return* rv for each security. For our purposes, we prefer to express the portfolio loss in terms of relative change instead of actual difference. Therefore, we define our random variable of interest starting from (2.30), in the following Definition.

Definition 2.6 (**Loss of a linear portfolio** $\mathcal{P}$)**.** Consider an investment universe of d available assets, each of those have a well defined value $S_{i,t}, i = 1, \ldots, d$ at any time t .

Let the vector of rvs $\boldsymbol{X}_{t+1} = (X_{1,t+1}, X_{2,t+1}, \ldots, X_{d,t+1})^\intercal$ represent the set of *negative (simple) returns* (i.e. losses) from the investment in each individual securities within our universe. Consider a fixed portfolio of the d securities and denote by λ_i the number of shares of asset i one hold in the portfolio at time t. Then, the *rv loss of the linear portfolio* $\mathcal{P}$, with allocation $\boldsymbol{w}_t = (w_{1,t}, w_{2,t}, \ldots, w_{d,t})^\intercal \in \mathbb{R}^d$, is

$$
L_{t+1}(\boldsymbol{w}) = -(V_{t+1} - V_t)/V_t = \sum_{i=1}^{d} \omega_i X_{i,t+1} \quad , 0 \leqslant wi \leqslant 1 \ , i = 1, \ldots, d, \tag{2.31}
$$

simply a linear combination of the loss from the d assets, each with weight $\omega_i = \tilde{\omega}_i / V_t = (\lambda_i S_{i,t})/V_t$ given by the proportion of the portfolio value invested in security i at time t.

Based on models and statistical estimates (computed from data) for $\boldsymbol{X}_{t+1}$, the problem is now clear: find the cdf (or some synthetic characteristic, as a risk measure) of our portfolio loss random variable L_{t+1}[21]. This rv depends also on the chosen allocation among the assets available in our universe, as identified by the vector of weights $\boldsymbol{w}$.

Remark 2.2 (Unchanged allocation)**.** Note that our definition of the portfolio loss implicitly assumes that the composition of the portfolio remains unchanged over the uni-

[21] In general for easy of notation we omit the dependence of the loss random variable on the vector $\boldsymbol{X}_{t+1}$.

tary time horizon. While unproblematic for daily losses this assumption becomes increasingly unrealistic for longer time horizons.

Remark 2.3 (Time Notation)**.** In developing the formulas we have implicitly considered a fixed horizon Δ and assumed that time is measured in units of this horizon Δ. We can therefore introduce a time series notation, moving from a generic process $Y(s)$ to the time series $(Y_t)t \in \mathbb{N}$ with $Y_t := Y(t\Delta)$. The loss of the portfolio is then written as

$$L_{t+1} := L_{[t\Delta,(t+1)\Delta]} = -(V_{t+1} - V_t).$$

For instance, in market risk management we often work with financial models where the calendar time s is measured in years and interest rates and volatilities are quoted on an annualized basis. If we are interested in daily losses we set $\Delta = 1/365$ or $\Delta \approx 1/250$; the latter convention is mainly used in markets for equity derivatives since there are approximately 250 trading days per year. The rvs V_t and V_{t+1} then represent the portfolio value on days t and $t+1$, respectively, and L_{t+1} is the loss from day t to day $t+1$.

Remark 2.4 (Aggregation of risk factors)**.** Considering our particular choice of *risk factors*, and thus *risk factor changes* (2.25), and that $\mathcal{P}$ linearly depends on those underlying *risk factors*, the structure of the portfolio value (2.27) we find is simple. Therefore, also the loss of the linear portfolio has immediate and clear interpretation (Equations (2.30) and (2.31)). However, this approach could be generalised to cases where portfolios $\mathcal{P}$ of financial positions (such as a collection of stocks or bonds, a book of derivatives, a collection of risky loans or even a financial institution's overall position in risky assets) might be more complicated. Thus, function (2.27) can be highly non-linear. For instance, this especially holds true in cases where derivative financial instruments (like options) are involved. In this situation, the standard procedure is, if f is differentiable, to consider a first-order approximation of the loss in (2.29) of the form

$$L_{t+1} = -(f(t+1, \boldsymbol{Z}_t + \mathbf{Y}_{t+1}) - f(t, \boldsymbol{Z}_t)) \approx L^{\Delta}_{t+1} = -\left(f_t(t, \boldsymbol{Z}_t) + \textstyle\sum_{i=1}^{d} f_{z_i}(t, \boldsymbol{Z}_t) Y_{i,t+1} \right)$$

where the subscripts to f denote partial derivatives. The first-order approximation is convenient as it allows us to represent the loss as a linear function of the risk-factor changes. The quality of the approximation is higher if the portfolio value is almost linear in the risk factors (i.e. if the function f has small second derivatives) and if Y_{t+1} is stochastically small. The latter implies that one-period changes in the risk factors are small; this is acceptable if we are measuring risk over a short horizon and we are in "normal" periods but not in "extreme" periods. Note that it is especially for the latter that QRM is needed!

Remark 2.5 (Unconditional and Conditional loss distribution)**.** Here, "the cdf" can be interpreted unconditionally or conditionally on some family of σ-algebras (a filtration of historical information). In risk management we often have to decide if we are interested in the conditional or the unconditional distribution of losses. Both are relevant for risk-management purposes, but it is important to be aware of the distinction between the two concepts. The differences between conditional and unconditional loss distributions are strongly related to time series properties of the series of risk-factor changes $(\boldsymbol{Y}_t)_{t\in\mathbb{N}}$. Suppose that the risk-factor changes form a *stationary time series* with stationary distribution $F_{\boldsymbol{Y}}$ on $\mathbb{R}^d$. Essentially, this means that the distribution of $(\boldsymbol{Y}_t)_{t\in\mathbb{N}}$ is invariant under shifts of time and most time series models used in practice for the modelling of risk-factor changes satisfy this property. Now fix a point in time t (current time), and denote by $\mathcal{F}_t$ the sigma field representing the publicly available information at time t. Typically, $\mathcal{F}_t = \sigma(\{\boldsymbol{Y}_s : s \leqslant t\})$, the sigma field generated by past and present risk-factor changes, often called the history, up to and including time t. Denote by $F_{\boldsymbol{Y}_{t+1}|\mathcal{F}_t}$ the conditional distribution of $\boldsymbol{Y}_{t+1}$ given current information $\mathcal{F}_t$. In most stationary time series models relevant for risk management, $F_{\boldsymbol{Y}_{t+1}|\mathcal{F}_t}$ is not equal to the stationary distribution $F_{\boldsymbol{Y}}$. An important example is provided by the popular models from the GARCH family. In this class of model the variance of the conditional distribution of $\boldsymbol{Y}_{t+1}$ is a function of past risk-factor changes and possibly of its own lagged values. On the other hand, if $(\boldsymbol{Y}_t)_{t\in\mathbb{N}}$ is an independent and identically distributed (i.i.d.) series, we obviously have $F_{\boldsymbol{Y}_{t+1}|\mathcal{F}_t} = F_{\boldsymbol{Y}}$. The *conditional loss distribution* $F_{L_{t+1}|\mathcal{F}_t}(l)$ is defined as the distribution under $F_{\boldsymbol{Y}_{t+1}|\mathcal{F}_t}$. Formally we have,

$$F_{L_{t+1}|\mathcal{F}_t}(l) = P(L_{t+1} \leqslant l \mid \mathcal{F}_t)$$

i.e. the conditional loss distribution gives the conditional distribution of the L_{t+1} in the next time period given current information $\mathcal{F}_t$. The *unconditional loss-distribution* $F_{L_{t+1}}$ on the other hand is defined as the distribution under the stationary distribution $F_{\boldsymbol{Y}}$ of risk-factor changes. It gives the distribution of the portfolio loss if we consider a generic risk-factor change $\boldsymbol{Y}$ with the same distribution as $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_t$. The unconditional loss distribution is of particular interest if the time horizon over which we want to measure our losses is relatively large, as is frequently the case in credit risk management and insurance. For now, we proceed referring to the latter, as we intend the application of our methodology mainly for the actuarial field.

2.6.2 Portfolio optimisation problem definition

The goal of our risk averse investor is to find the optimal allocation $\boldsymbol{w}^*$, which minimises the risk associated with the linear portfolio $\mathcal{P}$, whose loss random variable $L_{t+1}(\boldsymbol{w})$ is expressed in (2.31). Instead of using variance we refer to downside risk measures as either the VaR_α or the expected shortfall ES_α.

Clearly, we do not know the exact portfolio loss distribution. The time series $(L_t)_{t\in\mathbf{N}}$ (stationary or not) is driven by a d-dimensional stochastic process $(\boldsymbol{X}_t)_{t\in\mathbf{N}}$ and a deterministic portfolio structure f, that in our case is linear. One has to model the cdf $F_{L_{t+1}}(x) = P(L_{t+1} \leqslant x)$ under various assumptions on the input. Only for the most trivial case, like $\boldsymbol{X}_{t+1}$ is multivariate normal, can this be handled analytically. In practice, one could:

- Model the whole loss distribution. Find stochastic models for $(\boldsymbol{X}_t)_{t\in\mathbf{N}}$ closer to reality, in particular beyond Gaussianity but for which $F_{L_{t+1}}$ can be calculated and estimated readily. This is a non-trivial task in general as it would involve the model of both the marginals and their dependence structure via *copulas*;

- Summarise loss risk. Rather than aiming for a full model for $F_{L_{t+1}}$, estimate some characteristics of $F_{L_{t+1}}$, namely estimate a risk measure, relevant in our case for solving the portfolio choice problem[22],

We follow the second approach as we do not need to model the whole portfolio loss distribution: extreme events occur when a risk takes values from the tail of its distribution and EVT attempts to provide us with the best possible estimate of the tail area regardless of its central part.

With reference to the classical portfolio optimization problem of Markowitz type in (2.6), we replace the variance by a generic quantile risk measure ρ with confidence level α and we look at the conservative version of the optimisation problem which minimizes the objective function

$$
\begin{aligned}
&\boldsymbol{\omega}^* = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \quad \rho_\alpha(L_{t+1}(\boldsymbol{w})) \\
&\sum_{i=1}^{d} \omega_i = 1, \\
&\omega_i \geqslant 0 \ , \ i \in \{1, \ldots, d\}
\end{aligned}
\tag{2.32}
$$

The conditions imply that the portfolio is fully invested and no short selling is allowed and the confidence level is a design parameter. The formulation is consistent with the safety first principle introduced by (Roy, 1952), considering investors that act primarily to guard against catastrophic losses. We conduct the portfolio risk minimisation with respect to both Value at risk and Expected shortfall (in the next formula, we omit subscript $t+1$ for readability)

[22] Or for solvency capital calculations in case we were using this methodology to model the Insurer's predictive aggregate loss distribution.

$$\begin{aligned}
&\underset{\boldsymbol{\omega}}{\operatorname{argmin}} \quad ES_\alpha(L(\boldsymbol{w})) && \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \quad VaR_\alpha(L(\boldsymbol{w})) \\
&= \underset{\boldsymbol{\omega}}{\operatorname{argmin}}\, \mathbb{E}[L(\boldsymbol{w}) \mid L(\boldsymbol{w}) > VaR_\alpha(L(\boldsymbol{w}))] && = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \quad F^{-1}_{L(\boldsymbol{w})}(\alpha) \\
&\sum_{i=1}^{d} \omega_i = 1, && \sum_{i=1}^{d} \omega_i = 1, \\
&\omega_i \geqslant 0 \ , \ i \in \{1, \dots, d\} && \omega_i \geqslant 0 \ , \ i \in \{1, \dots, d\}
\end{aligned} \tag{2.33}$$

When asset returns $\boldsymbol{X}_{t+1}$ are assumed to be multivariate normal, the optimal allocation of the minimum risk portfolio is the straightforward solution of a system of linear equations. Indeed, because of the special nature of the normal distribution[23], allocations which minimize the portfolios variance also minimize quantile-based measures of risk such as VaR_α and ES_α. Under this assumption we would simply obtain the *minimum variance portfolio*.

However, we have widely discussed that time series of financial asset values exhibit heavy tails and thus, relaxing the normality assumption that characterised the MPV, we apply the tools of EVT to estimate the two risk measures. In general, two approaches are available: one univariate and the other multivariate.

The former is the one we rely on and is called the *structure variable method* (SVM) and considers focusing on a univariate functional of a random vector $\boldsymbol{X}$ (Coles et al., 2001) instead of modelling its multivariate distribution. For our purposes, we choose the portfolio loss operator previously defined in (2.28) and in (2.31) as the structure variable

$$L_{t+1}(\boldsymbol{\omega}) = \Delta_{[t]}(\boldsymbol{\omega}, \boldsymbol{X}_{t+1}) = \sum_{i=1}^{d} \boldsymbol{\omega}_i X_{i,t+1} \tag{2.34}$$

$\Delta_{[t]} : \mathbb{R}^d \to \mathbb{R}$. Observe that, while $\boldsymbol{X}_{t+1}$ is a random vector, $L_{t+1}(\boldsymbol{\omega}) = \Delta_{[t]}(\boldsymbol{\omega}, \boldsymbol{X}_{t+1})$ is a univariate random variable.
The use of the SVM allow us to focus and model $L_{t+1}(\boldsymbol{\omega})$ instead of modelling $\boldsymbol{X}_{t+1}$.

In this context, techniques from univariate EVT can be used to model the tail of the portfolio loss distribution $F_{L_{t+1}(\boldsymbol{\omega})}$ and to estimate the two portfolio risk measures $VaR_\alpha(L_{t+1}(\boldsymbol{\omega}))$ and $ES_\alpha(L_{t+1}(\boldsymbol{\omega}))$ for fixed large confidence levels α. Then the minimum risk portfolio is the one with optimal allocation $\boldsymbol{\omega}^*$ which minimizes either $VaR_\alpha(L_{t+1}(\boldsymbol{\omega}^*))$ or $ES_\alpha(L_{t+1}(\boldsymbol{\omega}^*))$, for a fixed choice of confidence level α. We investigate how the optimal portfolio allocation changes as a function of α.

Specifically, we apply the *threshold exceedances method* of univariate extremes and approximate the excess distribution over a high threshold, of the structure variable $L_{t+1}(\boldsymbol{\omega})$ using the generalized Pareto distribution.

[23]In particular, this is due to the fact that linear combination of the components of a multivariate normal distribution are again normally distributed.

2.6.3 GPD model for the portfolio loss distribution

In general, the minimization of risk, either VaR_α or ES_α, requires the specification of the distribution of portfolio losses. Since this distribution is unknown, we rely on asymptotic results from EVT to specify the form of the tails of the distribution.

Treating each portfolio $L_{t+1}(\boldsymbol{\omega})$ in (2.34) as a different structure variable, when varying $\boldsymbol{\omega} \in \mathbb{R}^d$, univariate extreme value theory provides the necessary tools to model the tail distribution of each of those rv.

Setting (**Sequence of portfolio negative returns**). Let $(L_t(\boldsymbol{\omega}))_{t\in N}$ be a sequence of identically distributed random variables, with unknown distribution function $F(x) = P(L(\boldsymbol{\omega}) \leqslant x)$, representing the negative returns on a portfolio of investments with asset allocation $\boldsymbol{\omega} \in \mathbb{R}^d$.

At this point, the key passage is to refer to the assumption formalised in 2.1, in order to be allowed to rely on all the results presented in Sections 2.4.1 and 2.5.

Assumption (**Portfolio loss distribution $\boldsymbol{F \in MDA(H_\xi)}$**). Our model for the portfolio loss rv $L(\boldsymbol{\omega})$ assumes that its unknown distribution $F_{L(\boldsymbol{\omega})} = F \in MDA(H_\xi)$. Therefore, applying Theorem 2.2, the distribution of all losses exceeding some large threshold u (namely *excess distribution* over the threshold u) is approximately a generalized Pareto distribution GPD (as formalised in Assumption 2.1 for a generic risk X_i with distribution F).

More formally, for a high threshold u let

$$F_u(x) = P(L(\boldsymbol{\omega}) - u \leqslant x \mid L(\boldsymbol{\omega}) > 0) = G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & ,\xi \neq 0, \\ 1 - exp(-x/\beta) & ,\xi = 0, \end{cases}$$

for $0 \leqslant x < x_F - u$ and some $\xi \in \mathbb{R}$ and $\beta > 0$.

A GPD model for the excess losses is used to model the tail of the underlying portfolio loss distribution F. Thus, we can rely on the results found in Section 2.5 to estimate associated risk measures. In view of (2.18), for $y > u$, u large, we have

$$F(y) = 1 - \bar{F}(u)\left(1 + \xi\frac{y-u}{\beta}\right)^{-1/\xi}, \ u \leqslant y < x_F \tag{2.35}$$

The distribution F of $L(\boldsymbol{\omega})$ is arbitrary for $y < u$. Setting $F(y_\alpha) = \alpha$ and solving for y_α (namely $y_\alpha = VaR_\alpha(L(\boldsymbol{\omega}))$), we obtain

$$VaR_\alpha(L(\boldsymbol{\omega})) = u + \frac{\beta}{\xi}\left(\left(\frac{1}{\bar{F}(u)}(1-\alpha)\right)^{-\xi} - 1\right) \tag{2.36}$$

provided $VaR_\alpha(L(\boldsymbol{\omega})) > u$. The expected shortfall $VaR_\alpha(L(\boldsymbol{\omega})))$ can also be obtained by using the model of the tail of the portfolio loss distribution given in (2.35),

$$ES_\alpha(L(\boldsymbol{\omega})) = \frac{VaR_\alpha(L(\boldsymbol{\omega}))}{1-\xi} + \frac{\beta - u\xi}{1-\xi} \tag{2.37}$$

The restriction $\xi < 1$ implies that the heavy-tailed distribution must have at least a finite mean.

2.7 A Practical Discussion of EVT-based modelling

Suppose we have data from an unknown underlying distribution F, which we assume lies in the domain of attraction of an extreme value distribution H_ξ for some ξ. If the data are realizations of i.i.d. variables, or variables from a process with an *extremal index* (θ such as GARCH), the implication of the EVT theory is that:

- The true distribution of the n-block maximum M_n can be approximated for large enough n by a three-parameter generalised extreme value (GEV) distribution $H_{\xi,\mu,\sigma}$ (*Block maxima method*);
- The true distribution of the realisations of the loss rvs exceeding a suitable threshold u can be approximated, for a large enough threshold, by a two parameters generalised pareto distribution (GPD) $G_{\xi,\beta}$ (*Treshold exceedances*).

The shape parameter ξ in the generalized Pareto distribution $G_{\xi,\beta}$ is the same parameter as in the generalized extreme value distribution function $H_{\xi,\mu,\sigma}$. The scale parameter,

$$\beta = \sigma + \xi(u - \mu)$$

of $G_{\xi,\beta}$ is a function of the threshold u, and of the location, scale and shape parameters of the generalized extreme value distribution function $H_{\xi,\mu,\sigma}$.

The consistency and asymptotic efficiency of the resulting MLEs can be established for the case when $\xi > -1/2$ using results in Smith (1985) and for $\xi > -1$ referring to Zhou (2009, 2010), thus virtually covering all the cases of interest for financial applications. Note also that, in the case of dependent data ($\theta < 1$), somewhat larger block sizes than are used in the i.i.d. case may be advisable; dependence generally has the effect that convergence to the GEV distribution is slower, since the effective sample size is $n\theta$, which is smaller than n.

2.7.1 Estimation and Step procedure

We have a sample of multivariate observations of the random vector $\boldsymbol{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t}) \in \mathbb{R}$ representing the negative returns of the d assets in our universe

$$\tilde{\boldsymbol{X}}_s = (\tilde{X}_{1,s}, \tilde{X}_{2,s}, \ldots, \tilde{X}_{d,s}) \ , \ s \in \{1, \ldots, n\}$$

where

$$\tilde{X}_{i,s} \ , \ i \in \{1, \ldots, d\} \ , \ s \in \{1, \ldots, n\}$$

is the s^{th} observation of the ith marginal component of the random vector of losses from the d assets in our universe (i.e. realisation at time s of the ith asset's negative return rv). For fixed allocation $\boldsymbol{\omega} \in \mathbb{R}$, we then have n observations

$$\{\tilde{\Delta}_{[s-1]}(\tilde{\boldsymbol{X}}_s, \boldsymbol{\omega}) \ , \ s = 1, \ldots, n\} = \{\tilde{L}_s(\boldsymbol{\omega}) = \textstyle\sum_{i=1}^{d} \boldsymbol{\omega}_i \tilde{X}_{i,s} \ , \ s = 1, \ldots, n\}$$

of the structure variable that is the portfolio loss rv. Extreme value theory provides a methodology for estimating the risk measures

$$\widehat{VaR}_\alpha(L(\boldsymbol{\omega})) \text{ and } \widehat{ES}_\alpha(L(\boldsymbol{\omega}))$$

for α large (close to one) starting from those realised values.

Assumption (Data distribution $\boldsymbol{F} \in \boldsymbol{MDA(H_\xi)}$). Suppose that for each asset allocation $\boldsymbol{\omega} \in \mathbb{R}$ our data

$$\{\tilde{L}_s(\boldsymbol{\omega}) = \textstyle\sum_{i=1}^{d} \boldsymbol{\omega}_i \tilde{X}_{i,s} \ , \ s = 1, \ldots, n\}$$

are drawn from a sequence of identically distributed portfolio loss rvs, with unknown underlying distribution $F_{L(\boldsymbol{\omega})} = F$ that lies in the domain of attraction of an extreme value distribution H_ξ for some ξ.

We can therefore estimate the GPD model for the tail of the portfolio loss distribution (presented in Section 2.6.3) on those realisation exceeding a fixed high threshold u via maximum likelihood (Section 2.4.1). In particular, a number N_u of the sample observations of our loss rvs will exceed the threshold u

$$\{\tilde{L}_j(\boldsymbol{\omega}) = \textstyle\sum_{i=1}^{d} \boldsymbol{\omega}_i \tilde{X}_{i,j} \ , \ j = 1, \ldots, N_u\}$$

For each of those we calculate the amount of excess loss $\{\tilde{L}_j(\boldsymbol{\omega}) - u, j = 1, \ldots, N_u\}$. Once we have estimated the parameter of the GPD $\hat{\xi}, \hat{\beta}$ by fitting the distribution over the N_u exceedances, we use them to get an estimate for the risk measures (Section 2.5.4).

In view of (2.23) and (2.24), for confidence levels $\alpha \geqslant 1 - N_u/n$, we obtain the point estimators of the risk measures calculated over the portfolio (with allocation $\boldsymbol{\omega}$) loss distribution

$$\widehat{VaR}_\alpha(L(\boldsymbol{\omega})) = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u}(1-\alpha) \right)^{-\hat{\xi}} - 1 \right) \tag{2.38}$$

$$\widehat{ES}_\alpha(L(\boldsymbol{\omega})) = \frac{\widehat{VaR}_\alpha(L(\boldsymbol{\omega}))}{1-\hat{\xi}} + \frac{\hat{\beta} - u\hat{\xi}}{1-\hat{\xi}} \tag{2.39}$$

As we can notice from the formulas above, for each vector of weights $\boldsymbol{\omega}$ (representing the allocation among the d assets) we have a different linear portfolio, and thus we get a different time series of realisations of the related loss random variable.

Setting 2.1 (**Model estimation - Step procedure**). For every allocation $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)$

(i) Data. Starting from the time series of asset (negative) returns $\tilde{\boldsymbol{X}}_s$, $s \in \{1, \ldots, n\}$ we obtain the time series of portfolio (negative) returns $\{\tilde{L}_s(\boldsymbol{\omega}) = \sum_{i=1}^d \boldsymbol{\omega}_i \tilde{X}_{i,s}$, $s = 1, \ldots, n\}$;

(ii) Parameters estimation. Choice of the threshold u and estimation of the parameters $(\hat{\xi}, \hat{\beta}, \widehat{\bar{F}(u)})$ fitting the GPD distribution of the excess loss over the threshold;

(iii) Risk measures. Estimators of $\widehat{VaR}_\alpha(L(\boldsymbol{\omega}))$ and $\widehat{ES}_\alpha(L(\boldsymbol{\omega}))$ are immediately given by substituting estimated parameter values into (2.38) and (2.39).

The optimal portfolio allocation $\boldsymbol{\omega}^*$ among the d assets, is the vector of weights that leads to minimum portfolio risk, assessed in terms of minimum value of the risk measures estimated in step (iii).

2.7.2 Threshold choice

Before one is allowed to proceed with the model estimation, a suitable, high enough, threshold u needs to be chosen (point (ii) in the list).

This is a very delicate issue concerning statistical methods of EVT, since this choice entails a trade-off between bias and variance of the parameters estimate. On the one hand, we would like u as large as possible in order to reduce the bias incurred from applying an asymptotic theorem and eventually including ordinary observations (not extreme). On the other, when u is too large, we are left with few, if any, observations to estimate the parameters, compromising their statistical accuracy. In addition, the parameter β in the GPD model has a theoretical dependence on the threshold chosen. Indeed, it appears in the theory as a function of the threshold u (see Eq. (2.13) and Theorem 2.2). In practice, we would expect the estimated β parameter to vary with threshold choice.

Methods for selection of the threshold u may be found for example in Embrechts et al. (1997), Coles et al. (2001). Below we briefly present the main alternatives.

Conventional Choices

Many authors do not address explicitly the problem of the choice of the cut-off for the data they are handling, but simply follow either common sense choices or suggestions retrieved in the literature. A widely used suggestion in this respect is that the number of data falling in the tail should not be higher than 10-15% and a rule of thumb value of 5-10% is often used;

Graphical Methods

These graphical tools can be very helpful in the choice of u, subtracting it from the domain of pure convention or common sense, though they are still far from a mathematical standard of rigour and still allow for considerable degrees of arbitrariness.

