Rapporto n. 3 2010
dmsia unibg.it

M. Rocco
Exteme value Theory for Finance: A Survey

Serie Ricerca

Dipartimento
di Matematica, Statistica,
Informatica e Applicazioni
"Lorenzo Mascheroni"

UNIVERSITÀ DEGLI STUDI DI BERGAMO

Extreme Value Theory for Finance: A Survey

Marco Rocco
Dipartimento di Matematica, Statistica, Informatica e Applicazioni
Università degli Studi di Bergamo
Via dei Caniana 2, 24127 Bergamo, Italy
e-mail: marco.rocco@unibg.it

Abstract

Extreme value theory is concerned with the study of the asymptotical distribution of extreme events, that is to say events which are rare in frequency and huge with respect to the majority of observations. Statistical methods derived from this theory have been increasingly employed in finance, especially in the context of risk measurement. The aim of the present study is two-fold. The first part delivers a critical review of the theoretical underpinnings of extreme value theory. The second part provides a survey of some major applications of extreme value theory to finance, namely its use to test between different distributional assumptions for the data, Value-at-Risk and Expected Shortfall calculations, asset allocation under safety-first type constraints and the study of contagion and dependence across markets under stress conditions.

JEL C10, C16, G10, G20, G21

Keywords extreme value theory, risk management, fat-tailed distributions, Value-at-Risk, systemic risk, asset allocation

Acknowledgements The present paper was written during a research stay at the Economic Outlook and Monetary Policy Department of the Economic Research Area of the Bank of Italy. I am grateful to Giuseppe Grande (Bank of Italy) for several helpful comments and suggestions concerning this work.

Contents

Introduction

Uncertainty and risk[1] enter the economy and the global financial system at least in a two-fold manner, both through the temporal component (with time accounting for any possible change to the present state of nature) and through the cross-sectional dimension (mainly related to imperfect knowledge of the states of nature themselves). This feature not just being merely of theoretical concern is proved by many concrete events. During the last two decades several crises have affected financial markets, e.g. the stock market crash of October 1987 (with the S&P500 index falling more than 20% on a single day, the Black Monday, October 19), the Asian financial crisis of 1997-1998, the hedge fund crisis of 1998 and the credit crisis begun in 2007. Thus, the problem of forecasting risk in order to prevent negative events to impact against one's own investments, or, in a broader sense, the problem of managing risk, has become a major concern for both financial institutions and market regulators. As a consequence, also due to a sharp deepening of the theoretical framework and remarkable improvements in computational tools, risk management is striving to assess itself as an independent scientific discipline, though naive enthusiasm misinterpreting it is likely to fade away after the latest crisis[2].

In order to pursue its aim, the first task of risk management is to try to quantify and measure risk. Ideally, the best and most informative risk measure for the returns of some financial activity is given by the whole tail of the distribution of the returns themselves. Anyhow, this distribution is in general:

(a) unknown (we only know the time series of the returns and we can just guess the exact distribution from it);

(b) difficult to deal with (in the sense that it contains, in a certain sense, too much information, while a tractable risk measure should make some synthesis of the richness of data that characterizes the exact distribution).

As an answer to the latter issue, several risk measures have been adopted during the last sixty years, the main ones being: volatility, in the framework of portfolio selection as depicted by Markowitz (1952); Value-at-Risk (VaR, for short), first used after the 1987 crash, subsequently developed and made public by J. P. Morgan in 1994 and now at the core of international regulation as settled by the first pillar of the Basel II Accord; Expected Shortfall (ES), probably a better risk measure rather than VaR, since it really takes into account the whole of the tail of the distribution and, moreover, it fulfils the properties required by a coherent risk measure, according to the definition of Artzner *et al.* (1999).

[1] As it is well known, the two concepts do not necessarily coincide. Anyway, we will not discuss here the variety of definitions and interpretations the notion of risk has undergone in the financial literature.

[2] Many people (among scholars, professionals and journalists) have addressed the issue of the failure of quantitative methods of finance and risk management in adequately preventing, or at least forecasting, the crisis (see Embrechts, 2009 for a defence of risk management and an attempt at clarifying the actual scope of its techniques). In fact, it looks like finance as a science is growing up: after the seminal success of its mathematical formulation and fast developments of the discipline, a period of positivistic self-confidence followed (misinterpreting the actual scope of the new techniques and underestimating the limitations implicit in the assumptions on which the theory relies), which could now be downsized, after the new insights brought about by the crisis. This could be a great opportunity for the growth of finance as a science and for the development of an epistemologically mature approach towards it (as it happened to other disciplines, like physics, for instance, which represents the prototype of the modern western concept of science).

Though many contributions question the effectiveness (or even the appropriateness) of these three quantities (especially the first two) as risk measures, they are still the standard measures one has to confront with. Compared to the tail of the distribution, they summarize all the information contained in it in a single number. Therefore they are much easier to deal with and provide a tool apt to decision making. Anyway, the former issue we have identified above, i.e. the lack of knowledge concerning the distribution of returns, leaves us with the problem of how actually computing these measures.

The fact that VaR and ES only deal with extreme quantiles of the distribution[3], disregarding the centre of the distribution itself, suggests the use of extreme value theory (EVT, from now on) as an effective tool for providing reliable estimates of them.

EVT is a well developed theory in the field of probability, that studies the distribution of extreme realizations of a given distribution function, or of a stochastic process, satisfying suitable assumptions. The foundations of the theory were set by Fisher and Tippett (1928) and Gnedenko (1943), who proved that the distribution of the extreme values of an iid sample from a cumulative distribution function F, when adequately rescaled, can converge (and indeed does converge, for the majority of known cumulative distribution functions F) to one out of only three possible distributions[4].

The most powerful feature of this result is the fact that, to some extent, the type of asymptotic distribution of extreme values does not depend on the exact cumulative distribution function F of returns. This allows one to neglect the precise form of F, following a non-parametric or a semi-parametric method to estimate Value-at-Risk. This is particularly important, according to item (a) above and according to the fact that, though it is known that financial time series usually exhibit skewed and fat-tailed distributions, there is no complete agreement on what distribution could fit them best.

Moreover, in principle EVT-based estimates of VaR could be more precise and reliable than the usual ones, given that EVT directly concentrates on the tails of the distribution, thus avoiding a major flaw of parametric approaches, whose estimates are somehow biased by the credit they give to the central part of the distribution, thus underestimating extremes and outliers, which are exactly what one is interested in when calculating VaR.

Finally, a third reason that makes EVT especially promising when trying to measure risk is the possibility it provides to concentrate on each one of the two tails of the distribution independently, thus allowing a flexible approach which can take skewness (a typical feature of financial time series) of the underlying distribution into account.

These three main advantages of an EVT approach to risk management could be summarized in the motto of DuMouchel (1983): "Letting the tails speak for themselves", which is particularly appealing as risk management is mainly focused on avoiding big unexpected losses and sudden crashes rather than long sequences of medium-sized losses (this is mainly a consequence of the market crises experienced in the recent past and the empirical observation that the final position of a portfolio is more affected by a few extreme movements in the market rather than by the sum of many small movements[5]).

[3] See section 7 below for the definition of VaR and ES.

[4] See section 1.1 below for a mathematically precise statement of this theorem.

[5] "Whilst it is hard to determine the holding period of an average investor, we know that portfolio performance, and therefore the investor's outcomes, are dependent on a few days of trading. Average daily returns are near zero but we know on any one day that extreme and large scale returns can occur." (Cotter, 2007a, p. 1).

Estimation of VaR and ES is nowadays the most popular application of EVT to finance, but it is not the only possible one. A critical (though not complete) survey of the main financial applications of EVT is the goal of the present paper. The material is organized in two parts. The first one is devoted to recall the basic elements of extreme value theory, highlighting its theoretical assumptions and analysing the main issues concerning its translation into a set of statistical methods. The second part reviews the application of these methods to empirical research in finance, showing how mathematical machinery elaborated in the first part can cast new light on some relevant financial matters (the aim of this part is rather to put some order in the growing literature on the subject, than to give a detailed account of each single contribution). Since readers with different backgrounds can be more interested in the former or in the latter part, we premise the paper with a quite detailed non-technical overview that enables anyone to focus directly on the paragraphs he/she is mainly interested in.

Non-technical Overview

Part I – Basic Theoretical Tools and Methodological Issues

Section 1 – Different Approaches to EVT

Given an unknown distribution F (think of returns on some financial activity, for instance), extreme value theory (EVT) is interested in modelling the tail of F only, without committing itself to making any specific distributional assumption concerning the centre of the distribution.

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

The difference between the two parametric approaches is related to the dissimilar (though complementary) meaning that they assign to the notion of "extreme value".

Consider N iid random variables X_i, $i = 1, \ldots, N$, representing *positive* losses.

(a) The first parametric approach, the *block maxima* method (see 1.2), divides the given N observations sample into m subsamples of n observations each (*n-blocks*) and picks the maximum M_k ($k = 1, \ldots, m$) of each subsample (a so-called *block maximum*).