Hill plots When using the *Hill estimator* to estimate the tail index $\alpha = 1/\xi$[24] of a given distribution, this is a very common method to determine a good choice for the cut-off (k that is the number N_u of the exceedances over the threshold u). It is a graphical representation, in a coordinate system, of $\{(k, \hat{\alpha}^H_{k,n}) : k = 1, \ldots, n\}$ the estimates of the tail index as a function of the cut-off. Regions of the plot that are approximately close to be horizontal lines indicate values of k for which theestimate $\hat{\alpha}^H_{k,n}$ is essentially stable with respect to the choice of the cut-off;

Pickand plots Analogously, this method can be employed when dealing with the *Pickands estimator* that was studied by Pickands (1975), and unlike the Hill estimator, can be used to estimate the shape parameter of any of the three extreme value distributions[25]. In this graph the set $\{(k, \hat{\alpha}^P_{k,n}) : k = 1, \ldots, \lfloor n/4 \rfloor\}$, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$, is plotted;

Mean excess plots The mean excess plot is a tool widely used in the study of risk, insurance and extreme values. One use is in validating a generalized Pareto model for the excess distribution. It consists in plotting $\{(X_{i,n}, e_n(X_{i,n})) : i = 2, \ldots, n\}$, where $X_{i,n}$ denotes the ith order statistic. If the data have an excess distribution over high thresholds which is distributed according to a GPD, then the resulting plot will be roughly linear (for the higher order statistics) and the slope of the line will provide evidence concerning the sign of the shape parameter ξ (an upward line corresponding to $\xi > 0$, i.e. the Fréchet case; an horizontal plot, to $\xi = 0$; a downward plot, to $\xi < 0$.

[24] Please refer to Section 3.3.2 Chapter 3 for further details on the Hill estimator.

[25] Pickands estimator is used in applied papers such as Longin (2005), which employs both Hill and Pickands estimators, and Ho (2008).

Monte Carlo Simulation and MSE Minimization

A largely used selection method for automatically obtaining the cut-off between the centre and the tail of the distribution is the one proposed by Koedijk et al. (1990), Jansen and de Vries (1991) and subsequently employed in many papers concerning financial applications of EVT, such as Longin and Solnik (2001), Haile and Pozo (2006), Vilasuso and Katz (2000). This method sets the minimization of the mean squared error (MSE) as the optimality criterion for the choice of the cut-off, thus taking simultaneously into account both the bias and the inefficiency components. In practice, the selection of the optimal cut-off is based on the following algorithm:

(i) Simulate data via Monte Carlo from a known distribution F in the domain of attraction of a Fréchet distribution with tail index α;
(ii) Compute the Hill estimator of α for different choices of k;
(iii) Choose the value of k that has minimal MSE.

The simulation from a known distribution is required. The usual choice for F is the cdf of a Student t, which allows for different degrees of heavy tailedness and is standard when modelling financial data. However, this method has two major drawbacks. The first one is the distributional assumption in step (i), since it seems denying one of the most appealing features of EVT, that is the possibility of modelling the tails without knowing the exact underlying distribution. In addition, there is no guarantee that the optimal choice for k, when dealing with simulated data, coincides with the optimal choice induced by real data.

The second drawback is related to the choice of MSE minimisation as an optimality criterion (and thus it affects data-driven methods as well). On the one hand, Manganelli and Engle (2001) point out that the evaluation of the optimal cut-off based on this criterion yields a biased estimator. On the other hand, if the i.i.d. hypothesis that we have assumed up to now fails, Kearns and Pagan (1997) show that actual standard errors are greater than those predicted by the asymptotic theory for independently distributed random variables. Therefore, when the data are not i.i.d., an MSE based optimality criterion is questionable.

Data Driven Algorithms

To overcome the need of a distributional assumption in the simulation based MSE minimization, several algorithms have been proposed which endogenously generate the selection of an optimal cut-off, relying on the real data. We refer the reader to the paper by Lux (2000), that provides a review and a comparative analysis of the main ones. He focuses on five methods: we bring to attention the method proposed by Hall (1990) and Danielsson and de Vries (1997b), which employs a bootstrap approach to estimating MSE, and the method by Beirlant and Novak (2006), which is based on an iterative regression approach.

Setting the threshold u in our model

Within our optimal portfolio choice model, we are required to choose a large u for each allocation $\boldsymbol{\omega}$ to be able to apply the methodology and obtain the risk measures estimates we need to identify the minimum risk portfolio. As we can see in the modelling step procedure presented in Setting 2.1, in order to solve the portfolio optimisation problem, an estimation of the tail distribution across many different portfolio allocations is necessary. Therefore, we are not focussing on just one random variable but we are modelling every portfolio loss rv resulting from all the allocation vector considered by the optimum search algorithm. In theory, based on the above-mentioned available methods, each of these allocations requires a careful selection of the choice of u. Considering our specific settings, this approach would be intractable in terms of managing such high problem dimension.

Setting 2.2 (**Threshold choice**)**.** With regard to the choice of a suitable threshold, considering our problem specification, we refer to the so called "conventional choice method". In practice, we choose u to be some high quantile of the empirical distribution. Keeping in mind the trade-off between bias and variance in choosing u and based on the best practice found in the related literature, we take u as the 95% quantile of the empirical distribution of $L(\boldsymbol{\omega}), \boldsymbol{\omega} \in \mathbb{R}^d$.

2.7.3 Simulation study

In this section we perform a simulation study via Monte Carlo simulation to better understand the sensitivity of the EVT threshold exceedances approach to the choice of the threshold. Both the non parametric (Hill estimator) and fully parametric (GPD model) are considered and compared.

In particular, we perform first the estimation of the shape parameter ξ and then the estimation of the high quantile VaR_α. The two estimators are compared using mean squared errors ($MSEs$) in both situations; we recall that the MSE of an estimator $\hat{\vartheta}$ of a parameter ϑ is given by $MSE(\hat{\vartheta}) = \mathbb{E}(\hat{\vartheta} - \vartheta)^2 = (\mathbb{E}(\hat{\vartheta} - \vartheta))^2 + var(\hat{\vartheta})$, and thus has the well-known decomposition into squared bias plus variance. A good estimator should keep both the bias term $\mathbb{E}(\hat{\vartheta} - \vartheta)$ and the variance term $var(\hat{\vartheta})$ small. Since analytical evaluation of bias and variance is not possible, we calculate Monte Carlo estimates by simulating 1000 datasets in each experiment. The estimates are calculated using the Hill method and the GPD method based on different numbers of upper order statistics (or differing thresholds) and try to determine the choice of k (or N_u) that is most appropriate for a sample of size n. Note that the parameters of the GPD are always determined by ML. In the case of estimating VaR we also compare the EVT estimators with the simple empirical quantile estimator.

The simulation experiment is conducted as follows. We assume that we have a sample of 1000 i.i.d. data from a t distribution with four degrees of freedom and we want to estimate ξ, the reciprocal of the tail index, which in this case we know is equal to 0.25[26]. The Hill estimate is constructed for k values in the range $\{2, \ldots, 200\}$ and the GPD estimate is constructed for k (or N_u) values in $\{30, 40, 50, \ldots, 400\}$. The results are shown in Figure 2.2.

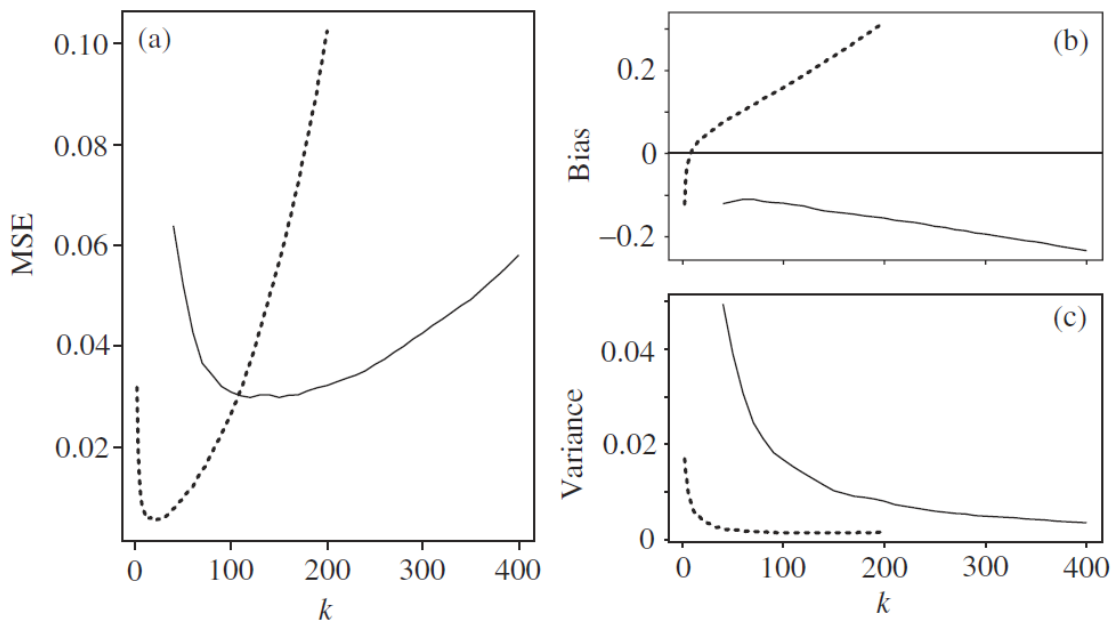


Figure 2.2: Comparison of (a) estimated MSE, (b) bias and (c) variance for the Hill (dotted line) and GPD (solid line) estimators of ξ as a function of k (or N_u), the number of upper order statistics from a sample of 1000 t-distributed data with four degrees of freedom.

The t distribution has a well-behaved regularly varying tail and the Hill estimator gives better estimates of ξ than the GPD method, with an optimal value of k around 20-30. The variance plot shows where the Hill method gains over the GPD method; the variance of the GPD-based estimator is much higher than that of the Hill estimator for small numbers of order statistics. The magnitudes of the biases are closer together, with the Hill method tending to overestimate ξ and the GPD method tending to underestimate it. If we were to use the GPD method, the optimal choice of threshold would be one giving 100-150 exceedances.

The conclusions change when we attempt to estimate the 99%VaR; the results are shown in Figure 2.3. The Hill method has a negative bias for low values of k but a rapidly growing positive bias for larger values of k; the GPD estimator has a positive bias that grows much more slowly; the empirical method has a negative bias. The GPD attains its lowest MSE value for a value of k around 100, but, more importantly, the MSE is very

[26]The distribution function F_ν of a t distribution has tail index ν (see Example 7.29 McNeil et al., 2005).

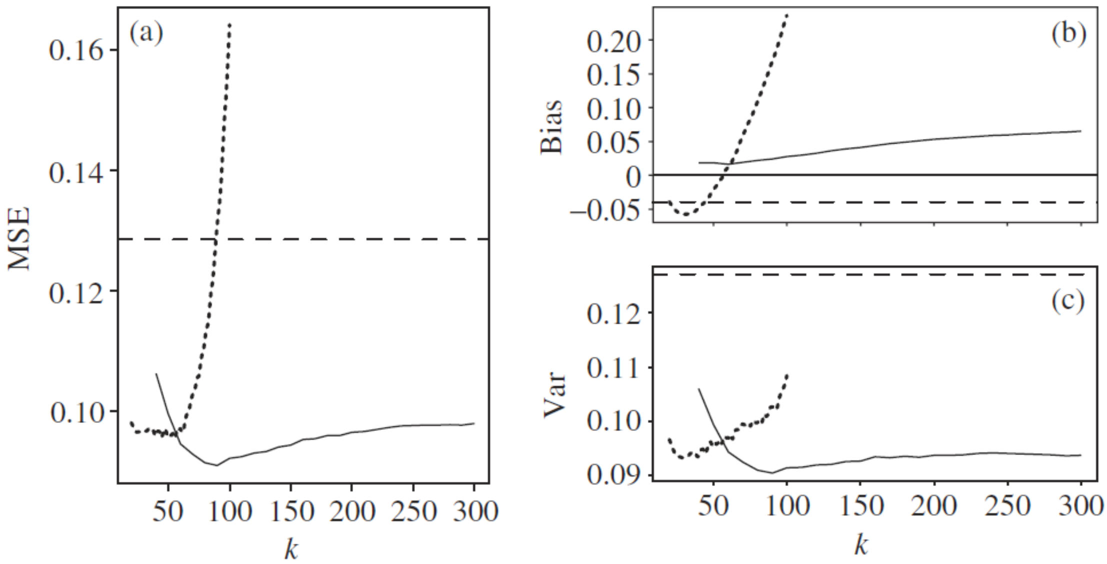


Figure 2.3: Comparison of (a) estimated MSE, (b) bias and (c) variance for the Hill (dotted line) and GPD (solid line) estimators of $VaR_{0.99}$, as a function of k (or N_u), the number of upper order statistics from a sample of 1000 t-distributed data with four degrees of freedom. Dashed line also shows results for the (threshold-independent) empirical quantile estimator.

robust to the choice of k because of the slow growth of the bias. The Hill method performs well for $20 \leqslant k \leqslant 75$ (we only use k values that lead to a quantile estimate beyond the effective threshold $X_{k,n}$) but then deteriorates rapidly.

Both EVT methods obviously outperform the empirical quantile estimator. Given the relative robustness of the GPD-based tail estimator to changes in k, the issue of threshold choice for this estimator seems less critical than for the Hill method.

2.7.4 Optimisation algorithm

The problem that motivates the introduction of an optimisation algorithm is the relevance of the dimensionality issue in portfolio selection, since, as the number of assets grows, the so called *brute force approach* to the calculation of the optimal portfolio becomes soon intractable.

Indeed, when looking at the bivariate case, two assets X_1 X_2, we could consider n_ω different portfolio allocations of the form $\omega_j = (\omega_{1,j}, \omega_{2,j}), j = 1, \ldots, n_\omega$ where $\omega_{2,j} = 1 - \omega_{1,j}$, and take $\omega_{1,j}$ to form an evenly spaced grid over the unit interval; that is, $\omega_{1,j} = (j-1)\delta_\omega$ where $\delta_\omega = 1/(n_\omega - 1)$. Since the computational burden in the two asset problem is minimal, considering $\delta_\omega = 1\%$ is reasonable as gives a grid of $n_\omega = 101$ points. Then proceeding with the estimation over the possible allocations, included in the grid, we can identify the minimum risk portfolio.

However, we are interested in constructing optimal portfolios of multiple assets. In this case, the described methodology becomes intractable. For instance, with three assets, the same 1% incremental step gives 5151 different possible allocations. Thus, we refer and use a two stage methodology instead. The first stage uses the sampling algorithm of Bensalah et al. (2002) to pick an appropriate starting position. The second stage, that is an improvement over Bensalah's algorithm provided by Bradley and Taqqu (2004b), uses an incremental trade algorithm to step away from the starting allocation in the direction of highest marginal utility.

The algorithm stops when there is no longer any marginal benefit to trading. We compare optimal allocations under the assumption of normality and under the incremental trade algorithm.

Let α be the design parameter and take the risk to be VaR_α or ES_α. Our universe consists of d securities and we consider portfolios in a subset of the d-dimensional unit cube, that is, $\boldsymbol{\omega}^* \in [0,1]^d$ such that $\sum_{i=1}^d \omega_i = 1$ and $\omega_i \geqslant 0$. To get a starting position, we randomly sample from portfolios in this set and compute their risk. We also consider the normal allocation $\boldsymbol{\omega}^{\mathcal{N}}$ and the d portfolios that consist of only one among the d assets in turn assets as a possible starting position. Specifically, we take as a starting position the allocation $\boldsymbol{\omega}_{st} \in [0,1]^d$, where $\boldsymbol{\omega}_{st}$ has minimal risk over all n_s sample allocations

$$\boldsymbol{\omega}_j = (\omega_{1,j}, \ldots, \omega_{d,j}), j = 1, \ldots, n_s$$

The n_s sample allocations are constructed as follows. For the jth sample, take

$$\begin{aligned}
\omega_{1,j} &\sim U([0,1]) \\
\omega_{2,j} &\sim U([0,1-\omega_{1,j}]) \\
&\vdots \\
\omega_{d-1,j} &\sim U([0,1-\omega_{1,j}-\cdots-\omega_{d-2,j}]) \\
\omega_{d,j} &= U([0,1-\sum_{i=1}^{d-1}\omega_{i,j}])
\end{aligned}$$

We fit the tail of the structure variable $L(\boldsymbol{\omega}_j)$ for each sample allocation $\boldsymbol{\omega}_j, j = 1, \ldots, n_s$. For each high quantile of interest α, the risk, $VaR_\alpha(L(\boldsymbol{\omega}_j))$ or $ES_\alpha(L(\boldsymbol{\omega}_j))$, is estimated accordingly. The starting position $\boldsymbol{\omega}_{st}$ is the allocation which minimizes the risk over all sample allocations $\boldsymbol{\omega}_j, j = 1, \ldots, n_s$.

Once a starting position $\boldsymbol{\omega}_{st}$ is obtained, we apply an incremental trade algorithm to find the minimum risk portfolio. It needs to be applied separately for VaR_α and ES_α. Let δ_ω be the size of the incremental trade to be considered. The algorithm takes steps of size δ_ω in each asset away from its current position (buying and selling). This amount δ_ω bought or sold in the ith asset leaves the portfolio over or under invested by δ_ω. This amount is then spread across all possible (except the ith) assets in order to leave the portfolio fully invested. The resulting portfolio has an incremental trade of δ_ω in the ith asset and a very minor change (relative to δ_ω) in the remaining assets.

More precisely, suppose we want to buy δ_ω in the ith asset. The allocation in the ith asset would go from $\omega_i \to \omega_i + \delta_\omega$, but this would leave the resulting portfolio over-invested by δ_ω. To overcome this, we use δ_ω as a design parameter and pick the additional allocation δ such that

$$\frac{\omega_i + \delta}{1 + \delta} = \omega_i + \delta_\omega$$

and then replace the weight ω_i by $\omega_i + \delta$. Next, we normalize all weights

$$\boldsymbol{\omega}_j = (\omega_1, \dots, \omega_{i-1}, \omega_i + \delta, \omega_{i+1}, \dots, \omega_d)$$

by $1 + \delta$. We follow a similar procedure for selling the ith asset $\omega_i \to (\omega_i - \delta)/(1 - \delta)$ and repeat the procedure for all d assets. At each iteration, the algorithm picks the trade which is most risk reducing. It continues in this fashion until the risk is minimized. If, for any given stage, no trade provides a reduction in risk, then we continue considering trades of marginally larger size δ_ω for a fixed number of times. This is done to avoid stopping at a local minimum. Local extrema of the objective function are due to the fact that one performs a separate statistical analysis at each point of the allocation space. Estimation error alone leads to local extrema of the objective function. When there is no marginal benefit to continued trading, the trading stops and the optimal allocation $\boldsymbol{\omega}^*$ is returned. Note that although the algorithm currently only considers the minimization of risk, it is straightforward to include other factors in the utility function. For example, it is straightforward to add expected returns to the utility function.

Chapter 3

Extreme Value Dependence

The concept of diversification is fundamental to the decision of how best to allocate assets and this is also true in the classical MVO. The benefit of diversification is partly controlled by the dependence between the assets under consideration. We perform a parallel analysis to identify the correct extremal dependence structure of the portfolio's constituents and to understand how this may impact the optimal asset allocation problem presented in the previous chapter.

In this chapter, we present a multivariate framework for studying rare events in finance, particularly useful for this purpose, and suitable when asset returns consist of more than one component, as in the case of a portfolio. The methodology is based on multivariate extreme value theory (MEVT) and, specifically, on the results of Poon et al. (2003, 2004). We believe that this line of research is of particular interest to international portfolio managers, who systematically monitor correlations between various stock markets, and the success of their policies is intimately linked to the stability of the estimated dependencies.

3.1 Multivariate Extremes

Financial risk models are inherently multivariate. As an example, the value change of a portfolio of traded instruments over a fixed time horizon depends on a random vector of risk-factor changes or returns. A stochastic model for a random vector can be thought of as simultaneously providing probabilistic descriptions of the behaviour of the components of the random vector and of their dependence or correlation structure. Therefore, estimating dependence between risky asset returns is essential in portfolio theory and many other finance applications such as hedging, spread analysis and valuation of exotic options written on more than one asset. In general, the modelling of interdependence between risk factors or risk-factor changes is a relevant basic question in Quantitative risk management.

A great part of the financial community uses linear correlation to describe any measure of dependence. Correlation lies at the heart of the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), where its use as a measure of dependence between financial instruments is essentially founded on an assumption of multivariate normally distributed returns. Without the restrictive assumption of normality, is linear correlation still an appropriate measure of dependence? Outside the elliptical world correlation must be used with care (Embrechts et al., 1999) and unfortunately, most dependent real world risks do not appear to have an elliptical distribution.

The conventional dependence measure, Pearson correlation, though widely used is appropriate only for linear association. It is constructed as an average of deviations from the mean, the weight given to extreme realizations is the same as for all of the other observations in the sample. If the dependence characteristics for extreme realizations differs from all others in the sample, the conclusions drawn from the Pearson correlation could result in a financial institution risking bankruptcy. Thus modern risk management calls for an understanding of stochastic dependence going beyond simple linear correlation.

For all of these reasons, alternative approaches to understanding and modelling dependency has been introduced: *copulas*[1] and various other dependence concepts (including, for instance, *comonotonicity*, *rank correlation* and *tail dependence*)[2], useful to risk management practitioners. The multivariate extreme value theory (MEVT) framework offers further tools to study the dependence structure when one is particularly concerned with extreme values, and need asymptotic measures of tail dependence to perform the analyses. MEVT can be used to model the tails of multivariate distributions and one way of understanding it, is as the study of copulas which arise in the limiting multivariate distribution of component-wise block maxima (see Section 4.5.3 in Chapter 4).

Understanding the dependence between extreme values in a multivariate random sample is of great interest and it also plays an important role in terms of affecting the benefit of diversification that characterise an asset allocation problem. Multivariate extreme value theory yields the ideal modelling environment to tackle this aspect within our model, as we clearly do not rely on elliptical risks assumptions; moreover, MEVT also represents the natural evolution of the methodology we have already referred to in the univariate case. Basically dependence structures can be grouped into four types: independent, perfect dependent, asymptotic independence and asymptotic dependence.

The distinction between the two asymptotic dependence structures occurs as both variables approach their respective upper limits. Most finance articles adopt MEVT models that assume asymptotic dependence without making clear distinction between those possible dependence structures.

[1]The word copula first appeared in the statistics literature Sklar (1959), although similar ideas and results can be traced back to Hoeffding (1940).

[2]See for example Embrechts et al. (2003) for further details.

3.1.1 Literature on Extremal dependence across markets

The issue of extreme dependence among financial time series can be specifically addressed when studying correlation and contagion among markets, in particular during period of economic difficulty.

It has been widely documented in many empirical studies that equity markets around the world exhibit a higher correlation during periods of crises. Although this is hardly disputed, there is a rising controversy on the economic interpretation that should be given to it.

Changes in the correlation coefficient of financial asset returns have been typically associated with the notion of contagion. This term refers to the spread of downside market shocks from one country to another and it can be explained from the existence of real and financial linkages between the countries or the behaviour, rational and irrational, of international investors[3]. However, before one proceeds with the explanations that have been offered to the correlation breakdown issue, there must be an agreement that this breakdown has really occurred.

In this context, the econometric investigation of the stability of the correlation coefficient has highlighted two major problems. First, the choice of "conditioning" the correlation index on periods of high volatility is not the appropriate one if we intend to test for a correlation breakdown. A second, and perhaps more critical, issue is related to the suitability of the Pearson correlation index as a statistical measure of dependence when returns are not drawn form the class of elliptical distributions, a distinct member of which is the multivariate normal distribution[4]. It is known, for example, that the conditional correlation of a multivariate normal distribution tends to zero as the threshold used to define the tails tends to infinity. This contradicts however the widely held view that correlation across markets increases dramatically in the presence of large negative shocks.

For this reason there is a stream of literature addressing this issue by making use of asymptotic results from the multivariate EVT which hold for a wide range of parametric distributions. Two main approaches have been adopted to study this aspect. Firstly, dependence between different countries has been analysed in detail for given financial sectors. Secondly, dependence among different financial markets within the same country was also of interest to many authors.

The paper of Longin and Solnik (2001) falls within the first type and it applies multivariate EVT to test the validity of the common belief that correlation between stock markets increases during volatile periods. They use multivariate threshold exceedances,

[3]See Karolyi (2003) for a survey on the issue.

[4]Embrechts et al. (2003, 2002) discuss the role of the linear correlation coefficient and other measures of dependence outside the class of elliptical distributions.

modelling dependence via the effective, though parsimonious, logistic model, in which a single parameter accounts for dependence and this parameter is related to correlation by a simple formula[5]. Fitting a bivariate model to pairs of monthly equity index returns they find more dependence in the tails of the distribution than can be explained by the multivariate normal distribution. Furthermore, the correlation pattern of threshold exceedances is asymmetric and it depends on the chosen threshold both in size and in sign. In particular, they find that correlation of exceedances is increasing in the absolute value of the threshold, if the threshold itself is negative; otherwise, it is decreasing. Moreover, it appears that it is a bear market, rather than volatility per se, that is the driving force in increasing international correlation.

In a complementary way, Bekiros and Georgoutsos (2008) study the correlation of extreme returns between seven Asia-Pacific stock markets and the U.S. They use the same methodology as Longin and Solnik (2001) recurring to a bivariate logistic dependence function to model threshold exceedances, in order to produce a ranking of Asia-Pacific countries in three risk categories (low, medium and high risk groups). In other words, they attempt to answer the question whether the contagion of financial crises is more or less the same for the seven Asia-Pacific countries.

They also examine whether there is any validity to the argument that the Asia-Pacific capital markets belong to a distinct cluster of markets, where the other two could be the U.S. and Europe. A clustering analysis shows that the Asia-Pacific countries do not belong to a distinct block based on the extreme correlations they have estimated within the group. The conclusion is that Japan and U.S. exhibit varying degrees of extreme correlation with the other markets and therefore their investors can benefit from diversifying their portfolios with assets from the Asia-Pacific stock markets. This remains true even during crisis periods as they conducted a sensitivity analysis estimating the model with two distinct sets of data, one including the 1987 crash and the other on the period immediately after it, yielding to similar results. The fact that these results are close to the correlation estimated via standard unconditional and conditional (GARCH models) methods provides evidence against contagion having occurred during the 1990s crises within the Asia-Pacific stock markets (namely no correlation breakdown has been observed).

The choice of the previous articles to measure extreme dependence by means of the correlation coefficient is however questionable.

Poon et al. (2003, 2004) apply new dependence measures from multivariate EVT (the coefficient of upper tail dependence and a complementary measure) to quantify dependence between equity markets from the G-5 countries[6]. They find little evidence of asymp-

[5] They use a transformation of the dependence parameter of the bivariate logistic distribution to estimate the conditional correlation between pairs of assets in the tails of the distribution.

[6] They apply the pair of dependence measures on daily data of stock index returns of the five largest

totic dependence between markets and they conclude that the asymptotic dependence between the European countries (United Kingdom, Germany and France) has increased over time but that asymptotic independence between Europe, United States and Japan best characterizes their stock markets behaviour. This is in agreement with our results.

Regarding the second aspect of dependence, namely dependence across different kinds of financial activities (cross-asset or inter-sector dependence), Hartmann et al. (2004) study contagion risk between stock markets and government bond markets of the G-5 countries. They characterize asset return linkages during periods of stress by a nonparametric extremal dependence measure and, contrary to correlation analysis, this is not predisposed toward the normal distribution and can allow for nonlinear relationships. Aiming at determine whether contagion is present between markets and across asset classes, their estimates for the G-5 countries suggest that simultaneous crashes between stock markets are much more likely than between bond markets.

Finally, for policy purposes, a particular interest is attached to studies regarding currency crises contagion and cross-country dependence in the banking sector (see for example Haile and Pozo, 2008; Garita and Zhou, 2009; Hartmann et al., 2005; Chan-Lau et al., 2007; Pais and Stork, 2011). We are not giving further details as it is not within our scope, but it is worth to be mentioned.

In general, there are two classes of extreme value dependence, *asymptotic dependence* and *asymptotic independence*, for which the characteristics of events behave quite differently as the events become more extreme. Both forms of extremal dependence permit dependence between moderately large values of each variable, but the very largest values from each variable can occur together only when the variables exhibit asymptotic dependence. Conventional multivariate extreme value theory has emphasized the asymptotically dependent class resulting in its wide use in most of the finance applications. If the series are truly asymptotically independent, such an approach will result in the over-estimation of extreme value dependence, and consequently of the financial risk.

We use the set of dependence measures from EVT proposed by Ledford and Tawn (1997) and Coles et al. (1999) to distinguish between asymptotically dependent or independent variables, and to quantify the degree of extremal dependence between international equity markets. The estimation methodology introduced by Poon et al. (2003, 2004) is applied. Our focus is on dependence estimation in the bivariate context, as this is the most common framework due to the complexity arising in higher dimensions.

---

stock markets.

3.2 Measures of Extreme value dependence

The joint distribution of a set of variables can be separated into their respective marginal distributions and the dependence structure among them. We focus on studying the dependence among extreme realizations of different components of a multivariate process. In the study of multivariate dependence structure, it is helpful to remove the influence of marginal aspects first by transforming the raw data to a common marginal distribution. After such a transformation, differences in distributions are purely due to dependence. Thus, suppose we have a bivariate returns (X, Y), we transform them to unit Fréchet marginals (S, T) as follows:

$$S = -1/\log F_X(x) \qquad \text{and} \qquad T = -1/\log F_Y(y) \tag{3.1}$$

where F_X and F_Y are the respective marginal distribution functions for X and Y. The Fréchet transformation is used because of the widely documented fat-tail distributions for risk asset returns. Consequently S and T have the distribution function $F(s) = e^{-1/s}$ for $s > 0$, so $P(S > s) = P(T > s) = s^{-1} + O(s^{-2}) \sim s^{-1}$ as $s \to \infty$. The variables (S, T) possess the same dependence structure as (X, Y) since

$$P(q) = P(F(T) > q \mid F(S) > q) = P(Y > F_Y^{-1}(q) \mid Y > F_X^{-1}(q)) \tag{3.2}$$

In practice, the values of F_X and F_Y, used in the transformation, are obtained using the empirical distribution functions of the separate variables.

We will focus our discussions on the dependence estimation in the bivariate context, though the ideas and techniques extend naturally to higher dimensions. The extension to general multivariate distributions can be theoretically conceivable, but it is often demanding from a computational viewpoint and it is not always completely straightforward. Thus, in practice, most applications restrict to k-variate EVT with $k = 2$ or $k = 3$.

The variables S and T are said to be *asymptotically independent* if $P(q)$ has a limit equal to zero as $q \to 1$. If the limit in Equation (3.2) is non zero instead, S and T are described as *asymptotically dependent.* Values of $P(q)$ greater than, equal to, and less than $(1 - q)$ indicate positive dependence, independence, and negative dependence, respectively, at percentile q.

In next Sections 3.2.1 and 3.2.2, we present measures of the degree of asymptotic dependence and asymptotic independence, respectively. Although it will be seen that these measures have some similarities with the co-skewness measure, they are not exactly the same. Positive co-skewness describes the case where large positive realizations of one series coincide with large absolute realizations of another series.

The extremal measures here can be used to measure dependency of the same or different signs and they provide two further advantages: they are invariant to the marginal distributions of the random variables and they give the most relevant summary of the dependence between variables for extreme values as they quantify the characteristics of $P(q)$ as $q \to 1$. If tail events are systematic as well, one might expect the extremal dependence between the asset returns and the market factor returns to be priced and command a risk premium. The extremal dependence measures we analyse provide the flexibility for testing such a pricing relationship separately for positive, negative, or unsigned systematic tail dependence.

3.2.1 The conventional approach χ

To understand extremal dependence, one must first appreciate that the form and degree of such dependence determine the chance of obtaining large values of both variables. As S and T are on a common scale, events of the form $(S > s)$ and $(T > s)$, for large values of s, correspond to equally extreme events for each variable. As all such probabilities will tend to zero as $s \to \infty$ it is natural to consider conditional probabilities of one variable given that the other is extreme.

Specifically, consider the behaviour of $P(T > s \mid S > s)$ for large s. If (S, T) are *perfectly dependent* then $P(T > s \mid S > s) = 1$. In contrast, if (S, T) are *exactly independent* then $P(T > s \mid S > s) = P(T > s)$, which tends to 0 as $s \to \infty$.
The first nonparametric measure of dependence we introduce χ, is based on $P(q)$ in Eq. (3.2).

Definition 3.1 (χ dependence measure). Let be (X, Y) a bivariate rv, and let be (S, T) the transformation into unit Fréchet marginals.

$$\chi = \lim_{q \to 1} P(q) = \lim_{s \to \infty} P(T > s \mid S > s) = \lim_{s \to \infty} \frac{P(T > s, S > s)}{P(S > s)} \tag{3.3}$$

and $0 \leqslant \chi \leqslant 1$. We say that variables are *asymptotically dependent* if $\chi > 0$, *asymptotically independent* if $\chi = 0$ and *perfectly dependent* if $\chi = 1$.

Clearly χ measures the degree of dependence that persists to the limit. Recalling that, motivated by MEVT, we standardise the marginals to have unit Fréchet distribution

$$P(S \leqslant s) = P(T \leqslant s) = e^{-1/s} \text{ for } s > 0$$

Hence

$$P(S > s) = P(T > s) \sim s^{-1} \text{ as } s \to \infty.$$

Perfect dependence ($S = T$ a.s.) corresponds to

$$P(S > s, T > s) = P(S > s) \sim s^{-1}$$

and exact independence corresponds to

$$P(S > s, T > s) = P(S > s)^2 \sim s^{-2}$$

The dependence measure χ presented in (3.3) is a version of the so called *coefficient of upper tail dependence*, that is invariant with respect to the marginal distributions of the random variables as it was obtained conveniently standardizing the marginal distributions.

Definition 3.2 (**coefficient of upper/lower tail dependence**)**.** Let X_1 and X_2 be rvs with cdfs F_1 and F_2. The *coefficient of upper tail dependence* of X_1 and X_2 is

$$\lambda_u := \lambda_u(X_1, X_2) = \lim_{q \to 1^-} P(X_2 > F_{X_2}^{-1}(q) \mid X_1 > F_{X_1}^{-1}(q))$$

provided a limit $\lambda_u \in [0, 1]$ exists. If $\lambda_u \in (0, 1]$, then X_1 and X_2 are said to show upper tail dependence or extremal dependence in the upper tail; if $\lambda_u = 0$, they are asymptotically independent in the upper tail. Analogously, the *coefficient of lower tail dependence* can be defined

$$\lambda_l := \lambda_l(X_1, X_2) = \lim_{q \to 0^+} P(X_2 \leqslant F_{X_2}^{-1}(q) \mid X_1 \leqslant F_{X_1}^{-1}(q))$$

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

Example 3.1 (Logistic model)**.** An example of a non-trivial asymptotically joint distribution is the logistic model in the bivariate extreme value, see Tawn (1988) and Longin and Solnik (2001), which for unit Fréchet margins has

$$P(S \leqslant s, T \leqslant t) = exp\{-(s^{-1/\alpha} + t^{-1/\alpha})^\alpha\} \tag{3.4}$$

with $0 < \alpha \leq 1$. When $\alpha = 1$, the variables are exactly independent and $\chi = 0$. When $\alpha < 1, \chi = 2 - 2^\alpha$ and the variables are asymptotically dependent to a degree depending on α.

Generally, when $\chi = 0$ the two random variables are not necessarily also exactly independent. For example, if the dependence structure is that of a bivariate normal random variable with any value for the correlation coefficient less than one, then $\chi = 0$ (Sibuya, 1959). These variables are not exactly independent if $\rho \neq 0$.

When exact independence of the variables is rejected, traditional multivariate extreme value methods assume $P(T > s \mid S > s) = \chi > 0$ for all large s. However, if the true distribution of the variables is asymptotically independent, this approach will overestimate $P(T > s \mid S > s)$ and all other probabilities of joint extreme events. The degree of bias will depend on the difference between the estimated χ and the true value of $P(T > s \mid S > s)$, which is determined by the value of s and the rate at which $P(T > s \mid S > s) \to 0$ as $s \to \infty$.

3.2.2 An alternative measure of dependence $\bar{\chi}$

Ledford and Tawn (1996, 1997); Bruun and Tawn (1998); Bortot and Tawn (1998) have provided a range of extremal dependence models, derived from a different form of multivariate limit theory, that describe dependence but have $\chi = 0$. Although the random variables are asymptotically independent in this case, different degrees of dependence are attainable at finite levels of s. Based on these studies, Coles et al. (1999) suggest a new dependence measure.

Definition 3.3 (**$\bar{\chi}$ asymptotic independence measure**). Let be (X, Y) a bivariate rv, and let be (S, T) the transformation into unit Fréchet marginals.

$$\bar{\chi} = \lim_{s\to\infty} \frac{2\log P(S > s)}{\log P(S > s, T > s)} - 1 \tag{3.5}$$

and $-1 < \bar{\chi} \leqslant 1$, is an appropriate *measure of asymptotic independence* as it gives the rate that $P(T > s \mid S > s) \to 0$ approaches 0. For *perfect dependence*, $P(T > s, S > s) = P(S > s)$ and $\bar{\chi} = 1$. For *independence*, $P(T > s, S > s) = [P(S > s)]^2$, so $\log P(S > s, T > s) = 2\log P(S > s)$ and $\bar{\chi} = 0$. Hence values of $\bar{\chi} > 0$, $\bar{\chi} = 0$ and $\bar{\chi} < 0$ correspond, respectively, to when (S, T) are *positively associated*, *exactly independent*, and *negatively associated in the extremes*. For the bivariate normal dependence structure $\bar{\chi} = \rho$ is the correlation coefficient.

The pair of dependence measures $(\chi, \bar{\chi})$ together provide all the necessary information for characterising the form and the degree of extremal dependence.

- For asymptotically dependent variables, $\bar{\chi} = 1$ with the degree of asymptotic dependence given by $\chi > 0$;
- For asymptotically independent variables, $\chi = 0$ with the degree of dependence given by $\bar{\chi}$.

It is important to test if $\bar{\chi} = 1$ first, before drawing conclusions about asymptotic dependence based on estimates of χ.

One could wrongly think that asymptotic independence means independence also at a sub-asymptotic level, but this is not the case. Also rvs positively associated in extremes and rvs negatively associated in extremes, lead to asymptotic independence.

The only degree of dependence between moderately large values of each rv, that is strong enough to effectively becoming asymptotic dependence ($\chi > 0$), is perfect dependence in extremes ($\bar{\chi} = 1$).

3.3 Estimation and statistical inference

For estimating $\chi > 0$ and $\bar{\chi}$, weak assumptions are required (the first three of which are essentially the same as those required when estimating the univariate tail behaviour by extreme value models):

(i) The joint distribution (S, T) has a joint tail behaviour that is bivariate regularly varying, satisfying the conditions of Ledford et al. (1998);

(ii) Both $\chi > 0$ and $\bar{\chi}$ are limit properties, so it is necessary to assume that the sample characteristics of the empirical joint distribution, above some selected threshold, reflect the limiting behaviour;

(iii) The series has sufficient independence over time for the sample characteristics to converge to the population characteristics $\chi > 0$ and $\bar{\chi}$;

(iv) The marginal variables can be transformed to identically distributed Fréchet variables.

With the above assumptions, we use results in Ledford and Tawn (1996, 1997); Ledford et al. (1998) to estimate $\chi > 0$ and $\bar{\chi}$.

3.3.1 Coefficient of tail dependence

Ledford and Tawn (1996, 1997); Ledford et al. (1998) showed that under general regularity conditions the joint survivor function of the random vector (S, T) satisfies

$$P(T > s, S > s) \sim \mathcal{L}(s) s^{-1/\eta} \qquad \text{as} \qquad s \to \infty \tag{3.6}$$

$0 < \eta \leqslant 1$ is a constant and $\mathcal{L}(s)$ is a slowly varying function. The constant $\eta \in (0, 1]$ is called the *coefficient of tail dependence* and describes that nature of the dependence in the tails. A function $\mathcal{L}(y)$ is said to be slowly varying (at ∞) if

$$\lim_{y \to \infty} \frac{\mathcal{L}(ty)}{\mathcal{L}(y)} \to 1 \qquad \text{for all} \qquad t > 0$$

Notice that (3.6) implies that

$$P(T > s \mid S > s) \sim \mathcal{L}(s) s^{1-1/\eta} \qquad \text{as} \qquad s \to \infty \tag{3.7}$$

$0 < \eta \leqslant 1$. Therefore, perfect dependence corresponds to

$$\eta = 1 \text{ and } \mathcal{L}(y) = 1,$$

and exact independence corresponds to

$$\eta = 1/2 \text{ and } \mathcal{L}(y) = 1,$$

When $0 < \eta < 1/2$ the marginal components are negatively associated in the tail since the survivor function decays faster than when they are exactly independent. The slowly varying function $\mathcal{L}$ describes the strength of the dependence for a given level of η. In view of (3.7), the definition of χ in (3.3) implies that

$$\chi > 0 \text{ if and only if } \eta = 1 \text{ and } \mathcal{L}(y) \to c > 0.$$

From the representation (3.6) it follows that

$$\begin{aligned}
\bar{\chi} &= \lim_{s\to\infty} \frac{2\log P(S > s)}{\log P(S > s \mid T > s)} - 1 \\
&= \lim_{s\to\infty} \frac{2\log y^{-1}}{\log(\mathcal{L}(s)s^{-1/\eta})} - 1 \\
&= \lim_{s\to\infty} \frac{-2}{\frac{\log \mathcal{L}(s)}{\log y} - \frac{1}{\eta}} - 1 \\
&= 2\eta - 1
\end{aligned} \tag{3.8}$$

since $\lim_{s\to\infty} \frac{\log \mathcal{L}(s)}{\log y} = 0$ (see Proposition 0.8 of Resnick, 1987). Therefore, the coefficient of tail dependence η and the measure $\bar{\chi}$ are linearly related.

The relationship between the dependence measures χ, $\bar{\chi}$, η and $\mathcal{L}(y)$ may be summarized as follows:

$$\bar{\chi} = 2\eta - 1 \tag{3.9}$$

and

$$\chi = \begin{cases} c & \text{if} \quad \bar{\chi} = 1 \quad \text{and} \quad \mathcal{L}(s) \to c > 0 \\ 0 & \text{if} \quad \bar{\chi} = 1 \quad \text{and} \quad \mathcal{L}(s) \to 0 \\ 0 & \text{if} \quad -1 < \bar{\chi} < 1. \end{cases} \tag{3.10}$$

There is the possibility that $\eta = 1$ and $\mathcal{L}(s) \to 0$ as $s \to \infty$ leading to asymptotic independence. This boundary case cannot be recognised from data as the slowly vary function cannot be identified other than as a constant, and misspecification of the dependence structure in this situation is unlikely to be important.

In summary, the pair $(\bar{\chi}, \chi)$ provide an overview of extremal dependence. When the variables are asymptotically independent and $\chi = 0$, the value of $\bar{\chi}$ provides a measure of the strength of dependence. When the variables are asymptotically dependent $\bar{\chi} = 1$ and χ (through c) provides a measure of the strength of asymptotic dependence. Thus, we focus on inference for η and $\lim_{s\to\infty} \mathcal{L}(s)$, treating the slowly varying function as constant over some threshold u, i.e. $\mathcal{L}(s) = c$ for $s > u$.

In particular, when (S, T) have a Gaussian dependence structure with correlation ρ, $\chi = 0$ and $\bar{\chi} = \rho$ which provides a useful benchmark for measuring dependence across the class of asymptotically independent variables.

3.3.2 Hill estimator

In practice, is convenient to first estimate η and then estimate $\bar{\chi}$ through (3.9). If $\bar{\chi}$ is statistically indistinguishable from 1, then one estimates χ (through $\mathcal{L}(s) = c$).

To proceed with the estimation, we follow the methodology suggested by Poon et al. (2003, 2004) which is based on results from univariate EVT. The estimation of χ and $\bar{\chi}$ is based on fitting a univariate model for exceedances of a high threshold.

However, this time we are not applying the fully parametric GPD approach, as we did in the previous chapter, but we rely on the semiparametric one. A number of diagnostic techniques exist for threshold selection (reviewed in Section 2.7.2 Chapter 2). An important result of Gnedenko (1943) shows that the tail of a univariate heavy-tailed variable Z above a high threshold u satisfies the following relation.

Theorem 3.1 (Gnedenko (1943) Fréchet MDA)**.** *For heavy tailed distributions $\xi > 0$,*

$$F \in MDA(H_\xi) \Leftrightarrow \bar{F}(x) = x^{-1/\xi}\mathcal{L}(x) \tag{3.11}$$

for some function $\mathcal{L}$ slowly varying at ∞.

This means that distributions giving rise to the Fréchet case are distributions with tails that are regularly varying functions with a negative index of variation. Their tails decay essentially like a power function and the rate of decay $\alpha = 1/\xi$ is often referred to as the *tail index* of the distribution.

These distributions are the most studied distributions in EVT and they are of particular interest in financial applications because they are heavy-tailed distributions with infinite higher moments.

If X is a non-negative rv whose cdf F is an element of $MDA(H_\xi)$ for $\xi > 0$, then we have already shown that $E[X^k] = \infty$ for $k \geqslant 1/\xi$.

This form of the survivor function is usually assumed to hold for x large ($x > u$) and $\mathcal{L}(x) = c$ for some constant c. Estimation of the shape parameter is performed by using the semi-parametric estimator of Hill et al. (1975).

We make the modelling assumption that Equation (3.6) is an equality for $s > u$. Inference follows, using univariate extreme value techniques, by identifying that if $Z = \min(S, T)$ then

$$\begin{aligned} P(Z > z) = P(\min(S,T) > z) = P(S > z, T > z) &= \mathcal{L}(z)z^{-1/\eta} \\ &= cz^{-1/\eta}, \text{ for } z > u \end{aligned} \tag{3.12}$$

for some high threshold u. From the Eq. (3.12) and the univariate form in Eq. (3.11), it can be seen that $1/\eta$ is the tail index of the univariate variable Z, and so can be estimate using the Hill estimator.

3.3.3 $\bar{\chi}$ estimation

Given a sample of i.i.d. observation of the bivariate random vector (S, T)

$$(Y_{1,i}, Y_{2,i}), i = 1, \ldots, n,$$

Let be

$$Y_{\min}^{(1)} \geqslant Y_{\min}^{(2)} \geqslant \cdots \geqslant Y_{\min}^{(n)}$$

the corresponding order statistics of

$$Y_{\min,i} = \min(Y_{1,i}, Y_{2,i}), i = 1, \ldots, n,$$

Conditional on a number of exceedances over the threshold u equal to k ($\{K = k\}$), the joint density of the order statistics $(Y_{\min}^{(1)}, Y_{\min}^{(2)}, \ldots, Y_{\min}^{(j)})$ is given by

$$f_{(Y_{\min}^{(1)}, Y_{\min}^{(2)}, \ldots, Y_{\min}^{(j)})}(y_1, \ldots, y_k) = \frac{n!}{(n-k)!}(1 - c y_k^{1/\eta})^{n-k} c^k \eta^{-k} \prod_{i=1}^{k} y_i^{-(1/\eta+1)}, \tag{3.13}$$

for $u < y_k \leqslant \cdots \leqslant y_1$. Maximization of the conditional log-likelihood function associated with (3.13) yields the following estimators due to Hill et al. (1975)

$$\hat{\eta} = \frac{1}{k}\sum_{i=1}^{k} \log Y_{\min}^{(i)} - \log u, \tag{3.14}$$

$$\hat{c} = \frac{k}{n} u^{1/\hat{\eta}} \tag{3.15}$$

Estimation of the extremal dependence measures $\bar{\chi}$ and χ is now straightforward. Using (3.9), we estimate $\bar{\chi}$ by

$$\hat{\bar{\chi}} = 2\hat{\eta} - 1 = 2\left(\frac{1}{k}\sum_{i=1}^{k} \log Y_{\min}^{(i)} - \log u\right) - 1 \tag{3.16}$$

Under certain regularity conditions on the tail of F[7] and conditions on the rate that $k = k(n) \to \inf$, $k/n \to 0$ as $n \to \infty$, the estimator $\hat{\eta}$ is asymptotically normal such that

$$\sqrt{k(n)}(\hat{\eta} - \eta) \xrightarrow{d} \mathcal{N}(0, \eta^2) \tag{3.17}$$

where $\xrightarrow{d}$ denotes convergence in distribution. The asymptotically normality of $\bar{\chi}$ is ensured by results in Smith (1987). A natural estimator of the variance of $\hat{\bar{\chi}} = 2\hat{\eta} - 1$ is then given by

$$\widehat{Var}(\hat{\bar{\chi}}) = \frac{4\hat{\eta}^2}{k} = \frac{(\hat{\bar{\chi}} + 1)^2}{k} \tag{3.18}$$

[7] These regularity conditions involve a second-order regular variation assumption on the tail of F. See Sec. 6.4 of Embrechts et al. (1997) for details and references.

If $\hat{\bar{\chi}}$ is significantly less than 1 (i.e., if $\hat{\bar{\chi}} + 1.96\sqrt{\mathrm{Var}(\hat{\bar{\chi}})} < 1$) then we infer the variables to be asymptotically independent and take $\chi = 0$.

3.3.4 χ estimation

Only if there is no significant evidence to reject $\bar{\chi} = 1$, hence $\hat{\bar{\chi}}$ is statistically indistinguishable from one, we reject asymptotic independence and estimate χ.
The estimation of χ is made under the condition that $\bar{\chi} = \eta = 1$. Taking $\eta = 1$ in (3.15), together with (3.10) gives the estimator

$$\hat{\chi} = \hat{c} = \frac{k}{n}u \tag{3.19}$$

An estimator of $\mathrm{Var}(\hat{\chi})$ is available via the properties of the maximum likelihood estimator. Let $l(c; \eta = 1, Y_1, \ldots, Y_k)$ be the logarithm of the likelihood associated with (3.13) where we set $\eta = 1$. $1/\mathrm{Var}(\hat{c})$ equals the Fisher information of the parameter c and is given by

$$-\frac{\partial^2}{\partial c^2} l(c; \eta = 1, Y_1, \ldots, Y_k) = \mathbb{E}\left(\frac{c^2(c - Y_k)^2}{c^2 n 2ckY_k + kY_k^2}\right) \tag{3.20}$$

But recall that we are conditioning on there being k exceedances of u. Letting $Y_k = u$ in (3.20) and replacing c by its estimator $\hat{c} = ku/n$, a natural estimator of the variance of $\chi = c$ is given by

$$\widehat{\mathrm{Var}}(\hat{\chi}) = \frac{u^2 k(n-k)}{n^3} \tag{3.21}$$

Remark 3.1 (Threshold choice). Key to the success of the above methodology is the choice of u (or k), above which we assume the model of the tail holds. The choice of k corresponds to the choice of the high threshold u in the threshold excess modelling. This determines how many order statistics, $(Y^{(1)}_{\min}, Y^{(2)}_{\min}, \ldots, Y^{(j)}_{\min})$, are used in the estimation of the η and c in (3.14) and (3.15) respectively.