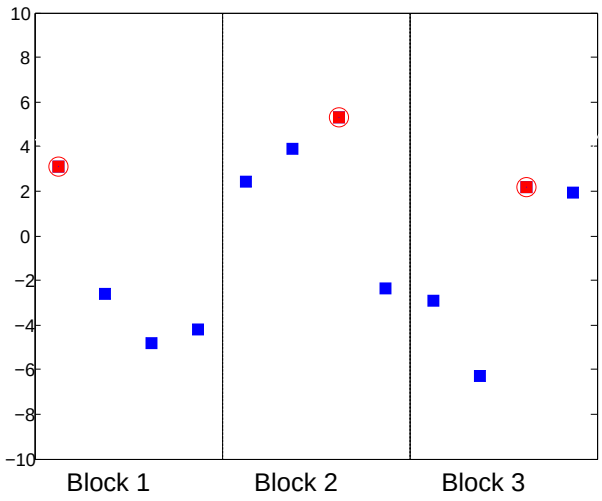

Figure 1: The block-maxima method.

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

The three asymptotic distributions of block maxima can be expressed in a unified manner by means of the *generalized extreme value* (GEV) distribution, a parametric expression depending on a real parameter ξ, known as the *shape parameter*. The three cases we have just mentioned correspond to $\xi > 0$ (called the Fréchet case), $\xi = 0$ (Gumbel) and $\xi < 0$ (Weibull), respectively.

(b) The second parametric approach, the *threshold exceedances* method (see 1.3), defines extreme values as those observations that exceed some fixed high threshold u. The aim of this method is therefore to model the distribution of the exceedances over u, that is to say, the random variables $Y_j = X_j - u$, for those observations (returns, in our case) X_j that exceed u (i.e. such that $X_j > u$).

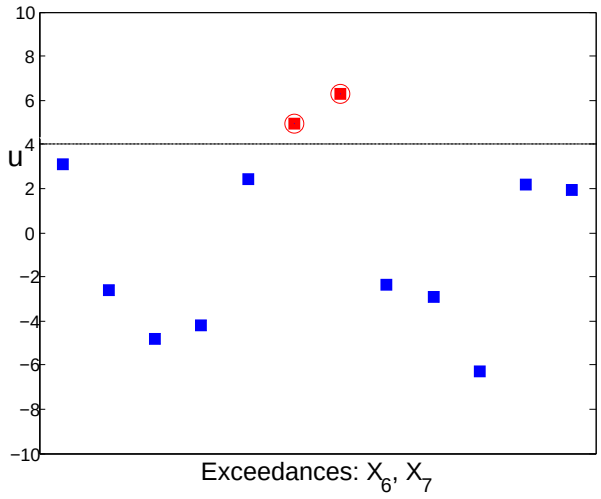

Figure 2: The threshold exceedances method.

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

(c) Both previous approaches are parametric (or semi-parametric), as they fit a parametric model (usually via maximum likelihood estimation) to the upper tail[6] of the distribution, though neglecting what happens at the centre of the distribution. Anyway, if one does not want to fit a model to the tail either, the possibility exists to directly estimate the shape parameter[7] ξ, pursuing a non-parametric approach (1.1). Several estimators accomplishing this task have been proposed, but the most frequently employed one is by far the Hill estimator.

[6] Analogous results hold for the lower tail as well.

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

Section 2 – Setting the Cut-off

Any of the three approaches to EVT we have considered in the previous section entails the choice of an adequate cut-off between the central part of the distribution and the upper tail, i.e. a point separating ordinary realizations of the random variable considered from extreme realizations of the same variable. When working with threshold exceedances, the cut-off is induced by the threshold u, while, in the block maxima method, it is implied by the number m of blocks (or the amplitude n of each block)[8].

This is a very delicate issue concerning statistical methods of EVT, since it entails a trade-off between bias and variance of the estimates of the shape parameter ξ. Indeed, reasoning for instance on threshold exceedances, if u is set too low (u_3 in the figure), many ordinary data are taken as extreme ones, thus yielding biased estimates. On the contrary, an excessively high threshold (as u_1) leaves us with scant extreme observations, not enough to obtain efficient estimates. In both cases, the resulting estimates are flawed and may lead to erroneous conclusions when assessing risk.