It is a delicate matter which typically relies on certain graphical techniques or sampling methods to determine the optimal choice of threshold, as we have already discussed in Section 2.7.2 Chapter 2[8]. Ideally, the choice of threshold u should be done for every pair of markets Y_1 and Y_2. Since in our application (Chapter 4) we will consider 12 markets, thus 66 different possible pairs of them, we make the pragmatic choice of u as the 95% quantile of the empirical distribution of $Y_{min,i}, i = 1, \ldots, n$.

First of all because is the typical choice for the application of EVT in literature, and secondly to be consistent with the choice we made to solve the optimal portfolio allocation problem.

[8]See Sec. 6.4 in Embrechts et al. (1997) and references therein for a detailed discussion on the choice of threshold.

Remark 3.2 (i.i.d. random variables). The whole precedent discussions apply to independent variables. When the variables are dependent, the statistical approaches for analysing maxima are unchanged and the limit distribution of the maximum is also a generalized extreme value distribution (see Section 7.1.3 in McNeil et al., 2005). Unlike the maximum over intervals method, temporal dependence due the use of threshold method adds some complications.

Ignoring the dependence and applying the methods as if the data were independent will lead to unbiased estimators but with standard errors that are too small (Kearns and Pagan, 1997). Several approaches may be used to overcome this problem: declustering of the exceedances of the threshold to produce approximately independent data (Davison and Smith, 1990), or adjusting the standard error within maximum likelihood framework (see Coles and Walshaw, 1994).

Chapter 4

Portfolio of International Equity Indices

This final chapter outlines the application of the methodology presented in Chapters 2 and 3 to a portfolio of international equity market indices.

The aim is to give an empirical perspective of the modelling framework and dependence analysis previously introduced. The choice of the asset universe is crucial for this purpose. Indeed, in our setting, performing an analysis to identify the correct extremal dependence structure of the portfolio's constituents, means studying the dependence between countries during extreme economic periods.

We link the two main topics, portfolio optimisation and extreme dependence analysis, trying to understand how the asymptotic dependence or independence of pairs of markets, may impact the optimal asset allocation problem.

The software used in this dissertation, to perform the extremal dependence analysis and to solve the EVT-based portfolio optimisation problem, was entirely and independently written by the author in MATLAB[1].

4.1 Empirical data

The methodology presented in the two previous chapter is tested on real data. This final chapter is dedicated to the presentation of analyses and results, when the EVT-based model is applied to a portfolio composed of international equity indices.

In the empirical application we consider the assets listed in Table 4.1. These are equity indices of twelve large capitalization international markets. We use *Thomson Reuters*

[1] We did not make use of any of the available software systems for EVT (like for instance EVIS, Extreme Values In S-Plus, developed at ETH Zurich, or XTREMES developed by Rolf Reiss and Michael Thomas at the University of Siegen in Germany.

Datastream as data source. Datastream produces many of their own stock indices, including total market indices representing all the stocks trading in a country's stock market.

We refer to these *Total market country indices* to perform our analyses. To retrieve them, the identifying codes are composed by: "Totmk" + "Country code". For instance **TOTMKUK** data type RI, represents the total return for the Total UK Market. We refer to data type RI for all the countries.

The total return index is a type of equity index that tracks both the capital gains of a group of stocks over time, and assumes that any cash distributions, such as dividends, are reinvested back into the index. Looking at an index's total return displays a more accurate representation of the index's performance. By assuming dividends are reinvested, the index effectively account for stocks that do not issue dividends and instead, reinvest their earnings within the underlying company.

Market	Min	Max	$Mean$	$stdev$	$skew$	$kurt$	MV
HK	-34.724	16.838	0.043	1.637	-1.444	31.762	1,994,437
JP	-16.000	12.000	0.031	1.356	-0.012	6.089	4,770,875
AU	-26.157	8.739	0.032	1.385	-1.152	20.102	1,133,719
BG	-11.146	10.204	0.034	1.180	-0.099	6.229	388,591
CN	-12.660	9.988	0.031	1.115	-0.576	12.087	1,661,925
FR	-10.142	11.234	0.034	1.328	-0.115	5.960	1,918,682
BD	-11.733	17.659	0.032	1.279	0.044	8.718	1,714,272
IT	-12.511	11.914	0.033	1.535	-0.106	5.135	518,132
NL	-10.852	10.727	0.036	1.244	-0.147	7.438	608,278
SW	-10.497	9.465	0.039	1.082	-0.206	6.039	1,442,935
UK	-13.528	12.543	0.032	1.217	-0.273	9.313	3,127,606
US	-18.705	11.518	0.039	1.081	-0.642	17.659	22,948,820

Table 4.1: Descriptive statistics for the raw data based on 9,565 observations of daily simple periodic returns (expressed as %) from international equity markets for the period from 02-Jan-1980 to 30-Aug-2016. The kurtosis is defined so that the kurtosis of the normal distribution is zero. Market value is reported in millions of US dollars.

Table 4.1 summarise the descriptive statistics of daily returns from 12 international equity markets that are taken into account. The simple returns data is calculated from daily closing price (adjusted for dividends and splits) time series of the total market indices covering the period from 02-Jan-1980 to 30-Aug-2016. Hence, the dataset consists of 9,565 data points for each market index time series.

These markets exhibit a negative skewness (except for Germany BD) and positive excess kurtosis, implying an increased probability of observing large losses. We have widely discussed how much such a departure from normality is critical to the safety-first investor. Therefore, a preliminary analysis of data shows the validity of the stylised facts, not just from a theoretical perspective, but also on the true data considered in our empirical experiment.

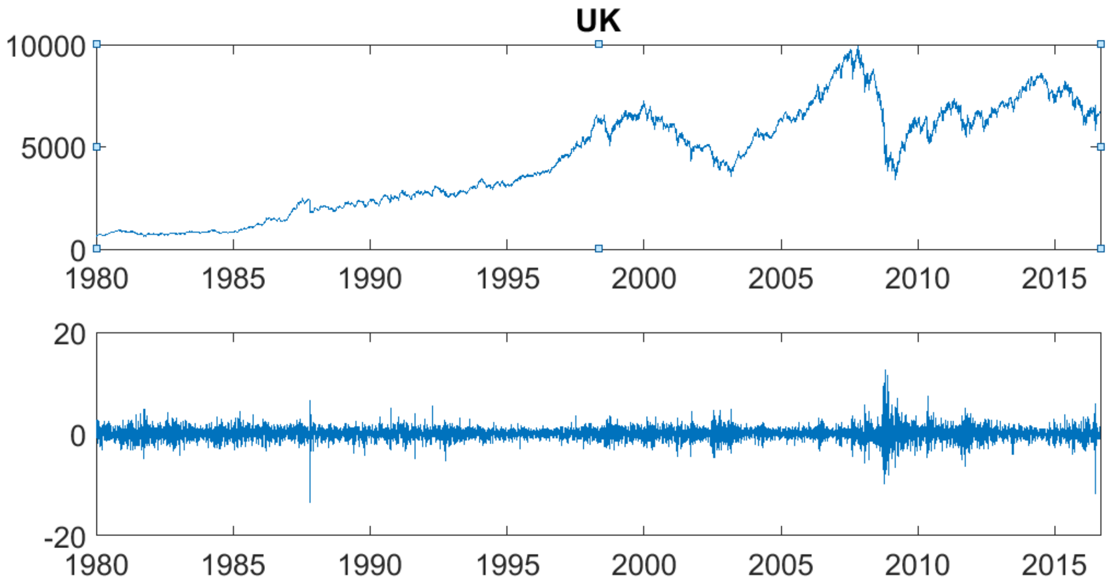
All the variables involved are downloaded from the data source in the same currency, US Dollars, as in our framework we are not modelling the dynamic of exchange rates. In this empirical application, we use the following definition of kurtosis such that the normal distribution has a kurtosis of zero.

$$K(X) = \frac{\mathbb{E}(X - \mu_X)^4}{(\mathrm{Var}X)^2} - 3$$

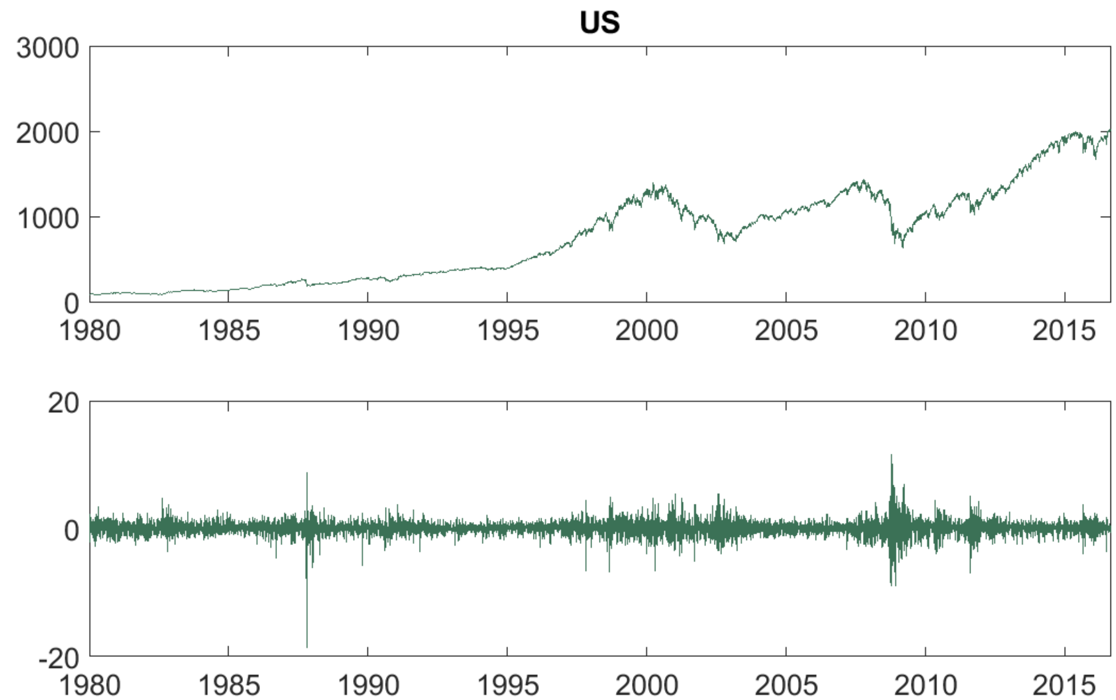
Heavy tailed distribution, thus will lead to positive value of kurtosis. Looking at Table 4.1, we can also notice that the Italian and Hong Kong stock markets are the most volatile (higher *stdev*), while Switzerland and US are the less risky in the sample, when referring to variance as risk measure. This evidence will be visible again, in the plots of the prices indices and of returns, since safer countries will display respectively a smoother price line and a narrower return series.

The price and return series for four out of twelve markets are displayed in Figure 4.1: two European, United Kingdom and Italy, one American, US, and finally one Asian market, Japan. The effects of the main financial crises can clearly be seen in all of them: the 1987 stock market crash; then the Asian financial crisis triggered in July 1997, the hedge-fund crisis during 1997-1998 and subprime crisis (2007-2009). In addition, the consequences of Brexit referendum in June 2016, are visible both in the UK market, plot 4.1a, and in the Italian one, 4.1d. In particular, the latter market observed, on "the day after" (24-Jun-2016), its minimum value on the whole time series considered.

A collection of the dates on which the minimum and maximum returns values of each daily market return time series were reached, are shown in Table 4.2. Canada, France and Netherlands were mainly hit, both in terms of losses and gains, during the credit crisis in 2008. The Asian crisis made the Asian markets within our sample, Hong Kong and Japan, reaching their maximum gain on the whole period considered.

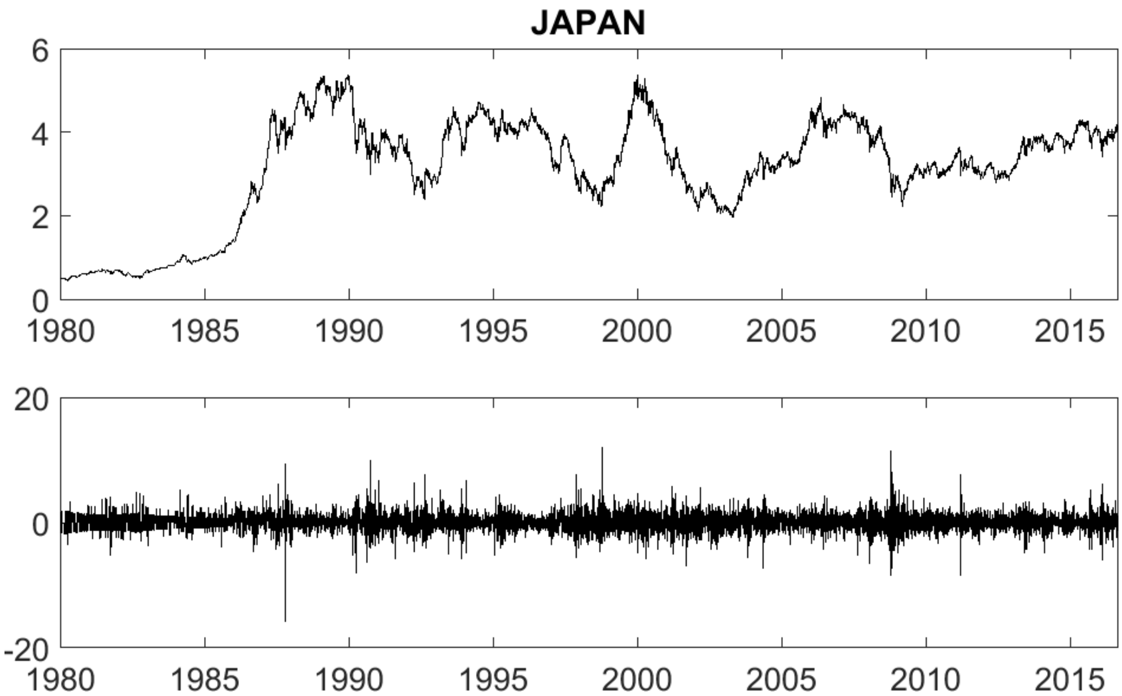


(a)

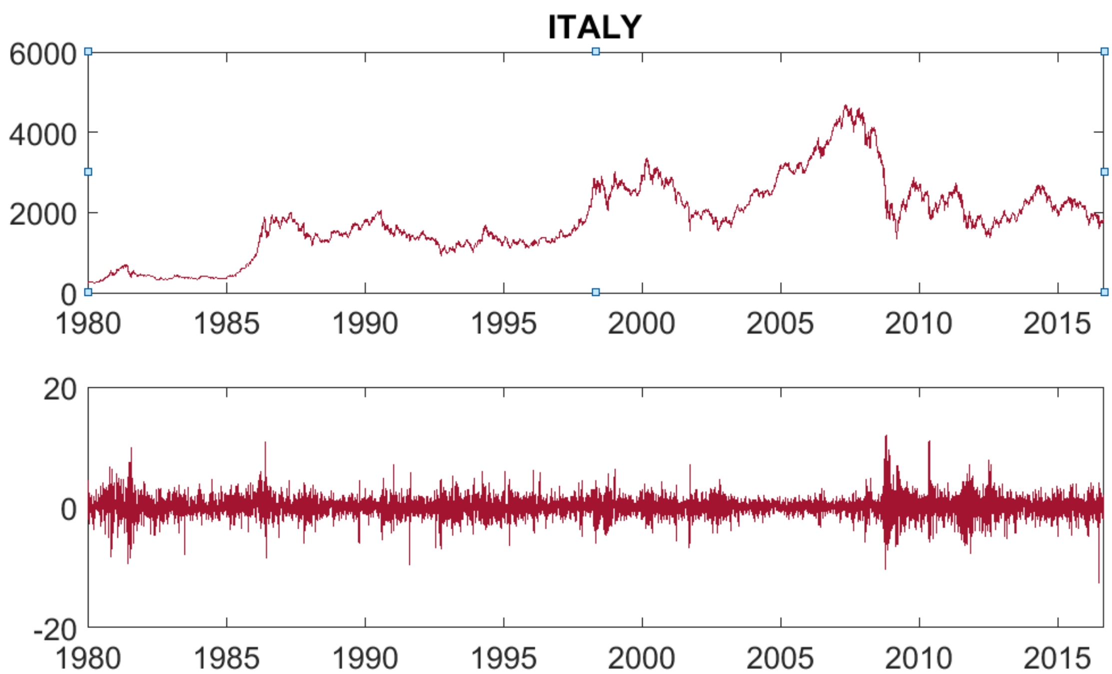


(b)

Figure 4.1: UK market-TOTMKUK and US market-TOTMKUS. The upper graphs show the closing share prices on the sampling period 02-Jan-1980 to 30-Aug-2016. The lower panel show the daily returns (expressed as %).



(c)



(d)

Figure 4.1: Japanese market-TOTMKJP and Italian Market-TOTMKIT. The upper graphs show the closing share prices on the sampling period 02-Jan-1980 to 30-Aug-2016. The lower panel show the daily returns (expressed as %).

Table 4.2: Dates on which minimum and maximum values of each daily market return time series were reached over the period 02-Jan-1980 to 30-Aug-2016.

Market	Min	Max	Market	Min	Max
HK	26/10/1987	29/10/1997	**BD**	19/08/1991	28/10/2008
JP	20/10/1987	07/10/1998	**IT**	24/06/2016	29/10/2008
AU	20/10/1987	13/10/2008	**NL**	06/10/2008	24/11/2008
BG	19/10/1987	10/05/2010	**SW**	19/10/1987	13/10/2008
CN	20/11/2008	14/10/2008	**UK**	20/10/1987	29/10/2008
FR	06/10/2008	24/11/2008	**US**	19/10/1987	13/10/2008

4.1.1 GPD estimation on the marginals

Following all the discussions in Chapter 2, and for being consistent throughout the analysis, we have chosen to set our threshold at the 95% percentile of the empirical distribution for all markets.

Thus we fit a GPD distribution to excess losses above this threshold, using formulas and methodology presented in Sections 2.4 and 2.5 in Chapter 2, via maximum likelihood. Henceforth, we are considering negative returns for each market, so that losses will be positive e we perfectly lie within the previously presented framework[2].

Given our sample of multivariate observations of the random vector representing negative returns of the 12 assets in our universe

$$\tilde{\boldsymbol{X}}_s = (\tilde{X}_{1,s}, \tilde{X}_{2,s}, \dots, \tilde{X}_{12,s}) \ , \ s \in \{1, \dots, n\}$$

We assume that the realisations of each component are drawn from a sequence of identically distributed rvs, with unknown underlying distribution $F^{(i)}, i = 1, \dots, 12$ and each of those lie in the domain of attraction of an extreme value distribution H_ξ for some ξ.

Therefore, all the results presented in Chapter 2 hold and we model the tails of every asset's return distribution in our sample $F^{(i)}(x), \ i = 1, \dots, 12, x \geq u$, by means of the EVT fully parametric GPD model.

Table 4.3, provides estimates of the shape parameter ξ for the left (Gains) and right (Losses) tails of the 12 international equity markets we consider. The shape parameter ξ is positive in all markets and in both tails. This is coherent with the heavy tailedness of the returns distributions showed by the kurtosis values calculated on the raw return series. For financial data we expect $\xi \in (0, 1)$, so that the tail is heavier than it would be for

[2] We apply the EVT based model always on the right tail of the distribution (positive values). Therefore, when analysing the Gain Tail, the sign of the negative returns will change accordingly. For the tails, losses and gains are reported as positive.

exponential decay distributions ($\xi > 0$), and the mean exists ($\xi < 1$). This is confirmed by our findings showed in Table 4.3. The relation $\xi < 1$ guarantees that the estimated expected shortfall, which is a conditional first moment, exists for both tails.

		Loss Tail					**Gain Tail**				
Market	$F(u)$	u	ξ	se ξ	β	se β	u	ξ	se ξ	β	se β
HK	0.95	2.373	0.246	0.053	1.129	0.078	2.402	0.209	0.056	0.975	0.070
JP	0.95	2.041	0.122	0.047	0.848	0.055	2.128	0.161	0.051	0.808	0.055
AU	0.95	2.017	0.248	0.051	0.859	0.058	2.071	0.195	0.056	0.758	0.054
BG	0.95	1.776	0.180	0.055	0.762	0.054	1.790	0.167	0.055	0.779	0.055
CN	0.95	1.604	0.229	0.055	0.836	0.059	1.606	0.232	0.055	0.669	0.047
FR	0.95	2.020	0.122	0.052	0.945	0.065	1.987	0.202	0.052	0.765	0.053
BD	0.95	1.976	0.143	0.054	0.844	0.059	1.934	0.214	0.054	0.711	0.050
IT	0.95	2.330	0.093	0.055	1.149	0.082	2.398	0.206	0.055	0.854	0.060
NL	0.95	1.856	0.144	0.054	0.925	0.065	1.847	0.204	0.054	0.781	0.055
SW	0.95	1.634	0.178	0.052	0.682	0.047	1.697	0.232	0.056	0.564	0.040
UK	0.95	1.828	0.193	0.050	0.790	0.053	1.807	0.198	0.051	0.710	0.048
US	0.95	1.597	0.256	0.055	0.683	0.048	1.611	0.221	0.058	0.673	0.049

Table 4.3: Parameter estimates of the generalized Pareto distribution for the loss and gain tails. "se" refers to standard errors. The threshold u was picked to correspond to the 0.95 quantile of the empirical distribution and is expressed as %. For both tails, losses and gains are reported as positive.

The GPD models we have fitted on each marginals show that all the equity market indices have finite variance as $\xi_i < 1/2, i = 1, \ldots, 12$ in both tails. All the markets have heavy-tailed Fréchet distributions with a finite fourth moment, excluding the US loss tail. Furthermore, we can notice that the GPD models for the loss tail of both Hong Kong and Australian markets, are close to having an infinite fourth moment. The obtained results agreed with one of the stilized facts, listed in Section 1.3.2 Chapter 1, stating that "the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied".

Looking at the threshold values (on the loss tail) we can say that Italy and Hong Kong, having the highest u values, are again the riskiest markets also considering a quantile based risk measure. Indeed, the u in our setting, corresponds to the value at risk at 95% confidence level calculated on the empirical distribution. US, Canada and Switzerland are the safest in this perspective.

Considering the results, we cannot generalise that the loss tail is heavier than the gain one for all indices. Actually for European countries, France, Germany, Italy and Netherlands, the opposite case seems to hold.

Now, suppose the risk manager wants to obtain value at risk and expected shortfall estimates of the returns on the indices at some high quantiles (higher than the threshold u). Parametric formulas are available for those risk measures, under the GPD modelling we are referring to (Section 2.5 in Chapter 2).

Loss tail risk measures					
Market	$\text{VaR}_{99\%}$	$\text{ES}_{99\%}$	Market	$\text{VaR}_{99\%}$	$\text{ES}_{99\%}$
HK	4.601	6.825	**BD**	3.505	4.747
JP	3.549	4.725	**IT**	4.325	5.796
AU	3.717	5.421	**NL**	3.531	4.892
BG	3.199	4.440	**SW**	2.906	4.011
CN	3.231	4.799	**UK**	3.319	4.654
FR	3.702	5.014	**US**	2.958	4.344

Table 4.4: VaR and ES values (expressed as %) at a confidence level $\alpha = 99\%$ for every equity market index in our universe. The focus is on the right tail as we are concerned by the consequences related to downside risk.

Table 4.4 shows the values of the Var and ES at a confidence level $\alpha = 99\%$ for every equity market index in our universe. Obviously, in this case we focus only on the right tail (loss) of the distribution as we are concerned by the consequences associated with downside risk. Looking at estimated VaR and ES values, we observe that the conclusions in terms of exposure to extreme losses, hence riskiness, among the considered countries are the same as before.

4.2 Extremal dependence analysis

This section presents the results obtained after analysing the extremal dependence structure between pairs of assets in our universe. Since the securities considered are total market index return, this would mean understanding dependence relations between countries when grouped in twos.

We focus on dependence estimation in the bivariate context, but the ideas and techniques can be extended to higher dimensions. The extension is theoretically conceivable, but often demanding from a computational viewpoint. We follow the methodology presented in Chapter 3 applying it to all the possible pairs of markets listed in Table 4.1. Hence, we consider 66 pairs in total and examine the nature of their extremal dependence for both the loss and gain tails of the distribution. Of course we aim at establishing which of the two classes of extreme value dependence, asymptotic dependence and asymptotic independence, characterise each couple.

In order to do that we refer to the pair of dependence measures $(\bar{\chi}, \chi)$, that together provide all the necessary information for characterising the form and the degree of extremal dependence. To briefly summarise their interpretation:

For asymptotically dependent variables, $\bar{\chi} = 1$, and the degree of extremal dependence is given by $\chi > 0$;

For asymptotically independent variables, $\chi = 0$, and the value of $\bar{\chi}$ provides the degree of dependence at a sub-asymptotic level. In particular, when $0 < \bar{\chi} < 1$ the markets are positively associated in the extremes, $\bar{\chi} = 0$ means independence, and finally $-1 < \bar{\chi} < 0$ negative association in the extremes.

First, we estimate $\bar{\chi}$ and its standard error (see Section 3.3.3 Chapter 3). Using the asymptotic normality of the estimator $\hat{\bar{\chi}}$ of $\bar{\chi}$, we reject asymptotic independence if $\hat{\bar{\chi}} \geqslant 1 - 1.96\sigma_{\hat{\bar{\chi}}}$. When asymptotic independence is rejected, we estimate $\bar{\chi}$ conditionally on $\bar{\chi} = \eta = 1$.

The results are presented in Table 4.6. This shows estimates of the classical dependence measure, that is the linear correlation ρ, and estimates of $\bar{\chi}$ and χ for both the tails. First, recall that if the dependence structure were Gaussian, then $\bar{\chi} = \rho$. Thus this value can be considered a benchmark.

Results shows that $\bar{\chi}_{Loss} > \rho$ for the loss tail in all 66 pairs of markets. Similar results hold for the the gain tail where again we have $\bar{\chi}_{Gain} > \rho$ for all 66 pairs of markets. This reinforces our previous conjecture that the Pearson correlation measure is a poor measure for tail dependence. Next, notice that $\bar{\chi}_{Loss} > \bar{\chi}_{Gain}$ in 55 out of the 66 markets. The dependence in the loss tail is thus generally greater than the dependence in the gain tail.

In Table 4.5 a summary of the findings regarding the extremal dependence structure is shown. In particular, it presents how many of the markets pairs show either asymptotic dependence or asymptotic independence within our sample; the results refer to both tails and are presented for different significance levels, $\alpha = 1\%$ and $\alpha = 5\%$, of the hypothesis test, with null $H_0 : \bar{\chi} = 1$. It is clear, that the dependence during bad states of the market is greater than the dependence found in very good economic periods (gain tail). Indeed, the number of couple of markets, being asymptotic dependent, is higher in the loss tail for both significance levels considered. When focussing on the loss tail and moving from a significance level of $\alpha = 5\%$ to $\alpha = 1\%$, the numbers of independent pairs drops from 32 (48% of the total) to 28 (42% of the total). Whereas, when looking at the gain tail, 61% of the total number of pairs are asymptotic independent with significance level $\alpha = 5\%$;

this percentage decreases to 52% at a significance level $\alpha = 1\%$.

While asymptotic independence between markets prevails in the gain tails, the same is not true in the loss one; where the situation is more balanced but between 52% ($\alpha = 5\%$) and 58% ($\alpha = 1\%$) of the market pairs are dependent in the very extreme part of the distribution. All the detailed results regarding the estimation of the dependence

Table 4.5: Number of markets pairs that show asymptotic dependence (first row) or asymptotic independence (second row) in the sample, looking at both tails and for different significance levels ($\alpha = 1\%$ and $\alpha = 5\%$,$H_0 : \bar{\chi} = 1$).

	Loss Tail		**Gain Tail**	
Test significance level	0.01	0.05	0.01	0.05
Asymptotic:				
Dependence	38	34	32	26
Independence	28	32	34	40

measures, and their standard errors, together with the p-values of the hypothesis tests are reported in Table 4.6. The values relative to χ are shown only for asymptotic dependent pairs of markets, i.e. for those that not significantly reject the null $H_0 : \bar{\chi} = 1$ (p-value>significance level). Notice that in red are the χ estimates for those pairs being asymptotic dependent when the test is performed at a significance level of $\alpha = 1\%$, but resulting significantly asymptotic independent when $\alpha = 5\%$; in black those whose $\bar{\chi}$ is statistically indistinguishable from one, thus asymptotic dependent, for $\alpha = 5\%$.

Recall that the χ estimate gives the probability of joint occurrence of the most extreme values. For example, an estimates of $\chi = 0.5$ in the period, would indicate that there is about a 50% chance that one country will experience the largest gain (fall) in the stock market, when the largest positive (negative) return is recorded in the other paired market. So χ is a true measure for systemic risk in international stock markets. The mean value of the χ dependence measure for the asymptotically dependent markets is 0.4545 (when test significance $\alpha = 5\%$), which corresponds to fairly strong dependence. The mean value of the χ dependence measure, for markets which are asymptotically dependent in the gain tail, is 0.4417 (when test significance $\alpha = 5\%$). This again corresponds to fairly strong asymptotic dependence.

Remark 4.1 (**MEVT and the common assumption of asymptotic dependence**). Since a non zero estimate of χ in the loss tail, corresponds to $52\% - 58\%$ of the total number of groups considered, we may conclude that the a priori modelling via MEVT,

Table 4.6: The traditional dependence measure of linear correlation ρ and extreme values theory dependence measures $\bar{\chi}$ and χ for both the Loss and Gain tails for pairs of assets in our universe. In red, the χ estimates for asymptotic dependent pairs when the hypothesis test has $\alpha = 1\%$, but resulting significantly asymptotic independent when $\alpha = 5\%$; in black those whose $\bar{\chi}$ is statistically indistinguishable from one (asymptotic dependent), for $\alpha = 5\%$.

		Loss Tail					**Gain Tail**				
Markets	ρ	$\bar{\chi}$	$\sigma_{\bar{\chi}}$	p-value ($H_0 : \bar{\chi} = 1$)	χ	σ_{χ}	$\bar{\chi}$	$\sigma_{\bar{\chi}}$	p-value ($H_0 : \bar{\chi} = 1$)	χ	σ_{χ}
HKvsJP	0.317	0.664	0.076	0.000			0.520	0.070	0.000		
HKvsAU	0.429	0.768	0.081	0.002			0.638	0.075	0.000		
HKvsBG	0.256	0.551	0.071	0.000			0.328	0.061	0.000		
HKvsCN	0.257	0.499	0.069	0.000			0.444	0.066	0.000		
HKvsFR	0.288	0.594	0.073	0.000			0.398	0.064	0.000		
HKvsBD	0.303	0.512	0.069	0.000			0.401	0.064	0.000		
HKvsIT	0.222	0.509	0.069	0.000			0.368	0.063	0.000		
HKvsNL	0.318	0.621	0.074	0.000			0.424	0.065	0.000		
HKvsSW	0.298	0.632	0.075	0.000			0.402	0.064	0.000		
HKvsUK	0.312	0.587	0.073	0.000			0.470	0.067	0.000		
HKvsUS	0.168	0.432	0.065	0.000			0.472	0.067	0.000		
JPvsAU	0.392	0.744	0.080	0.001			0.554	0.071	0.000		
JPvsBG	0.252	0.472	0.067	0.000			0.407	0.064	0.000		
JPvsCN	0.179	0.393	0.064	0.000			0.387	0.063	0.000		
JPvsFR	0.249	0.456	0.067	0.000			0.352	0.062	0.000		
JPvsBD	0.255	0.412	0.065	0.000			0.370	0.063	0.000		
JPvsIT	0.190	0.470	0.067	0.000			0.324	0.061	0.000		
JPvsNL	0.260	0.481	0.068	0.000			0.313	0.060	0.000		
JPvsSW	0.310	0.509	0.069	0.000			0.497	0.068	0.000		
JPvsUK	0.263	0.437	0.066	0.000			0.465	0.067	0.000		
JPvsUS	0.059	0.360	0.062	0.000			0.308	0.060	0.000		
AUvsBG	0.392	0.807	0.083	0.010			0.858	0.085	0.047	0.299	0.013
AUvsCN	0.375	0.752	0.080	0.001			0.751	0.080	0.001		
AUvsFR	0.396	0.814	0.083	0.013	0.324	0.014	0.742	0.080	0.001		
AUvsBD	0.389	0.804	0.083	0.009			0.645	0.075	0.000		
AUvsIT	0.349	0.733	0.079	0.000			0.639	0.075	0.000		
AUvsNL	0.437	0.832	0.084	0.023	0.343	0.015	0.731	0.079	0.000		
AUvsSW	0.418	0.850	0.085	0.038	0.321	0.014	0.737	0.079	0.000		

(*Table* 4.6 *Continued*). The traditional dependence measure of linear correlation ρ and extreme values theory dependence measures $\bar{\chi}$ and χ for both the Loss and Gain tails for pairs of assets in our universe. In red, the χ estimates for asymptotic dependent pairs when the hypothesis test has $\alpha = 1\%$, but resulting significantly asymptotic independent when $\alpha = 5\%$; in black those whose $\bar{\chi}$ is statistically indistinguishable from one (asymptotic dependent), for $\alpha = 5\%$.

		Loss Tail					**Gain Tail**				
Markets	ρ	$\bar{\chi}$	$\sigma_{\bar{\chi}}$	p-value ($H_0 : \bar{\chi} = 1$)	χ	σ_{χ}	$\bar{\chi}$	$\sigma_{\bar{\chi}}$	p-value ($H_0 : \bar{\chi} = 1$)	χ	σ_{χ}
AUvsUK	0.446	0.878	0.086	0.079	0.339	0.015	0.777	0.081	0.003		
AUvsUS	0.142	0.599	0.073	0.000			0.554	0.071	0.000		
BGvsCN	0.434	0.884	0.086	0.088	0.346	0.015	0.893	0.087	0.108	0.303	0.014
BGvsFR	0.717	1.000	0.093	0.500	0.541	0.024	0.963	0.090	0.340	0.494	0.022
BGvsBD	0.683	0.971	0.090	0.374	0.513	0.023	0.910	0.087	0.152	0.474	0.021
BGvsIT	0.586	0.874	0.086	0.070	0.438	0.020	0.832	0.084	0.023	0.397	0.018
BGvsNL	0.740	1.000	0.095	0.500	0.549	0.024	1.000	0.092	0.500	0.488	0.022
BGvsSW	0.698	1.000	0.093	0.500	0.495	0.022	1.000	0.092	0.500	0.451	0.020
BGvsUK	0.644	1.000	0.094	0.500	0.465	0.021	1.000	0.093	0.500	0.411	0.018
BGvsUS	0.335	0.853	0.085	0.042	0.301	0.013	0.814	0.083	0.013	0.280	0.012
CNvsFR	0.512	0.910	0.087	0.150	0.376	0.017	0.851	0.085	0.039	0.341	0.015
CNvsBD	0.498	0.963	0.090	0.340	0.376	0.017	0.897	0.087	0.119	0.336	0.015
CNvsIT	0.414	0.766	0.081	0.002			0.709	0.078	0.000		
CNvsNL	0.550	0.956	0.089	0.310	0.407	0.018	0.878	0.086	0.077	0.364	0.016
CNvsSW	0.461	1.000	0.092	0.500	0.334	0.015	0.731	0.079	0.000		
CNvsUK	0.557	0.982	0.091	0.421	0.402	0.018	0.962	0.090	0.335	0.343	0.015
CNvsUS	0.662	0.990	0.091	0.458	0.487	0.022	0.899	0.087	0.121	0.434	0.019
FRvsBD	0.767	1.000	0.095	0.500	0.574	0.026	0.974	0.090	0.388	0.549	0.024
FRvsIT	0.646	1.000	0.094	0.500	0.489	0.022	0.956	0.089	0.312	0.437	0.020
FRvsNL	0.810	1.000	0.096	0.500	0.609	0.027	1.000	0.092	0.500	0.577	0.026
FRvsSW	0.741	1.000	0.092	0.500	0.530	0.024	0.988	0.091	0.446	0.484	0.022
FRvsUK	0.727	1.000	0.093	0.500	0.528	0.024	0.953	0.089	0.299	0.499	0.022
FRvsUS	0.406	0.916	0.088	0.169	0.335	0.015	0.833	0.084	0.023	0.312	0.014
BDvsIT	0.618	0.976	0.090	0.394	0.472	0.021	0.894	0.087	0.110	0.419	0.019
BDvsNL	0.780	1.000	0.093	0.500	0.586	0.026	0.936	0.089	0.235	0.543	0.024
BDvsSW	0.752	1.000	0.093	0.500	0.529	0.024	0.932	0.088	0.219	0.503	0.022
BDvsUK	0.668	1.000	0.095	0.500	0.468	0.021	0.930	0.088	0.214	0.456	0.020
BDvsUS	0.428	0.965	0.090	0.348	0.345	0.015	0.888	0.086	0.098	0.326	0.015

(*Table* 4.6 *Continued*). The traditional dependence measure of linear correlation ρ and extreme values theory dependence measures $\bar{\chi}$ and χ for both the Loss and Gain tails for pairs of assets in our universe. In red, the χ estimates for asymptotic dependent pairs when the hypothesis test has $\alpha = 1\%$, but resulting significantly asymptotic independent when $\alpha = 5\%$; in black those whose $\bar{\chi}$ is statistically indistinguishable from one (asymptotic dependent), for $\alpha = 5\%$.

		Loss Tail					**Gain Tail**				
Markets	ρ	$\bar{\chi}$	$\sigma_{\bar{\chi}}$	p-value ($H_0 : \bar{\chi} = 1$)	χ	σ_{χ}	$\bar{\chi}$	$\sigma_{\bar{\chi}}$	p-value ($H_0 : \bar{\chi} = 1$)	χ	σ_{χ}
ITvsNL	0.640	0.969	0.090	0.365	0.480	0.021	0.875	0.086	0.073	0.425	0.019
ITvsSW	0.591	0.862	0.085	0.053	0.435	0.019	0.896	0.087	0.115	0.386	0.017
ITvsUK	0.582	0.960	0.090	0.328	0.427	0.019	0.847	0.084	0.035	0.378	0.017
ITvsUS	0.316	0.749	0.080	0.001			0.558	0.071	0.000		
NLvsSW	0.763	1.000	0.092	0.500	0.551	0.025	1.000	0.092	0.500	0.492	0.022
NLvsUK	0.790	1.000	0.094	0.500	0.571	0.025	1.000	0.095	0.500	0.521	0.023
NLvsUS	0.430	0.909	0.087	0.150	0.353	0.016	0.900	0.087	0.125	0.321	0.014
SWvsUK	0.679	1.000	0.093	0.500	0.465	0.021	0.982	0.091	0.422	0.448	0.020
SWvsUS	0.341	0.938	0.089	0.243	0.291	0.013	0.703	0.078	0.000		
UKvsUS	0.417	0.874	0.086	0.071	0.347	0.015	0.783	0.082	0.004		

relying on the assumption of asymptotic dependence, is inappropriate in many cases. Considering the difference in terms of asymptotic dependence we found in the two tails, this assumption would definitely not be appropriate with reference to the gain tail. Taking into account an asymmetric model would be advisable. Multivariate that assumes asymptotic dependency will overestimate portfolios that consist of asset returns that are asymptotically independent; thus is important to perform a preliminary analysis of the portfolio components to assess the extremal dependence structure.

Remark 4.2 (**Market volatility**). The number of cases of asymptotic dependence should probably be considered an upper bound since Poon et al. (2003, 2004) show that some of the extremal dependence is due to the non-i.i.d. nature of the data. In their investigation, the number of cases of asymptotic dependence drops when the market volatility is filtered from the individual return series. This indicates that volatility is a major contributing factor to the between-series extremal dependence.

Remark 4.3 (**Market integration**). With reference to similar studies conducted in the past, the degree of asymptotic independence those found in the data was higher[3], with a very few instances of asymptotic dependence between markets.

However, a subset of the markets analysed here and different time periods were actually examined, and mostly with data arriving at early 2000's. Therefore, our finding regarding an increase in dependence through time, can be easily attributed both to markets integration and to the subsequent financial crises, further bringing a contagion and systemic risk.

Remark 4.4 (**Dependence in the extremes**). Both forms of extremal dependence permit dependence between moderately large values of each variable, but the very largest values from each variable can occur together only when the variables exhibit asymptotic dependence.

The extremal measures can be used to measure dependency of the same or different signs. In particular for those pairs of market exhibiting asymptotic independence, i.e. $\chi = 0$, the value of $\bar{\chi}$ represents the degree of association of the rvs in the extremes.

In our data base, none of the asymptotic independent markets, are independent also when considering moderately large values. $\bar{\chi}$ assumes always positive values and thus showing rvs positively associated in the extremes.

Lastly, note the geographic relationship between markets which are asymptotically independent with most the other markets. These are mainly included in the first of the three tables 4.6 and are: Hong Kong, Japan and Australia. It seems that the cases of strong asymptotic independence occur mostly in markets which are not in close geographic proximity.

[3]See for example Poon et al. (2003),Poon et al. (2004), Bradley and Taqqu (2004b).

4.3 Portfolio optimisation results

In this section we present the results obtained by applying the EVT based asset allocation, presented in Chapter 2, to the Group of Five G-5 Countries. This name is given to the five industrialized nations that meet periodically to achieve a cooperative effort on international economic and monetary issues. The G-5 consists of five of the world's leading industrialised countries: France, Germany, Japan, the United Kingdom, and the United States.

Before proceeding we focus on the analysis of their extremal dependence structure. The main variables of interest are shown in Table 4.7.

These values reveal that Japan is the only country that is asymptotically independent of the other markets. Furthermore, as measured by $\bar{\chi}$ (loss tail), has also a low amount of dependence, in terms of positive correlation in extremes. This holds also when looking at the gain tail. Thus, in agreement with what has been said in the previous section, (positive) returns exhibits even lower sub-asymptotic dependence. All these facts support the holding of Japan for the benefit of diversification.

Table 4.7: Focus on the extremal dependence for the G-5 Countries. Classical dependence measure, linear correlation ρ, and EVT dependence measures $\bar{\chi}$ and χ for both the Loss and Gain tails are shown.

		Loss Tail					**Gain Tail**				
Markets	ρ	$\bar{\chi}$	$\sigma_{\bar{\chi}}$	p-value ($H_0 : \bar{\chi} = 1$)	χ	σ_{χ}	$\bar{\chi}$	$\sigma_{\bar{\chi}}$	p-value ($H_0 : \bar{\chi} = 1$)	χ	σ_{χ}
JPvsFR	0.249	0.456	0.067	0.000			0.352	0.062	0.000		
JPvsBD	0.255	0.412	0.065	0.000			0.370	0.063	0.000		
JPvsUK	0.263	0.437	0.066	0.000			0.465	0.067	0.000		
JPvsUS	0.059	0.360	0.062	0.000			0.308	0.060	0.000		
FRvsBD	0.767	1.000	0.095	0.500	0.574	0.026	0.974	0.090	0.388	0.574	0.026
FRvsUK	0.727	1.000	0.093	0.500	0.528	0.024	0.953	0.089	0.299	0.528	0.024
FRvsUS	0.406	0.916	0.088	0.169	0.335	0.015	0.833	0.084	0.023		
BDvsUK	0.668	1.000	0.095	0.500	0.468	0.021	0.930	0.088	0.214	0.468	0.021
BDvsUS	0.428	0.965	0.090	0.348	0.345	0.015	0.888	0.086	0.098	0.345	0.015
UKvsUS	0.417	0.874	0.086	0.071	0.347	0.015	0.783	0.082	0.004		

The other markets are asymptotically dependent and, in this case, looking at the value of χ, we can measure the strength of it. We notice that US has the lowest level of asymptotic dependence within the group, whereas France exhibits the strongest limiting dependence. The others, United Kingdom (UK) and Germany (BD), have got roughly the same dependence level.

Tables 4.8 and 4.9 show the results obtained from solving the optimisation problem, via the methodology described in detail in Chapter 2, when focussing on the G-5 Countries' market indices. The optimal solutions, represented by the allocation vector $\boldsymbol{\omega}^*_\alpha$, are displayed for several high standard confidence levels ($\alpha_1 = 97.5\%$, $\alpha_2 = 99\%$, $\alpha_3 = 99.5\%$) and very high ones ($\alpha_4 = 99.9\%$ and $\alpha_5 = 99.99\%$).

The output from our model is then compared to the optimal asset allocation we would find under the assumption of normally distributed return. The latter is represented by the vector $\omega^{\mathcal{N}}$. We can notice that it does not vary when changing the confidence level as, finding the optimal portfolio that minimise either VaR or ES, coincides with identifying the minimum variance portfolio.

For every confidence level, the risk measure's of each marginal distribution is showed in the column next to the the related optimal allocation vector. This gives us an idea about the riskiness of the equity market index as a univariate random variable. In particular, looking at this quantities, at the highest quantiles $\alpha = 99.99\%$, Japan is less risky than all other markets considering both risk measures.

As a result, Japan is held in the largest proportion in the minimum VaR_α and ES_α portfolios for $\alpha = 99.99\%$. The initial optimal allocation, for the lowest confidence level $\alpha = 97.5\%$ in the set, invests his largest share in the US market, followed by Japan index, the UK, Germany and finally France. The same investment order holds for both the risk measures, even if with different, although similar, proportions.
Moving towards higher α, the portfolio proportion invested in US index decreases, while the Japanese component gains ground, reversing the situation at $\alpha = 99.9\%$ when considering the minimum VaR portfolio, and becoming the most influential investment at $\alpha = 99\%$ when minimising the ES.

Thus, for the α most far in the tail, the minimum risk portfolio has its bigger share invested in the only market that, according to χ and $\bar{\chi}$ measures, is considered asymptotically independent with all the others. The second asset in terms of proportion in the portfolio composition is the US index that, despite being the riskiest when looking at the risk measure's values on the marginals, has the lowest degree of dependence with all the others, among the asymptotic dependent components.

We can notice that using ES as risk measure, the asset allocation quicker reacts to and absorbs the asymptotic dependence results of the countries whose indices are involved in the optimisation. This is due to the nature of the risk measure itself, that is by definition a mean of the extreme values lying in the tail of the distribution.

In general, however, it is difficult to interpret and truly understand the contribution of the changing relative marginal risks and the change in dependence. This is particularly

Table 4.8: Optimal VaR_α allocations of the two stage incremental trade algorithm for the G5 markets as a function of high confidence level α. The columns with heading VaR_α are the estimated value at risk of the market at the given confidence level. The last row shows the extra risk taken by using normal weights $w^{\mathcal{N}}$ instead of optimal weights w^*. The risk measure's values are expressed in % terms.

		$\alpha_1 = 97,5\%$		$\alpha_2 = 99\%$		$\alpha_3 = 99,5\%$		$\alpha_4 = 99,9\%$		$\alpha_5 = 99,99\%$	
	$w^{\mathcal{N}}$	$w^*_{\alpha_1}$	VaR_{α_1}	w^*_2	VaR_{α_2}	$w^*_{\alpha_3}$	VaR_{α_3}	$w^*_{\alpha_4}$	VaR_{α_4}	$w^*_{\alpha_5}$	VaR_{α_5}
US	46.08%	50.90%	2.115	44.18%	2.957	43.18%	3.739	31.71%	6.192	30.82%	12.026
JP	30.45%	30.71%	2.652	33.94%	3.546	34.55%	4.293	52.29%	6.290	57.22%	9.925
UK	15.09%	15.90%	2.414	16.91%	3.319	17.22%	4.118	5.71%	6.441	1.35%	11.308
FR	1.27%	0.85%	2.704	0.77%	3.701	0.78%	4.534	0.27%	6.763	3.41%	10.822
BD	7.10%	1.64%	2.591	4.19%	3.504	4.27%	4.279	10.02%	6.406	7.21%	10.444
$VaR_\alpha(\mathcal{N})$		1.6293		2.2446		2.8157		4.6047		8.8522	
$VaR_\alpha(\mathcal{O})$		1.6286		2.2166		2.7747		4.4427		7.2353	
$\frac{VaR_\alpha(\mathcal{N})}{VaR_\alpha(\mathcal{O})}$		1.0004		1.0126		1.0148		1.0365		1.2235	

Table 4.9: Optimal ES_α allocations of the two stage incremental trade algorithm for the G5 markets as a function of high confidence level α. The columns with heading ES_α are the estimated value at risk of the market at the given confidence level. The last row shows the extra risk taken by using normal weights $w^{\mathcal{N}}$ instead of optimal weights w^*. The risk measure's values are expressed in % terms.

		$\alpha_1 = 97,5\%$		$\alpha_2 = 99\%$		$\alpha_3 = 99,5\%$		$\alpha_4 = 99,9\%$		$\alpha_5 = 99,99\%$	
	$w^{\mathcal{N}}$	$w^*_{\alpha_1}$	ES_{α_1}	w^*_2	ES_{α_2}	$w^*_{\alpha_3}$	ES_{α_3}	$w^*_{\alpha_4}$	ES_{α_4}	$w^*_{\alpha_5}$	ES_{α_5}
US	46.08%	43.39%	3.211	34.14%	4.343	33.79%	5.395	30.89%	8.692	28.72%	16.534
JP	30.45%	36.28%	3.703	40.60%	4.722	40.19%	5.572	57.62%	7.847	58.94%	11.987
UK	15.09%	15.98%	3.532	20.64%	4.653	20.44%	5.643	2.05%	8.521	2.10%	14.549
FR	1.27%	0.79%	3.876	0.43%	5.013	1.43%	5.962	1.86%	8.502	1.39%	13.127
BD	7.10%	3.56%	3.680	4.19%	4.746	4.15%	5.651	7.59%	8.134	8.85%	12.849
$ES_\alpha(\mathcal{N})$		2.4297		3.2559		4.0226		6.4246		12.1273	
$ES_\alpha(\mathcal{O})$		2.4131		3.2113		3.9172		5.6547		8.6179	
$\frac{ES_\alpha(\mathcal{N})}{ES_\alpha(\mathcal{O})}$		1.0068		1.0139		1.0269		1.1362		1.4072	

true when considering the multi-asset setting like in our case.

A surprising result revealed in Tables 4.8 and 4.9 is the robustness of the classical approach and thus robustness of diversification, for standard confidence levels. Indeed, the last row in the table shows the extra risk taken by using normal weights $w^{\mathcal{N}}$ instead of optimal weights w^* resulting from our model. The extra risk is calculated as ratio between the risk measures of the optimal portfolios, respectively identified under normality assumption[4] and under GPD modelling[5]. The assumption of normality apparently leads to only a modest amount of unnecessary risk at the $\alpha = 97.5\%$, $\alpha = 99\%$ and $\alpha = 99.5\%$ quantiles. But when going further into the tail, at the highest $\alpha = 99.9\%$ and $\alpha = 99.99\%$, there is a significant higher risk exposure (as high as 40% of extra risk if considering the expected shortfall) when using the normal weights.

4.4 Threshold sensitivity analysis

As we discussed in Section 2.7.2 Chapter 2, the threshold choice it is in general a delicate issue when applying the threshold exceedances EVT method. The higher u is set, the smaller is the bias incurred from applying an asymptotic theorem, but, on the other side, we are left with few observations to estimate the parameters, and this could compromise their statistical accuracy.

Moreover, the situation in our model setting is even more complicated as, in order to solve the portfolio optimisation problem and identify the minimum risk portfolio, an estimation of the tail distribution across many different portfolio allocations is necessary to determine the risk measures' estimate. Due to the high problem dimension, we referred to the "conventional choice method" and, based on the best practice found in the related literature, we set u equal to 95% quantile of the empirical distribution of $L(\omega), \omega \in \mathcal{R}^d$.

In the previous section 2.7.3 Chapter 2, we already performed a simulation analysis to understand how, in general, the EVT threshold exceedances method is sensitive to the threshold choice. The results obtained in that case, were reassuring and supported the use of the GPD approach rather than the Hill method.

However, although widely used in literature, the use of a single percentile of the empirical portfolio distribution as a target to select the sample considered in the tail estimation might seem a strong assumption. Therefore, in this section, the influence of different thresholds on our EVT-based model is investigated. Given the importance of the ξ parameter in determining the weight of the tail and the relationship between quantiles and expected shortfalls, we first show how estimates of ξ vary as we consider a series of thresholds, and then we move to the risk measures' point estimates.

[4] The optimal portfolio under the assumption of normality is denoted by $\mathcal{N}$.

[5] The optimal portfolio found via our model is denoted by $\mathcal{O}$.

Setting. We perform a sensitivity analysis studying the optimal portfolio choice, among a basket of international equity indices referred to G-5 countries, made using our methodology but letting the threshold vary among different values. In particular, we consider a set of u values, when performing the GPD fitting of the tail's portfolio loss distribution, that represent the empirical α-percentiles[6] forming an evenly spaced sequence in the interval [90%,99%], where the step is $\delta_\alpha = 0.001$[7]. We take into account both the problems of minimising the VaR and the ES and construct the optimal portfolio satisfying downside risk requirements of different tightness. Specifically, we solve the optimisation looking at the same risk measures' confidence levels considered in the previous section, namely $\{97.5\%, 99\%, 99.5\%, 99.9\%, 99.99\%\}$.

This setting give us a sample of 956 exceedances at maximum, that progressively decrease when moving the threshold far in the tail. In figure 4.2 we show the estimate of ξ^* as the thresholds u changes. The values of the empirical distribution $F_n(u)$, are displayed on the horizontal axis while the ξ^* on the y axis. ξ^* represents the shape parameter of the GPD fitting the right tail distribution, over u, of the optimal portfolio (negative) return. The results are exhibited when the optimal portfolio is obtained minimising both VaR, continuous line on the left hand side, and ES, dashed line on the right hand side, with several levels of confidence (different risk measure's confidence level on each row).

First we notice that the estimates of ξ^* are clearly different when considering various risk measures and confidence levels.

This is because we obtain distinct minimum risk portfolios, and so optimal weights $\boldsymbol{\omega}^*$, based on the risk measure's choice and on the tightness of the downside risk requirement considered. Therefore, the parameters of the GPD, ξ^* included, that fits the tail of the minimum risk portfolio are obviously determined by these choices.

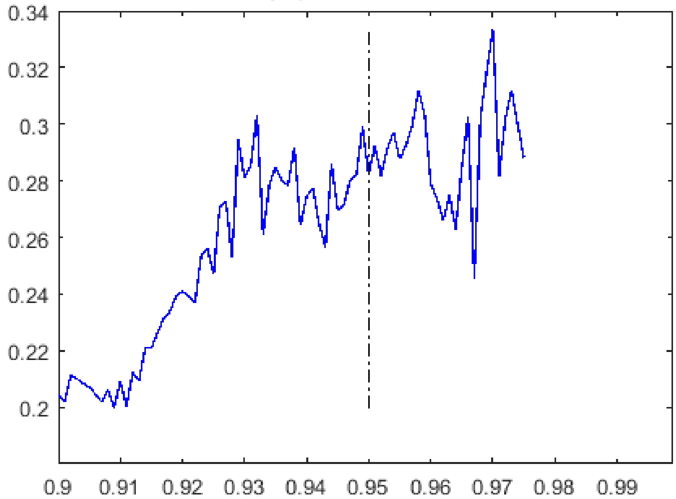
In all the cases, approximately around the threshold values corresponding to the $98^{th}-98.5^{th}$ empirical percentiles, the estimates exhibit a sudden change in direction and show a different evolution compared to the one was guiding the trend before. This is probably the result of the decreasing number of observations involved in the estimation, that leads to an increase in estimators' variability. ξ^* is always positive[8] and $\xi^* < 0.5$ indicates that all the portfolio loss distributions have finite variance.

We observe values $\xi^* \geqslant 0.5$ only when referring to the VaR as portfolio risk measure, and just for the very high threshold values where the estimators' accuracy might be compromised (in particular it happens in correspondence to the following confidence levels: $99\%, 99.5\%$ and 99.9%).

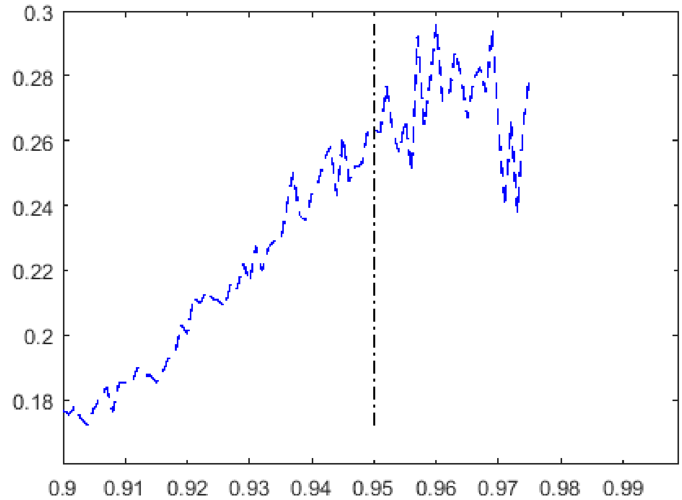
[6]Therefore, so that $F_n(u) = \alpha$, where $F_n(x) = \dfrac{1}{n}\sum_{s=1}^{n} \mathbb{1}_{[\tilde{L}_s(\boldsymbol{\omega}) \leq x]}$ is the empirical distribution of the portfolio loss.

[7]For a total of 100 possible values of the threshold.

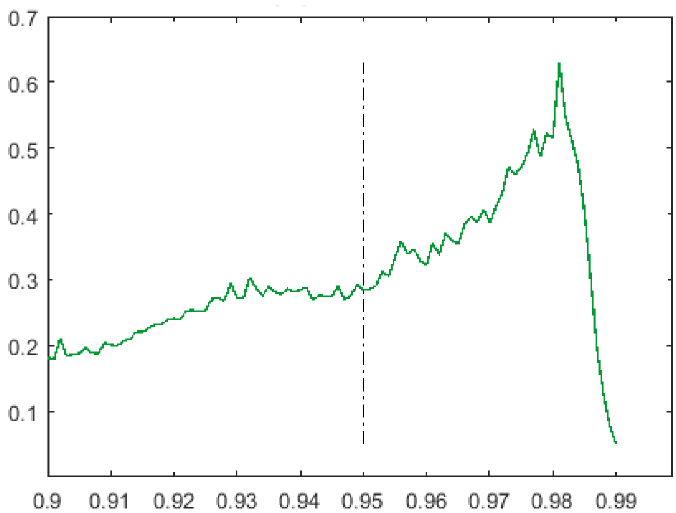
[8]Except for $\mathrm{VaR}_{99.99\%}(\mathcal{O})$ and $\mathrm{ES}_{99.99\%}(\mathcal{O})$ when the threshold assumes its highest values.



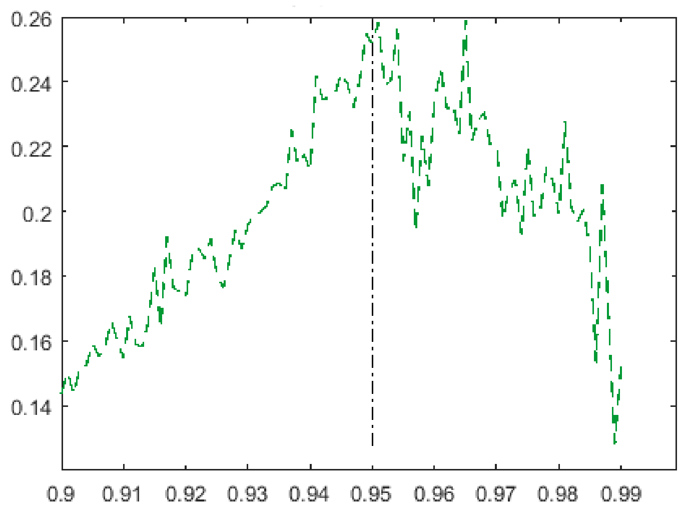
(a) ξ^*, $\mathrm{VaR}_{97.5\%}(\mathcal{O})$



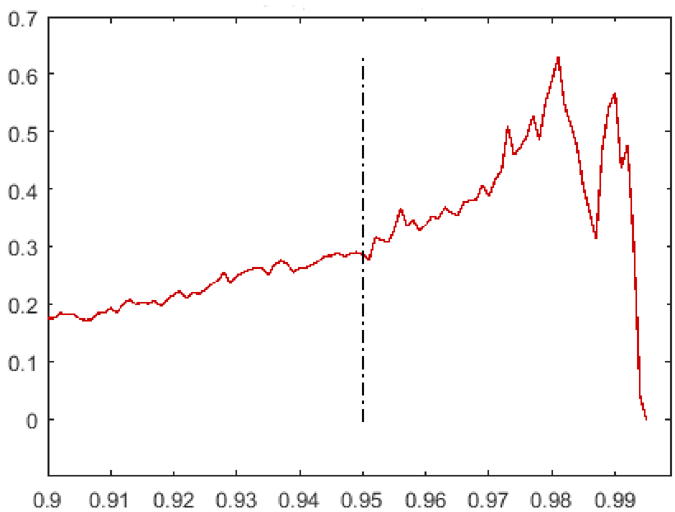
(b) ξ^*, $\mathrm{ES}_{97.5\%}(\mathcal{O})$



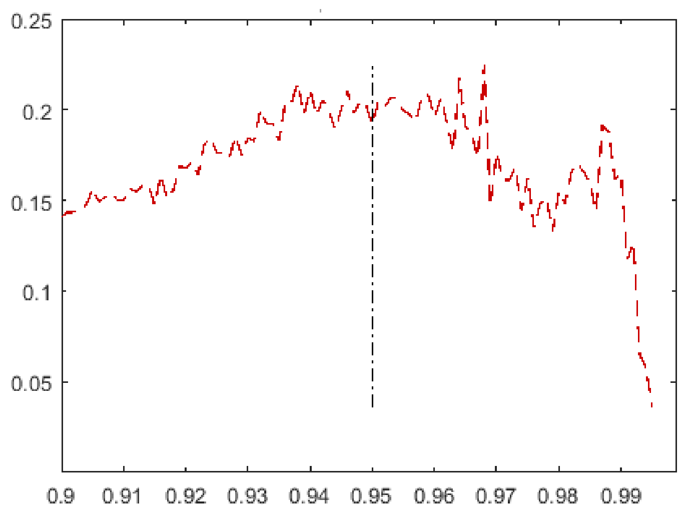
(c) ξ^*, $\mathrm{VaR}_{99\%}(\mathcal{O})$



(d) ξ^*, $\mathrm{ES}_{99\%}(\mathcal{O})$

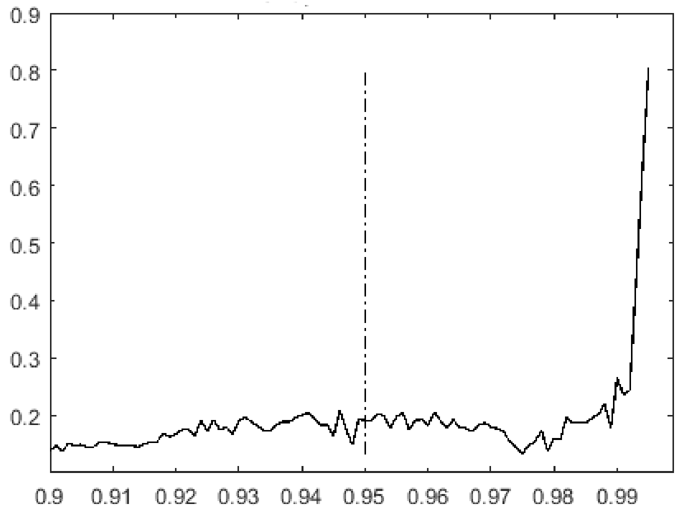


(e) ξ^*, $\mathrm{VaR}_{99.5\%}(\mathcal{O})$

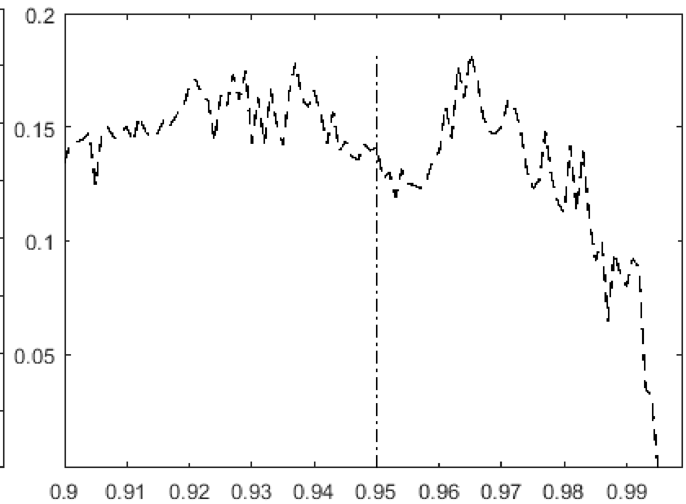


(f) ξ^*, $\mathrm{ES}_{99.5\%}(\mathcal{O})$

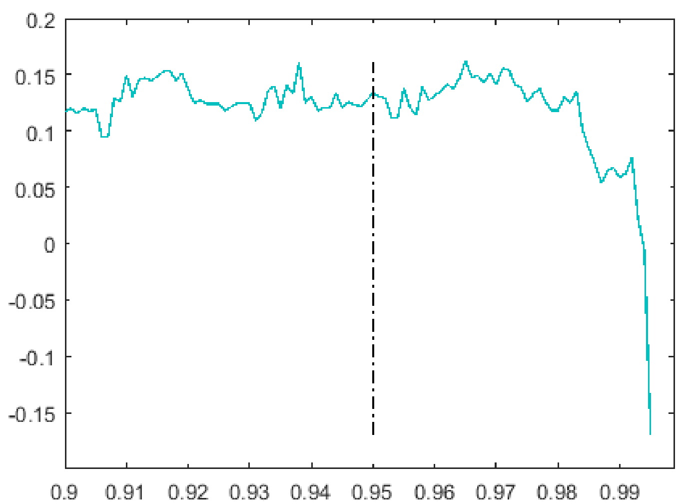
Figure 4.2



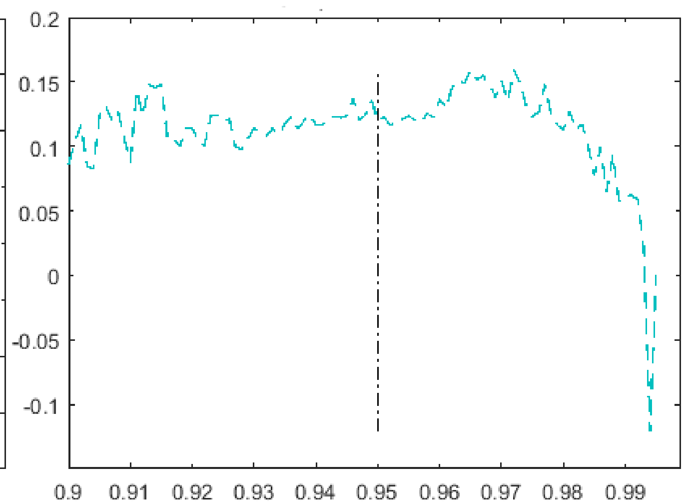
(g) ξ^*, VaR$_{99.9\%}(\mathcal{O})$



(h) ξ^*, ES$_{99.9\%}(\mathcal{O})$



(i) ξ^*, VaR$_{99.99\%}(\mathcal{O})$



(j) ξ^*, ES$_{99.99\%}(\mathcal{O})$

Figure 4.2: Estimate of ξ^* for different thresholds u (on the x axis we actually have $F(u)$). ξ^* is the shape parameter of the GPD fitting the distribution the Optimal Portfolio return over u. The results are shown when the optimal ptf is obtained minimising both VaR (continuous line, on the left hand side) and ES (dashed line, on the right hand side), and for several confidence levels (97.5%, 99%, 99.5%, 99.9%, 99.99%).

The estimates remain fairly constant around the values of approximately $0.12 - 0.14$ for both the risk measures with the two highest confidence levels (in 4.2g, 4.2h, 4.2i, and 4.2j). The method seems therefore quite robust when looking at the very extreme risks. In correspondence to the more standard quantiles, the trend of ξ^* for $\mathrm{VaR}_{97.5\%}(\mathcal{O})$ stabilise starting from $F_n(u) = 0.93$ around values of $0.28 - 0.3$ (4.2a). ξ^* related to the minimum risk portfolio $\mathrm{ES}_{97.5\%}(\mathcal{O})$, reaches its maximum values of $0.26 - 0.28$ when $F_n(u) \geq 0.95$ (4.2b). $\mathrm{VaR}_{99\%}(\mathcal{O})$ (4.2c) and $\mathrm{VaR}_{99.5\%}(\mathcal{O})$ (4.2e), although showing an overall increasing trend of ξ^*, are however quite stable around values of $0.25 - 0.3$ for $0.93 \leq F_n(u) \leq 0.96$. $\mathrm{ES}_{99.5\%}(\mathcal{O})$ remains confined for the whole u values in the interval $[0.15, 0.2]$ and, in particular, is quite steady on the value 0.2 when $0.93 \leq F_n(u) \leq 0.97$.

We also analyse the sensitivity of the two risk measures' point estimates, because this is the actual piece of information we are interested in. Figure 4.3 illustrates the point estimates of the optimal portfolio risk measures, VaR (solid line) and expected shortfall (dotted line), calculated using our EVT-based model. The results are always shown for several confidence levels, and when varying the threshold u.

It is clear that these estimates show definitely modest variability to the threshold choice, slightly bigger for the highest confidence interval (99.99%).

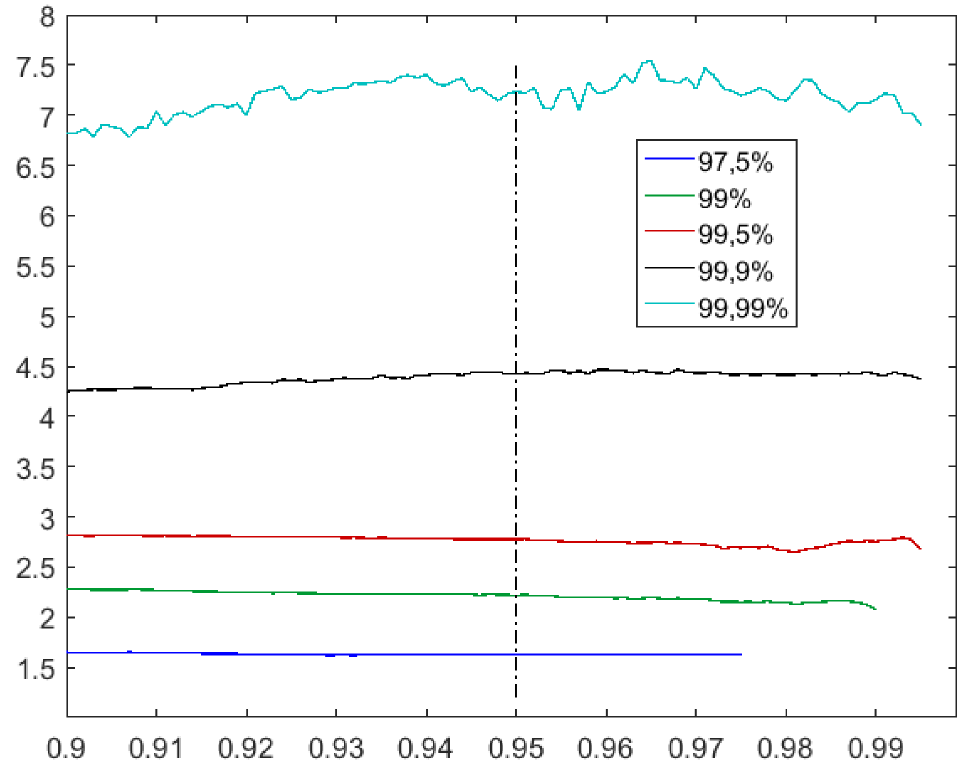
However, the reading of the graph might be impeded because of the different risk measures' scale. Therefore, in order to further investigate the first impression, we also display the same information similarly to what we have just done with the "optimal" shape parameters of the GPD. The results, separately illustrated by type of risk measure and by confidence level, are shown in figure 4.4[9].

The risk measures' estimate of the minimum risk portfolio, when considering a confidence level of 97,5%, are enclosed in an interval of width 0.03% for the VaR and 0.01% for the ES. Moreover, $\mathrm{VaR}_{97.5\%}(\mathcal{O})$ remain remarkably constant around 1.625% for $F_n(u) \geq 0.93$. If we consider $\mathrm{VaR}_{99\%}(\mathcal{O})$, $\mathrm{ES}_{99\%}(\mathcal{O})$, $\mathrm{VaR}_{99.5\%}(\mathcal{O})$, and $\mathrm{ES}_{99.5\%}(\mathcal{O})$, the intervals within the point estimates vary, are slightly bigger with a dimension of $0.1\% - 0.2\%$. Finally, when looking at the highest confidence levels, the trend showed by the estimates of the VaR and ES are quite similar. These exhibit modest variability[10] that essentially tracks the variability of the ξ^* estimates.

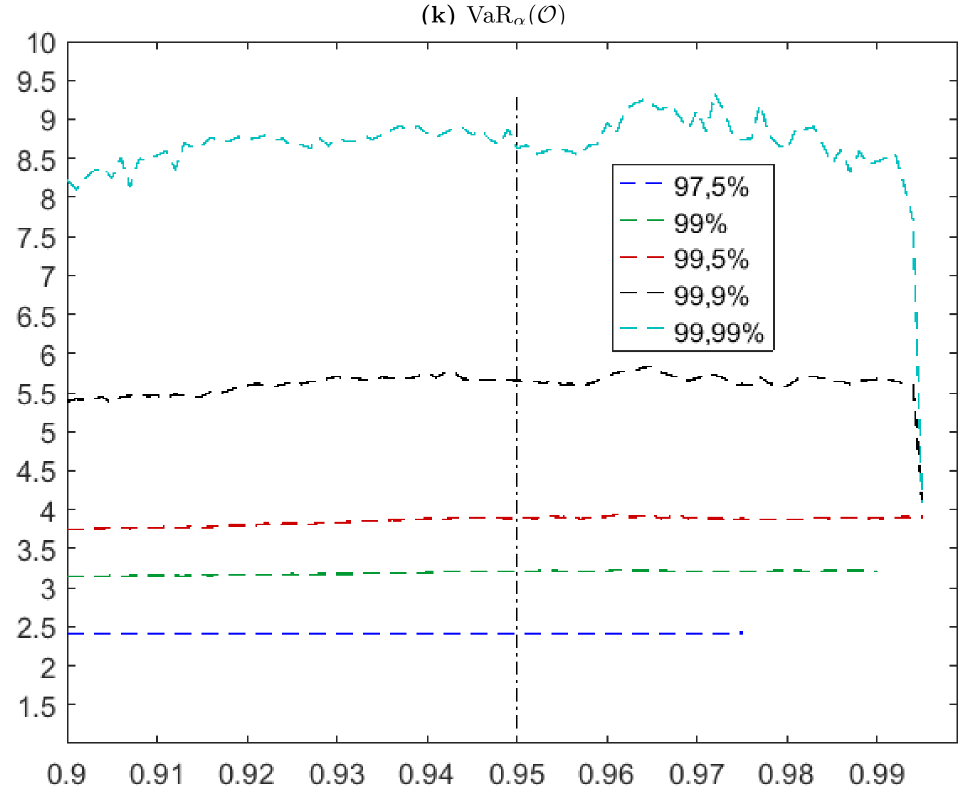
These pictures provide some reassurance that different thresholds do not lead to drastically different conclusions. Furthermore, the stability usually characterising the estimates correspondent to $\{u : 0.94 \leq F_n(u) \leq 0.97\}$, supports the threshold choice we made.

[9] In the graphs 4.4h and 4.4j we do not represent the two last point estimates of $\mathrm{ES}_{99.9\%}(\mathcal{O})$ and $\mathrm{ES}_{99.99\%}(\mathcal{O})$, correspondent to $F_n(u) = 99.5\%$ and equal to 4.07%, as this would alter the scale and we will not be able to appreciate the actual evolution. These two values are causing the sudden drop of the two series at the top in figure 4.3.

[10] The intervals containing these estimate are not wider than 1%.

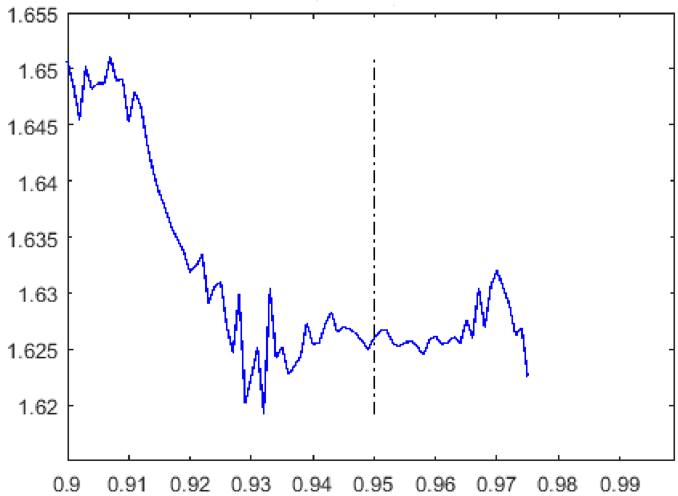


(k) $\mathrm{VaR}_\alpha(\mathcal{O})$

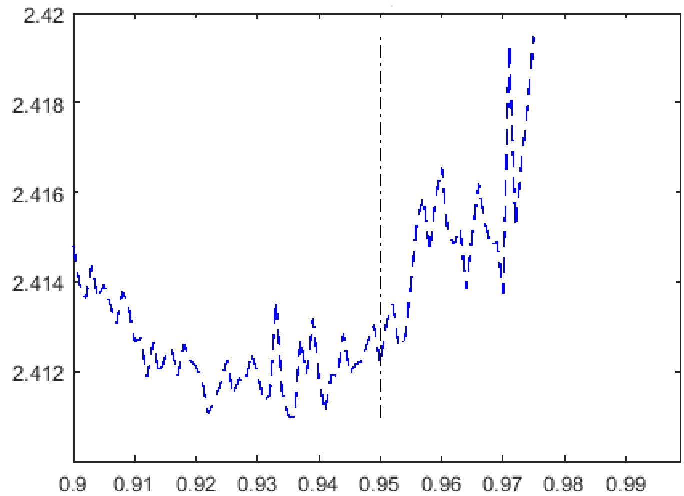


(l) $\mathrm{ES}_\alpha(\mathcal{O})$

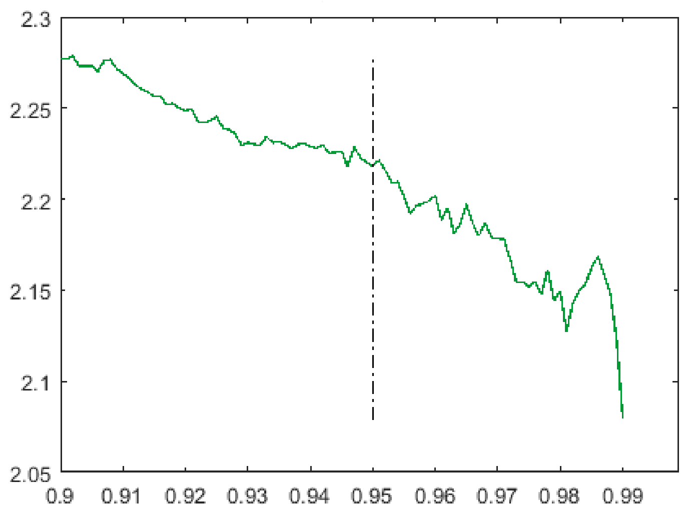
Figure 4.3: Point estimates of the Optimal portfolio risk measures, VaR (solid line) and expected shortfall (dotted line), using our EVT-based model, for several confidence levels, when varying the threshold u (on the x axis we actually have $F(u)$).



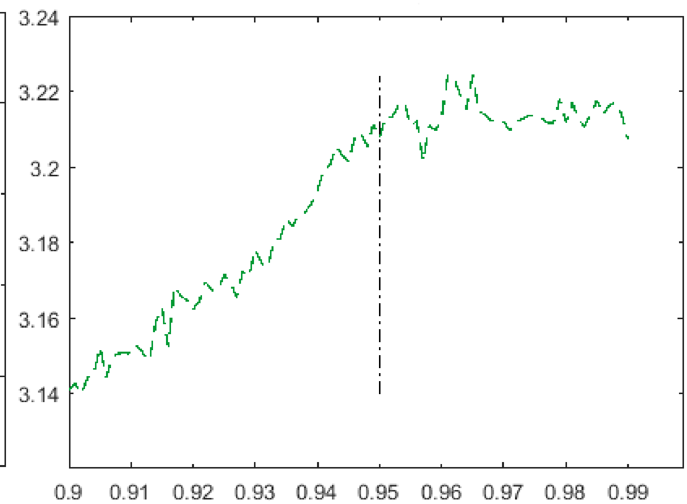
(a) $\mathrm{VaR}_{97.5\%}(\mathcal{O})$



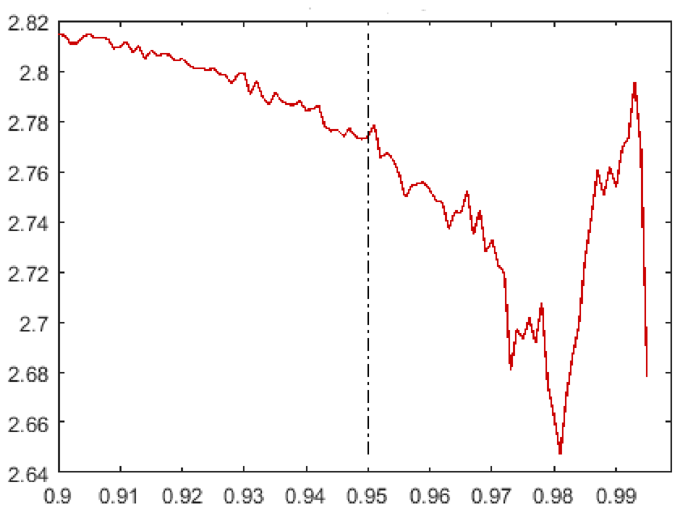
(b) $\mathrm{ES}_{97.5\%}(\mathcal{O})$



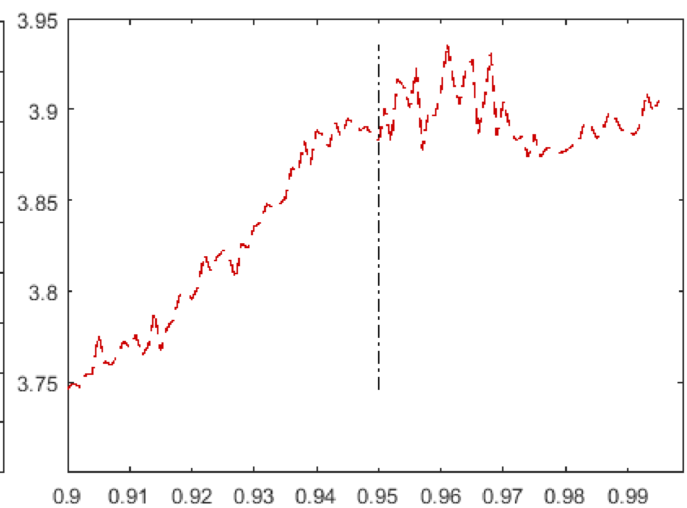
(c) $\mathrm{VaR}_{99\%}(\mathcal{O})$



(d) $\mathrm{ES}_{99\%}(\mathcal{O})$

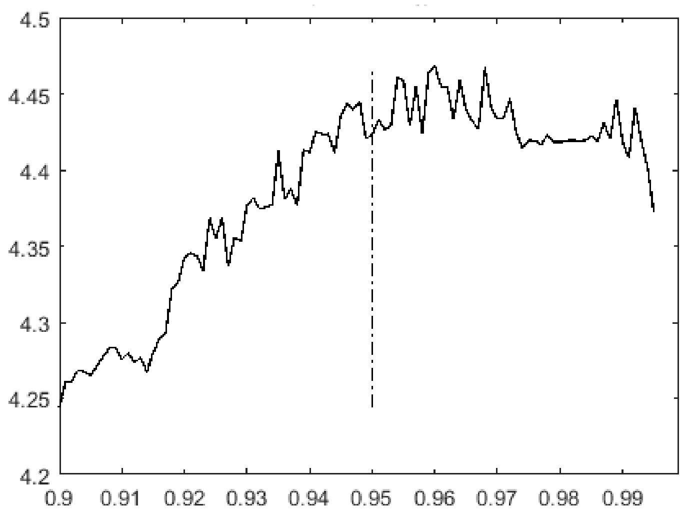


(e) $\mathrm{VaR}_{99.5\%}(\mathcal{O})$

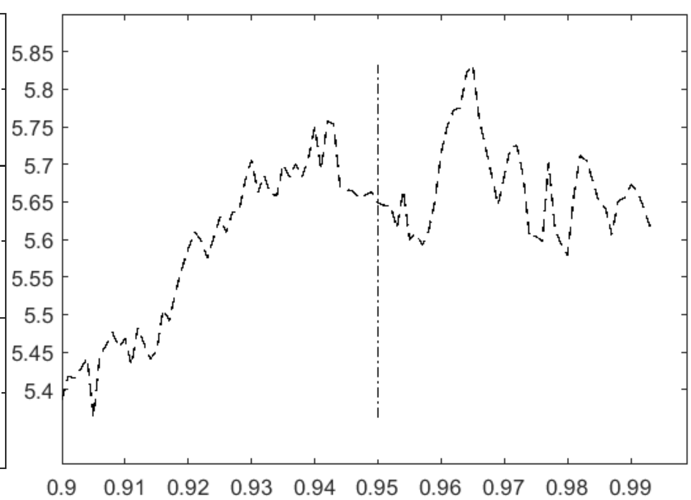


(f) $\mathrm{ES}_{99.5\%}(\mathcal{O})$

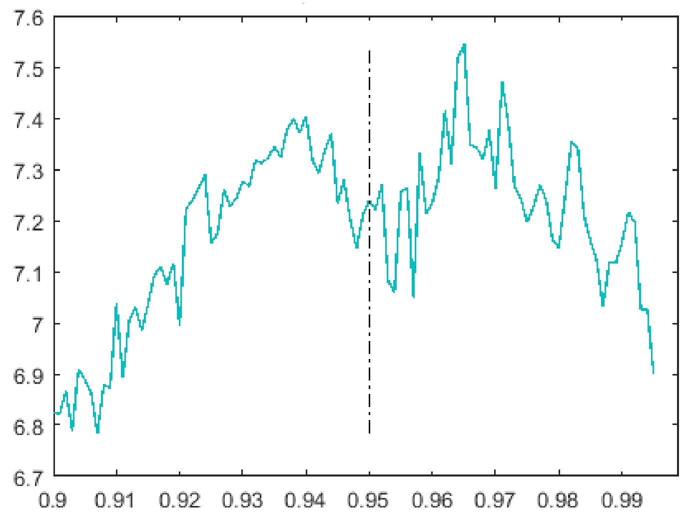
Figure 4.4



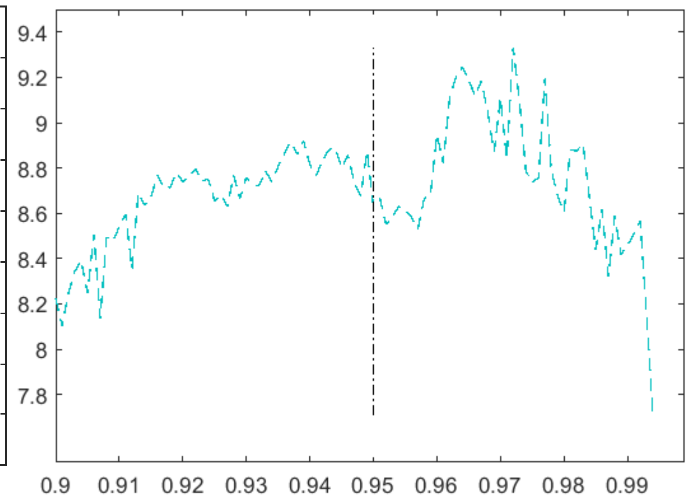
(g) $\mathrm{VaR}_{99.9\%}(\mathcal{O})$



(h) $\mathrm{ES}_{99.9\%}(\mathcal{O})$



(i) $\mathrm{VaR}_{99.99\%}(\mathcal{O})$



(j) $\mathrm{ES}_{99.99\%}(\mathcal{O})$

Figure 4.4: Point estimates of the Optimal portfolio risk measure, VaR (solid line on the left hand side) and expected shortfall (dotted line on the right hand side), using our EVT-based model, and for different thresholds u values (on the x axis we actually have $F(u)$). We separately illustrate the results for several confidence levels (97.5%, 99%, 99.5%, 99.9%, 99.99%). In the graphs 4.4h and 4.4j we do not represent the two last point estimates, correspondent to $F_n(u) = 99.5\%$ and equal to 4.07, as this would make them less legible.

4.5 Final remarks and further developments

The material contained in this dissertation could be further developed and extended in several directions. This section discusses in more details the reasons behind some methodology choices we made, points out limitations of the current work, and also outlines directions for future research.

4.5.1 Dynamic threshold

In section 4.4, we studied the impact of choosing different thresholds in our EVT-based model. The sensitivity was performed considering a set of values for the fixed (quantile based) threshold, chosen in order to model the tail of every possible portfolio loss distribution via the POT approach.

It would be very interesting to investigate how the model could incorporate a dynamic threshold, allowing its adjustment in response to the portfolio configurations selected at each step of the optimisation algorithm. A possible approach to improving and refining the model could be via the integration of factor models and tail index regression within the algorithm.

4.5.2 Conditional EVT approach.

Various stylized facts of empirical finance argue against the assumption that market return series are independent and identically distributed. While the correlation of market returns is low, the serial correlation of absolute or squared returns is high; returns show *volatility clustering*, namely the tendency of large values to be followed by other large values, although not necessarily of the same sign.

Therefore, in calculating daily VaR estimates for such risks, a large literature recognise that volatility of market instruments should be take into account. Intuitively, an extreme value in a period of high volatility appears less extreme than the same value in a period of low volatility. The most popular models for characterising this phenomenon are the stochastic volatility (SV) models.

An interesting approach to explore in future research would be enclosing such a dynamic risk measurement procedure in our model, to take into account the extreme risk over and above the volatility risk. A possible way to proceed is adopting the methodology suggested by Diebold et al. (2000) and McNeil and Frey (2000).

Their basic idealisation is that returns (X_t) follow a stationary time series model with stochastic volatility structure $X_t = \mu_t + \sigma_t Z_t$, with Z_t a strict white noise process of unknown distribution. Afterwards, they apply a two-step procedure to obtain conditional VaR and ES estimates, that consist in: first, pre-filtering the data through an $AR(1)-$

$GARCH(1,1)$ model, and obtaining estimates of the conditional volatility via pseudo-maximum-likelihood; second, applying EVT to the residual distribution[11]. In particular one could use historical simulation for the central part of the distribution and threshold methods from EVT for the tails, to estimate the distribution of the residuals. This approach would reflect two stylized facts exhibited by most financial return series, namely stochastic volatility and the fat-tailedness of conditional return distributions over short time horizons. We may test the approach on our equity index return series, and analyse via backtesting if this conditional methodology yields better estimates of VaR and expected shortfall than unconditional EVT[12].

A further interesting methodology to consider might be the nonparametric extension of the classical Peaks-Over-Threshold method from EVT, recently proposed by Chavez-Demoulin et al. (2014). This technique for the statistical estimation of conditional risk measures, is applicable to both stationary as well as nonstationary time series models[13].

Remark. The presence of stochastic volatility implies that returns are not necessarily independent over time. Hence, with such models there are two types of return distribution to be considered: the *conditional return distribution* where the conditioning is on the current volatility and the *unconditional (marginal or stationary) distribution* of the process. We have already discussed the two in Remark 2.5 Chapter 2. The GPD method, we referred to, when applied to threshold exceedances in a financial return series is essentially an unconditional method for estimating the tail of the P&L distribution and associated risk measures.

Both distributions are actually of relevance to risk managers. A risk-manager who wants to measure the market risk of a given portfolio is mainly concerned with a possible loss caused by an adverse market movement over the next day, given the current volatility background. The estimation of unconditional tails, instead, provides different, but complementary information about risk. In this case, we take the long-term view and attempt to assign a magnitude to a specified rare adverse event, such as a 5-year loss (the size of a daily loss which occurs on average once every 5 years. This kind of information may be of interest to the risk manager who wishes to perform a scenario analysis and get a feeling for the scale of worst case or stress losses. In general, this approach might more suitable in the credit and insurance field, being characterised by a long-term perspective.

The concern was raised that the use of conditional return distributions for market risk measurement might lead to capital requirements that fluctuate wildly over time and are

[11] Statistical tests and exploratory data analysis confirm that the error terms or residuals do form, at least approximately, i.i.d. series that exhibit heavy tails.

[12] In literature, McNeil and Frey's backtesting shows that it yields better estimates than unconditional EVT or GARCH-modelling with normally distributed error terms. Their analysis contradicts Danielsson et al. (1998), who state that an unconditional approach is instead better suited for VaR estimation than conditional volatility forecast.

[13] The stationarity assumption may be violated by erratic changes of regime.

therefore difficult to implement. However, besides providing a basis for the determination of risk capital, measures of market risk are also employed to give the management of a financial firm a better understanding of the riskiness of its portfolio. Therefore, from this point of view, as the riskiness of a portfolio does indeed vary with the general level of market volatility, the current volatility background should be reflected in the risk-numbers reported to management.

4.5.3 Multivariate extreme value theory and copulas

A key point of our work is that, in order to solve the optimisation problem, only univariate EVT is applied to model the tail of the portfolio loss using the POT paradigm. Therefore, the question of modelling the full joint distribution of asset returns, in particular their tails, using techniques motivated by MEVT and copula-based, is replaced by modelling the returns of the univariate portfolio itself.

The choice of reducing a multivariate problem to a univariate one, is justified by the computational hurdles posed by the optimisation problem.

Suppose we have a family of random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ representing d-dimensional losses at different points in time, where $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{id})^{\mathsf{T}}$.
We define the vector of componentwise block maxima to be $(M_{1n}, M_{2n}, \ldots, M_{dn})^{\mathsf{T}}$ where $M_{jn} = \max(X_{1j}, X_{2j}, \ldots, X_{nj})$ is the block maximum of the j^{th} component for a block of size n observations. Now consider the vector of normalized block maxima given by $((M_{1n}-b_{1n})/a_{1n}, \ldots, (M_{dn}-b_{dn})/a_{dn})^{\mathsf{T}}$, where $a_{jn} > 0$ and b_{jn} are normalizing sequences. If this vector converges in distribution to a non-degenerate limiting distribution then this limit must have the form

$$C\left(H_{\xi_1}\left(\frac{x_1-\mu_1}{\sigma_1}\right), H_{\xi_2}\left(\frac{x_2-\mu_2}{\sigma_2}\right), \ldots, H_{\xi_d}\left(\frac{x_d-\mu_d}{\sigma_d}\right)\right)$$

for some values of the parameters ξ_j, μ_j, σ_j and some copula C. It must have this form because of univariate EVT and each marginal distribution of the limiting multivariate distribution must be a GEV.
MEVT characterizes the copulas C which may arise in this limit, the so-called *MEV copulas*. It turns out that the limiting copulas must satisfy

$$C(u_1^t, \ldots, u_d^t) = C^t(u_1, \ldots, u_d) \text{ for } t > 0.$$

There is no single parametric family which contains all the MEV copulas, but certain parametric copulas are consistent with the above condition and might therefore be regarded as *natural* models for the dependence structure of extreme observations. In two dimensions the Gumbel copula is an example of an MEV copula[14].

[14] It is moreover a versatile copula and allow us to characterise the configurations of asymptotic depen-

Realistically, however, parametric models of this kind are only viable in a small number of dimensions. If we are interested in only a few risk factors and are particularly concerned that joint extreme values may occur, we can use such models to get useful descriptions of the joint tail. In very high dimensions there are too many parameters to estimate and too many different tails of the multivariate distribution to worry about. This is the so-called *curse of dimensionality*. In such situations collapsing the problem to a univariate problem by considering a whole portfolio of assets as a single risk and collecting data on a portfolio level seems more realistic.

This seems reasonable also in light of the results of Bradley and Taqqu (2004a). They perform a comparison of the performance of univariate and multivariate techniques for an asset allocation problem. Due to numerical constraints, they examine the bivariate case only and conclude that MEVT yields accurate results if the copula allows for asymptotic independence. MEVT models with a dependence function that prescribes only asymptotic dependence[15] may overstate the risk of a properly diversified portfolio. They also find that the multivariate and the univariate estimated risks that are in close agreement for a wide range of failure levels and allocations.

In particular, the use of a univariate approach, which is easier to implement, seems to be robust when we are not interested in particularly extreme quantiles.

4.5.4 Dynamic asset allocation

Finally, another interesting question is to look deeper at the implications the findings could have in terms of portfolio rebalancing. It would be relevant for asset managers' purposes, to develop the model towards a Dynamic Asset Allocation Framework.

A portfolio's asset allocation is the major determinant of a portfolio's risk-and-return characteristics. However, over time, asset classes produce different returns, so the portfolio's asset allocation changes. Therefore, to recapture the portfolio's original risk-and-return characteristics, it should be rebalanced.

The primary goal of a rebalancing strategy is to minimize risk relative to a target asset allocation, rather than to maximize returns. Considering the foundations characterising our model, it would probably fit well into this perspective.

dence we have analysed in Chapter 3. If the parameter β is 1 then $C_1^{Gu}(v_1, v_2) = v_1 v_2$ and this copula models *independence* of the components of a random vector $(X_1, X_2)^\intercal$. If $\beta \in (0,1)$ then the Gumbel copula models dependence between X_1 and X_2. As β decreases the dependence becomes stronger until a value $\beta = 0$ corresponds to perfect dependence of X_1 and X_2; this means $X_2 = T(X_1)$ for some strictly increasing function T. For $\beta < 1$ the Gumbel copula shows *tail dependence*, namely the tendency of extreme values to occur together

[15] For instance, the standard logistic model to multivariate extremes.

4.6 Conclusions

Recent studies have highlighted the importance of rare events in several fields such as insurance and portfolio choice. The extreme and rare events might be in the form of a large change in investment returns, a stock market crash, major defaults, or the collapse of risky asset prices other than men-made or natural catastrophes.

The problem of portfolio optimisation has always been a topic of strong interest for investors and fund managers, and the need to create resilient portfolios, reducing the impact of extreme adverse events, is a growing concern given the financial crises suffered in recent decades. The new risk-based solvency requirements for the banking and insurance sectors further motivates the search for more appropriate methodologies to cope with rare events which cause heavy losses.

Extreme Value Theory provides the solid mathematical and probabilistic fundamentals, essential in order to build statistical models describing such risky events. Wherever the tail of a loss distribution is of interest, whether for market, credit, operational or insurance risks, EVT provides simple but powerful tool for estimating quantile-based or mean-based risk measures according to the EVT framework, that should be more suitable when focussing on a random variable's tail. Employing other classical techniques might be more practical and intuitive, but could lead to an underestimation of the tail risk, which plays such a crucial role in the everyday operations of a financial institution.

We have investigated the asset allocation problem for risk-averse investors, whose goal is to minimise the quantile and mean based risk measures VaR_α and ES_α of the portfolio, to guard against the occurrence of large losses. The model relies on the foundation introduced by the safety first criterion.

Moreover, we went beyond the classical MVO and its assumptions using EVT to reproduce the fact that the returns are heavy-tailed (instead of assuming normality) and employing quantile-based risk measures (instead of the symmetric variance). The methodology allows to embody the risk preferences of an investor representing the asymmetric treatment of upside and downside risk. This behavioural profile has been observed in several studies. Furthermore, the application of tail risk measures to solve the asset allocation problem, reflects the impelling need of properly managing and considering also the extreme events. This characterises risk managers' mindset within the new solvency regime, thinking beyond what is expected on average.

Furthermore, considering that the concept of diversification is fundamental to the decision of how best to allocate assets and it is partly controlled by the dependence between the assets under consideration, we performed a parallel analysis to identify the correct extremal dependence structure of the portfolio's constituents.

Using dependence measures $\bar{\chi}$ and χ from Multivariate EVT we found that extreme dependence is much stronger in bear markets than in bull markets. Comparing our result with similar studies conducted in the past, in general, the correlation between volatilities has increased over time to produce asymptotically dependent stock markets. The finding of an increase in dependence through time, can be attributed both to markets integration and to the financial crises, further bringing a contagion and systemic risk. Moreover, it seems that the cases of strong asymptotic independence occur mostly in markets which are not in close geographic proximity.

The EVT based methodology was tested on real data and applied to a portfolio of international equity market indices. In particular the empirical results of the portfolio optimisation examined the Group of Five G-5 countries: France, Germany, Japan, the United Kingdom, and the United States.

We found that when markets are asymptotically independent, there is a greater benefit to diversification at the highest confidence levels α than there is at lower α. Our investigation revealed that for the confidence levels most far in the tail $\alpha = 99.9\%$ and $\alpha = 99.99\%$, the minimum risk portfolio has its bigger share invested in the only market that is asymptotically independent with all the others, i.e. Japan. The second asset in terms of proportion in the portfolio composition is the US index that, despite being the riskiest when looking at the risk measure's values on the marginal, has the lowest degree of dependence among the asymptotic dependent assets. Furthermore, the optimal allocation is quantile-based and changes when varying the confidence level of the risk measures. This is not true, when assuming normally distributed returns.

We can also notice that using ES as risk measure, the asset allocation quicker reacts to and absorbs the asymptotic dependence results of the countries whose indices are involved in the optimisation. This is due to the nature of the risk measure itself, that is by definition a mean of the extreme values lying in the tail of the distribution.

A surprising result of our investigation is the robustness of the assumption of normality on the allocation problem. For each allocation problem, we compare the risk of the portfolio using normal weights $\omega^{\mathcal{N}}$ to the risk of the optimal portfolio ω^*.
For various high levels of confidence α we disclose the ratio of the risks. This ratio is surprisingly close to one for all but the largest α. This means that for all, except for the highest confidence levels, the investor takes not significant extra risk by allocating his assets as though returns were normally distributed. However, as soon as we move to the highest confidence levels, the extra risk taken, when considering the normality assumption, jumps to value as high as 40%.

Therefore, increasing the effort during the modelling phase in terms of model complexity and accuracy, can lead to a significant ex-post reduction of the risk exposure.

Bibliography

Acerbi, C. and Székely, B. (2014). Back-testing expected shortfall. *Risk*. forthcoming, available at `https://arxiv.org/pdf/1410.4382.pdf`. (Cited on page 27).

Acerbi, C. and Székely, B. (2016). L'es est mort, vive l'es! *Talk at ETH Zurich*. (Cited on page 27).

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503. (Cited on pages 19, 20, and 21).

Ang, A., Chen, J., and Xing, Y. (2006). Downside risk. *Review of Financial Studies*, 19(4):1191–1239. (Cited on page 16).

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1997). Thinking coherently. *Risk*, 10:68–71. (Cited on page 19).

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3):203–228. (Cited on pages 19 and 21).

Arzac, E. R. and Bawa, V. S. (1977). Portfolio choice and equilibrium in capital markets with safety-first investors. *Journal of Financial Economics*, 4(3):277–288. (Cited on pages 16, 34, and 35).

Bali, T. G. (2003). An extreme value approach to estimating volatility and value at risk. *The Journal of Business*, 76(1):83–108. (Cited on page 15).

Balkema, A. and Embrechts, P. (2007). *High risk scenarios and extremes: a geometric approach*. Zurich Lectures in Advanced Mathematics. European Mathematical Society Publishing House.

Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, pages 792–804. (Cited on pages 45 and 48).

Bäuerle, N. and Müller, A. (2006). Stochastic orders and risk measures: consistency and bounds. *Insurance: Mathematics and Economics*, 38(1):132–148. (Cited on page 22).

BCBS (December 2014). Consultative document. fundamental review of the trading book: Outstanding issues. http://www.bis.org/bcbs/publ/d305.pdf. Basel Committee on Banking Supervision. Basel: Bank for International Settlements. (Cited on page 28).

BCBS (January 2016). Standards. minimum capital requirements for market risk. http://www.bis.org/bcbs/publ/d352.pdf. Basel Committee on Banking Supervision. Basel: Bank for International Settlements. (Cited on pages 26 and 28).

BCBS (May 2012). Consultative document: Fundamental review of the trading book. http://www.bis.org/publ/bcbs219.pdf. Basel Committee on Banking Supervision. Basel: Bank for International Settlements. (Cited on pages 19 and 26).

BCBS (October 2013). Second consultative document. fundamental review of the trading book. http://www.bis.org/publ/bcbs265.pdf. Basel Committee on Banking Supervision. Basel: Bank for International Settlements. (Cited on pages 19 and 26).

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of extremes: Theory and Applications*. John Wiley & Sons. (Cited on page 37).

Beirlant, J. and Novak, S. (2006). The magnitude of a market crash can be predicted. *Journal of banking & finance*, 30(2):453–462. (Cited on page 66).

Beirlant, J., Teugels, J. L., and Vynckier, P. (1996). *Practical analysis of extreme values*. Leuven University Press. (Cited on pages 37 and 45).

Bekiros, S. D. and Georgoutsos, D. A. (2008). The extreme-value dependence of asia-pacific equity markets. *Journal of Multinational Financial Management*, 18(3):197–208. (Cited on page 76).

Bellini, F. and Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, 15(5):725–733. (Cited on pages 26 and 27).

Bellini, F., Klar, B., Müller, A., and Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54:41–48. (Cited on pages 23 and 24).

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*. Princeton University Press. (Cited on page 23).

Bensalah, Y. et al. (2000). Steps in applying extreme value theory to finance: A review. *Bank of Canada*. Working paper.

Bensalah, Y. et al. (2002). Asset allocation using extreme value theory. *Bank of Canada*. Working paper. (Cited on pages 35, 39, and 70).

Bibby, B. M. and Srensen, M. (2003). Chapter 6 - hyperbolic processes in finance. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 211 – 248. North-Holland, Amsterdam. (Cited on page 14).

Biffis, E. (2013). Insurance for large-scale commercial risks: Understanding the risk of catastrophes. `https://wwwf.imperial.ac.uk/business-school/research/research-celebration/insurance-for-large-scale-commercial-risks/`. (Cited on page 2).

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654. (Cited on pages 9 and 15).

Blattberg, R. C. and Gonedes, N. J. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *The journal of business*, 47(2):244–280. (Cited on page 14).

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327. (Cited on page 15).

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, pages 542–547. (Cited on page 15).

Bortot, P. and Tawn, J. A. (1998). Models for the extremes of markov chains. *Biometrika*, 85(4):851–867. (Cited on page 81).

Bradley, B. O. and Taqqu, M. S. (2003). Financial risk and heavy tails. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 35 – 103. North-Holland, Amsterdam.

Bradley, B. O. and Taqqu, M. S. (2004a). Asset allocation when guarding against catastrophic losses: a comparison between the structure variable and joint probability methods. *Quantitative Finance*, 4(6):619–636. (Cited on pages 39 and 117).

Bradley, B. O. and Taqqu, M. S. (2004b). An extreme value theory approach to the allocation of multiple assets. *International Journal of Theoretical and Applied Finance*, 7(8):1031–1068. (Cited on pages 39, 70, and 102).

Brooks, C., Burke, S. P., Heravi, S., and Persand, G. (2005). Autoregressive conditional kurtosis. *Journal of Financial Econometrics*, 3(3):399–421. (Cited on page 11).

Bruun, J. T. and Tawn, J. A. (1998). Comparison of approaches for estimating the probability of coastal flooding. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):405–423. (Cited on page 81).

Cambou, M. and Filipović, D. (2015). Model uncertainty and scenario aggregation. *Mathematical Finance*. (Cited on page 24).

Castle, J. and Shephard, N. (2009). *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. OUP Oxford.

Chan-Lau, J. A., Mitra, S., and Ong, L. (2007). Contagion risk in the international banking system and implications for london as a global financial center. *IMF Working Papers*, pages 1–46. (Cited on page 77).

Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2015). An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*. (Cited on pages 23 and 28).

Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 181(1):44–52. (Cited on pages 23 and 115).

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer. (Cited on pages 37, 60, and 64).

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for Extreme Value Analyses. *Extremes*, 2(4):339–365. (Cited on pages 77 and 81).

Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds. *Applied Statistics*, 43(1):139–157. (Cited on page 87).

Consigli, G. (2002). Tail estimation and mean–var portfolio selection in markets subject to financial instability. *Journal of Banking & Finance*, 26(7):1355–1382. (Cited on page 38).

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236. (Cited on page 10).

Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6):593–606. (Cited on pages 23, 24, 25, and 29).

Cramér, H. (1930). *On the mathematical theory of risk*. Centraltryckeriet. (Cited on page 34).

Danielsson, J. (2011). *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab*, volume 588. John Wiley & Sons. (Cited on page 23).

Danielsson, J. and de Vries, C. (1997a). Beyond the sample: Extreme quantile and probability estimation. Tinbergen institute discussion papers, Tinbergen Institute. (Cited on pages 38 and 45).

Danielsson, J. and de Vries, C. G. (1997b). Tail index and quantile estimation with very high frequency data. *Journal of empirical Finance*, 4(2):241–257. (Cited on pages 37, 38, and 66).

Danielsson, J. and de Vries, C. G. (2000). Value-at-risk and extreme returns. *Annales d'Economie et de Statistique*, pages 239–270. (Cited on page 37).

Danielsson, J., Hartmann, P., and de Vries, C. (1998). The cost of conservatism: Extreme returns, value-at-risk, and the basle 'multiplication factor'. *Risk*, 11(1):101–103. (Cited on pages 37, 45, and 115).

Davis, M. H. (2016). Verification of internal risk measure estimates. *Statistics & Risk Modeling*. forthcoming, available at `https://arxiv.org/pdf/1410.4382.pdf`. (Cited on page 27).

Davison, A. C. (1984). Modelling excesses over high thresholds, with an application. In *Statistical extremes and applications*, pages 461–482. Springer. (Cited on page 45).

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 393–442. (Cited on pages 45 and 87).

De Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*. Springer Science & Business Media. (Cited on page 37).

Deaton, A. (1991). Saving and liquidity constraints. *Ecomometrica*, 59(5):221–248. (Cited on page 35).

Deaton, A. (1992). *Understanding consumption*. Oxford University Press. (Cited on page 35).

DellAquila, R. and Embrechts, P. (2006). Extremes and robustness: a contradiction? *Financial Markets and Portfolio Management*, 20(1):103–118.

Diebold, F. X., Schuermann, T., and Stroughair, J. D. (2000). Pitfalls and opportunities in the use of extreme value theory in risk management. *The Journal of Risk Finance*, 1(2):30–35. (Cited on pages 15 and 114).

Eberlein, E., Keller, U., and Prause, K. (1998). New insights into smile, mispricing, and value at risk: The hyperbolic model. *The Journal of Business*, 71(3):371–405. (Cited on page 14).

Embrechts, P. and Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Application*, 1:493–514. (Cited on page 26).

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*, volume 33. Springer-Verlag Berlin Heidelberg. 1st edition 1997. (Cited on pages 1, 37, 64, 85, and 86).

Embrechts, P., Lambrigger, D. D., and Wüthrich, M. V. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes*, 12(2):107–127.

Embrechts, P., Lindskog, F., and Mcneil, A. (2003). Chapter 8 - modelling dependence with copulas and applications to risk management. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 329 – 384. North-Holland, Amsterdam. (Cited on pages 74 and 75).

Embrechts, P., Liu, H., and Wang, R. (2016). Quantile-based risk sharing. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2744142`. (Cited on page 25).

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: Value at Risk and beyond*, pages 176–223. Cambridge University Press. (Cited on page 75).

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. (2014). An academic response to basel 3.5. *Risks*, 2(1):25–48. (Cited on page 28).

Embrechts, P., Resnick, S., and Samorodnitsky, G. (1998). Living on the edge. *Risk*, 11(1):96–100. (Cited on pages 37 and 45).

Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41. (Cited on pages 37, 45, and 74).

Embrechts, P., Wang, B., and Wang, R. (2015a). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19(4):763–790. (Cited on pages 23, 24, and 25).

Embrechts, P., Wang, R., et al. (2015b). Seven proofs for the subadditivity of expected shortfall. *Dependence Modeling*, 3(1):126–140. (Cited on pages 17, 19, and 20).

Emmer, S., Kratz, M., and Tasche, D. (2015). What is the best risk measure in practice? a comparison of standard measures. *Journal of Risk*, 18(2):31–60. (Cited on pages 24, 25, 27, 28, and 29).

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007. (Cited on page 15).

European Parliament (2009a). Omnibus II directive 2014/51/EC. `http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.153.01.0001.01.ENG`. (Cited on page 3).

European Parliament (2009b). Solvency II directive 2009/138/EC. `http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009L0138`. (Cited on pages 3 and 19).

Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., and Focardi, S. M. (2007). *Robust portfolio optimization and management*. John Wiley & Sons. (Cited on pages 9 and 43).

Falk, M., Hüsler, J., and Reiss, R. (2011). Laws of small numbers: Extremes and rare events. 3rd edition. First edition in 1994. (Cited on page 37).

Fama, E. F. (1963). Mandelbrot and the stable paretian hypothesis. *The journal of business*, 36(4):420–429. (Cited on page 10).

Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business*, 38(1):34–105. (Cited on page 10).

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press. (Cited on page 44).

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and osband's principle. *Annals of Statistics*, 44(4):1680–1707. (Cited on page 27).

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447. (Cited on pages 21 and 22).

Frittelli, M. and Gianin, E. R. (2002). Putting order in risk measures. *Journal of Banking & Finance*, 26(7):1473–1486. (Cited on page 21).

Garita, G. and Zhou, C. (2009). Can open capital markets help avoid currency crises? Dnb working papers, Netherlands Central Bank, Research Department. (Cited on page 77).

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453. (Cited on pages 44, 46, and 84).

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762. (Cited on pages 26 and 27).

Granger, C. W. J. (2009). In praise of pragmatic econometrics. In *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, chapter 10, pages 266–290. Oxford University Press, Oxford. (Cited on page 16).

Haile, F. and Pozo, S. (2008). Currency crisis contagion and the identification of transmission channels. *International Review of Economics & Finance*, 17(4):572–588. (Cited on page 77).

Haile, F. D. and Pozo, S. (2006). Exchange rate regimes and currency crises: an evaluation using extreme value theory. *Review of International Economics*, 14(4):554–570. (Cited on page 66).

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of multivariate analysis*, 32(2):177–203. (Cited on page 66).

Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42(6):1887–1896. (Cited on pages 23, 24, and 29).

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust statistics: the approach based on influence functions*. John Wiley & Sons.

Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton university press. (Cited on page 23).

Hartmann, P., Straetmans, S., and de Vries, C. (2004). Asset market linkages in crisis periods. *Review of Economics and Statistics*, 86(1):313–326. (Cited on page 77).

Hartmann, P., Straetmans, S., and De Vries, C. G. (2005). Banking system stability: A cross-atlantic perspective. Technical report, National Bureau of Economic Research. NBER Working Paper No. 11698. (Cited on page 77).

Harvey, C. R. and Siddique, A. (1999). Autoregressive conditional skewness. *Journal of financial and quantitative analysis*, 34(04):465–487. (Cited on page 11).

Harvey, C. R. and Siddique, A. (2000). Conditional skewness in asset pricing tests. *The Journal of Finance*, 55(3):1263–1295. (Cited on page 11).

Hill, B. M. et al. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174. (Cited on pages 84 and 85).

Ho, T.-K. (2008). Extremal analysis of currency crises in taiwan. *Applied Economics*, 40(9):1175–1186. (Cited on page 65).

Hoeffding, W. (1940). *Massstabinvariante korrelationstheorie*. Teubner. Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität Berlin 5. (Cited on page 74).

Hosking, J., 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. (Cited on page 49).

Huang, C.-f. and Litzenberger, R. H. (1988). *Foundations for financial economics*. North-Holland. (Cited on page 9).

Huber, P. J. (1981). *Robust statistics*. Wiley, New york.

Huber, P. J. and Ronchetti, E. (2009). *Robust statistics*. Wiley, New york. 2nd edition. (Cited on pages 23, 24, and 29).

Hyung, N. and De Vries, C. G. (2007). Portfolio selection with heavy tails. *Journal of Empirical Finance*, 14(3):383–400. (Cited on page 38).

IAIS (December 2014). Consultation document. risk-based global insurance capital standard. `http://www.iaisweb.org/page/consultations/closed-consultations/risk-based-global-insurance-capital-standard`. International Association of Insurance Supervisors. (Cited on page 28).

Ingersoll, J. E. (1987). *Theory of financial decision making*, volume 3. Rowman & Littlefield. (Cited on page 9).

Jansen, D. (2001). Limited downside risk in portfolio selection among us and pacific basin equities. *International Economic Journal*, 15(4):1–39. (Cited on page 38).

Jansen, D. and de Vries, C. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *The Review of Economics and Statistics*, pages 18–24. (Cited on page 66).

Jansen, D. W., Koedijk, K. G., and De Vries, C. G. (2000). Portfolio selection with limited downside risk. *Journal of Empirical Finance*, 7(3):247–269. (Cited on page 38).

Jobst, N. J. and Zenios, S. A. (2001). The tail that wags the dog: Integrating credit risk in asset portfolios. *The Journal of Risk Finance*, 3(1):31–43. (Cited on page 11).

Johansen, A. and Sornette, D. (1999). Critical crashes. *arXiv preprint cond-mat/9901035*. (Cited on page 12).

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737. (Cited on page 11).

Karolyi, G. A. (2003). Does international financial contagion really exist? *International Finance*, 6(2):179–199. (Cited on page 75).

Kearns, P. and Pagan, A. (1997). Estimating the density tail index for financial time series. *Review of Economics and Statistics*, 79(2):171–175. (Cited on pages 50, 66, and 87).

Kiesel, R., Rühlicke, R., Stahl, G., and Zheng, J. (2012). Conceptualizing robustness in risk management. *Available at SSRN 2065723*. (Cited on page 24).

Kiesel, R., Rühlicke, R., Stahl, G., and Zheng, J. (2016). The wasserstein metric and robustness in risk management. *Risks*, 4(3):32. (Cited on page 23).

Knetsch, J. L. and Thaler, R. H. (1990). Daniel kahneman. *The Journal of Political Economy*, 98(6):1325–1348. (Cited on page 35).

Koedijk, K. G., Schafgans, M. M., and De Vries, C. G. (1990). The tail index of exchange rate returns. *Journal of international economics*, 29(1-2):93–108. (Cited on page 66).

Kolm, P. N., Tütüncü, R., and Fabozzi, F. J. (2014). 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research*, 234(2):356–371.

Kon, S. J. (1984). Models of stock returns - a comparison. *The Journal of Finance*, 39(1):147–165. (Cited on page 14).

Kou, S. and Peng, X. (2014). On the measurement of economic tail risk. *Available at SSRN 2381651*. (Cited on pages 23 and 24).

Kou, S., Peng, X., and Heyde, C. C. (2013). External risk measures and basel accords. *Mathematics of Operations Research*, 38(3):393–417. (Cited on pages 23 and 24).

Krätschmer, V., Schied, A., and Zähle, H. (2012). Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *Journal of Multivariate Analysis*, 103(1):35–47. (Cited on page 24).

Krätschmer, V., Schied, A., and Zähle, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastic*, 18(2):271–295. (Cited on pages 23, 24, and 26).

Kratz, M., Lok, Y. H., and McNeil, A. J. (2016). Multinomial var backtests: A simple implicit approach to backtesting expected shortfall. *arXiv preprint*. Available at `https://arxiv.org/abs/1611.04851`. (Cited on page 27).

Lambert, N., Pennock, D. M., and Shoham, Y. (2008). Eliciting properties of probability distributions: the highlights. *ACM SIGecom Exchanges*, 7(3):9. (Cited on page 26).

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187. (Cited on pages 81 and 82).

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499. (Cited on pages 77, 81, and 82).

Ledford, A. W., Tawn, J. A., et al. (1998). Concomitant tail behaviour for extremes. *Advances in applied Probability*, 30(1):197–215. (Cited on page 82).

Levy, H. and Sarnat, M. (1972). Safety firstan expected utility principle. *Journal of Financial and Quantitative Analysis*, 7(03):1829–1834.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The review of economics and statistics*, pages 13–37. (Cited on page 9).

Longin, F. (2005). The choice of the distribution of asset returns: How extreme value theory can help? *Journal of Banking & Finance*, 29:1017–1035. (Cited on page 65).

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The journal of finance*, 56(2):649–676. (Cited on pages 66, 75, 76, and 80).

Longin, F. M. (2000). From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance*, 24(7):1097–1130. (Cited on page 37).

Longin, F. M. (2001). Beyond the var. *The Journal of Derivatives*, 8(4):36–48. (Cited on page 37).

Lux, T. (2000). On moment condition failure in german stock returns: an application of recent advances in extreme value statistics. *Empirical economics*, 25(4):641–652. (Cited on page 66).

Mancini, L. and Trojani, F. (2011). Robust value at risk prediction. *Journal of Financial Econometrics*, 9(2):281–313.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419. (Cited on pages 8 and 10).

Manganelli, S. and Engle, R. F. (2001). Value at risk models in finance. *European Central Bank working paper*, (75). (Cited on page 66).

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1):77–91. (Cited on pages 8 and 33).

Markowitz, H. (1959). *Portfolio selection: efficient diversification of investments*, volume 16. Cowles Foundation Monograph, Yale University Press. (Cited on pages 9 and 16).

Markowitz, H. (2014). Mean–variance approximations to expected utility. *European Journal of Operational Research*, 234(2):346–355. (Cited on page 9).

Markowitz, H. M. (1999). The early history of portfolio theory: 1600-1960. *Financial Analysts Journal*, 55(4):5–16. (Cited on page 33).

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN bulletin*, 27(01):117–137. (Cited on page 37).

McNeil, A. J. (1998). Calculating quantile risk measures for financial return series using extreme value theory. *ETH Zürich, Departement Mathematik*. (Cited on page 37).

McNeil, A. J. (1999). Extreme value theory for risk managers. *ETH Zürich, Departement Mathematik*. (Cited on page 1).

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300. (Cited on pages 23, 37, 45, 114, and 115).

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press. 1st Edition. (Cited on pages 17, 23, 28, 37, 50, 55, 68, and 87).

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press. 2nd Edition. (Cited on pages 1, 23, and 37).

Neftci, S. N. (2000). Value at risk calculations, extreme events, and tail estimation. *The Journal of Derivatives*, 7(3):23–37. (Cited on pages 37 and 38).

Osband, K. (1985). *Providing incentives for better cost forecasting*. PhD thesis, University of California, Berkeley. (Cited on page 26).

Pais, A. and Stork, P. (2011). Contagion risk in the australian banking and property sectors. *Journal of Banking & Finance*, 35(3):512–531. (Cited on page 77).

Pesenti, S. M., Millossovich, P., and Tsanakas, A. (2016). Robustness regions for measures of risk aggregation. *Dependence Modeling*, 4(1). (Cited on pages 17, 24, and 25).

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131. (Cited on pages 45, 48, and 65).

Poon, S.-H., Rockinger, M., and Tawn, J. (2003). Modelling extreme-value dependence in international stock markets. *Statistica Sinica*, pages 929–953. (Cited on pages 73, 76, 77, 84, and 102).

Poon, S.-H., Rockinger, M., and Tawn, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Review of financial studies*, 17(2):581–610. (Cited on pages 73, 76, 77, 84, and 102).

Post, T. and Vliet, P. (2004). Conditional downside risk and the capm. *ERIM Report Series Research in Management*, (ERS-2004-048-F&A). (Cited on page 15).

Pownall, R. A. and Koedijk, K. G. (1999). Capturing downside risk in financial markets: the case of the asian crisis. *Journal of International Money and Finance*, 18(6):853–870. (Cited on pages 37 and 38).

Prause, K. (1999). *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures*. PhD thesis, Citeseer. (Cited on page 14).

Rachev, S. and Mittnik, S. (2000). Stable paretian models in finance. (Cited on page 15).

Rachev, S. T., Mittnik, S., and Paolella, M. S. (1998). Stable paretian modeling in finance: some empirical and theoretical aspects. In *A practical guide to heavy tails*, pages 79–110. Birkhauser Boston Inc. (Cited on page 15).

Reiss, R.-D., Thomas, M., and Reiss, R. (2007). *Statistical analysis of extreme values*, volume 3. Springer. 3rd edition. First edition in 1997. (Cited on page 37).

Resnick, S. I. (1987). Extreme values, regular variation and point processes. (Cited on pages 37 and 83).

Rocco, M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys*, 28(1):82–108. (Cited on page 15).

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. In *Journal of Risk*, volume 2, pages 21–42. (Cited on page 20).

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471. (Cited on page 20).

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of economic theory*, 13(3):341–360. (Cited on page 9).

Roy, A. D. (1952). Safety-first and the holding of assets. *Econometrica*, 20(3):431–449. (Cited on pages 16, 33, 34, 35, 53, and 59).

Rubinstein, M. (2002). Markowitzs portfolio selection: A fifty-year retrospective. *The Journal of finance*, 57(3):1041–1045. (Cited on page 9).

Sandström, A. (2010). *Handbook of solvency for actuaries and risk managers: theory and practice*. CRC Press, Tailor & Francis Group. (Cited on pages 19 and 28).

SCOR (2008). From principle-based risk management to solvency requirements. *Swiss solvency test documentation, Zurich*. (Cited on page 19).

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442. (Cited on page 9).

Sibuya, M. (1959). Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210. (Cited on page 80).

Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Publications de l'Institut de Statistique de l'Université Paris 8. (Cited on page 74).

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90. (Cited on page 62).

Smith, R. L. (1987). Estimating tails of probability distributions. *The annals of Statistics*, 15:1174–1207. (Cited on pages 53 and 85).

Steinbach, M. C. (2001). Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM review*, 43(1):31–85. (Cited on page 9).

Susmel, R. (2001). Extreme observations and diversification in latin american emerging equity markets. *Journal of International Money and Finance*, 20(7):971–986. (Cited on page 38).

Tarullo, D. et al. (2008). Banking on basel: The future of international financial regulation. *Peterson Institute Press: All Books*. (Cited on page 19).

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415. (Cited on page 80).

Tobin, J. (1958). Liquidity preference as behavior towards risk. *The review of economic studies*, 25(2):65–86. (Cited on page 9).

Tsafack, G. (2009). Asymmetric dependence implications for extreme risk management. *Journal of Derivatives*, 17(1):7.

Vilasuso, J. and Katz, D. (2000). Estimates of the likelihood of extreme returns in international stock markets. *Journal of Applied Statistics*, 27(1):119–130. (Cited on page 66).

Von Neumann, J. and Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton university press. (Cited on page 9).

Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100(4):794–815. (Cited on page 62).

Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 101(4):971–983. (Cited on page 62).

Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*. forthcoming, available at `https://arxiv.org/pdf/1303.1690.pdf`. (Cited on pages 26 and 27).