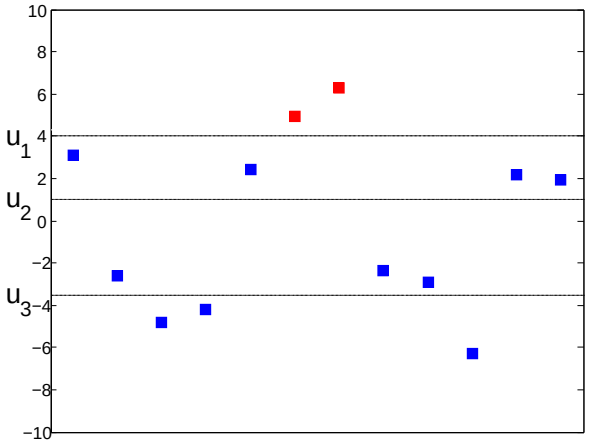

Figure 3: Choice of the threshold u.

Four main solutions to this problem can be found in the literature:

(a) using common sense-based choices of the cut-off (e.g., choose u in such a way that about 5%-15% of the data are thought of as extreme observations[9]);

(b) employing graphical methods (known as *Hill plots* when the Hill estimator is used), displaying the estimated values of ξ as a function of the cut-off, in order to find out some interval of candidate cut-off points that yield stable estimates of ξ (corresponding to a horizontal line in the Hill plot);

[8] Analogously, also the Hill estimator entails setting a cut-off, since it is based on order statistics and requires an explicit choice of one of them, which works as a cut-off.

[9] Notice that the very name of extreme value theory may be somewhat misleading on some occasions, given that, in practice, when fitting an EVT model to data, one often considers also observations that are not "extreme", if compared to the whole time series (for instance, one could argue that a cut-off yielding 10% of data as extreme includes in the tail of the distribution also quite ordinary observations).

(c) making Monte Carlo simulations and then choosing the cut-off that minimizes a statistical quantity (the mean squared error) that takes both bias and variance of the estimates into account;

(d) implementing algorithms (based for instance on the bootstrap method) that endogenously pick out the cut-off that is most suited to the data at hand.

Section 3 – Dependence in the Data

The basic assumption of EVT is that data be iid, which however is not true for most financial time series. Therefore, using EVT without properly considering the dependence structure of the data, yields incorrect estimates, possibly resulting in unexpected losses, or in excessively conservative positions (both to be avoided for risk management purposes).

To take dependence of the data into consideration, two main approaches are usually employed.

(a) If the time series is strictly stationary (as it is often assumed for financial time series), then an additional parameter can be estimated, the *extremal index* ϑ, which accounts for the clustering of extremal values due to dependence.

(b) Alternatively, one can explicitly model the dependence structure, fitting some GARCH-type model to the data. If the standardized residuals exhibit a roughly iid structure, one can then apply EVT to them, instead of directly to the data.

The latter method proved to work definitely well when using EVT for estimating quantile-based measures of risk, such as Value-at-Risk or Expected Shortfall, and it seems to be sufficiently robust to yield good estimates even in the case when the GARCH model is misspecified to some extent. A third method, based on a point process approach, can be employed to explicitly model the behaviour of extremes within clusters.

Section 4 – The Multivariate Setting

When studying extremes of multivariate time series, the dependence between extreme values of the different components of the series can also play a very important role.

In this respect, two main cases may occur, namely *asymptotically independent* and *asymptotically dependent* time series (4.1). In the first case, reasoning for simplicity on a bivariate time series, the two components of the series are not necessarily independent (and indeed they usually are not), but, as we consider "sufficiently high" thresholds (i.e., growing to infinity), the occurrence of an extreme realization of the first random variable does not affect the probability that we observe an extreme realization of the second variable as well. The opposite holds in the case of asymptotic dependence.

Explicit modelling of multivariate time series via *multivariate extreme value theory* (MEVT) can be accomplished by means of copula functions, according to an important result (see 4.2) stating that the asymptotic distribution of multivariate extremes is the combination of:

- the asymptotic distributions of extreme values of the margins, as derived in the univariate framework (section 1);
- the asymptotic behaviour of the copula.

The latter is less well understood than the former and in practice one often uses some easy way to implement copula to model extremal dependence, though it is not necessarily the correct one.

Finally, notice that both the block maxima method and the threshold exceedances approach can be generalized to the multivariate setting (4.2 and 4.3, respectively).

Section 5 – The Choice of the Data-Set

This section is devoted to a brief discussion of the issues concerning the frequency of the data employed in empirical studies using EVT and the choice of an appropriate time window for the observations. These topics can play a significant role when applying EVT, because of the dichotomy inscribed in its nature: on the one hand, EVT requires a lot of data, since its results are asymptotical, but, on the other hand, it is necessarily faced with a scarcity of data, given that it concentrates on the tails of the distribution and extreme events are, by definition, rare.

Part II – Financial Applications

Section 6 – Testing for Different Distributional Assumptions

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

As a consequence, a lot of research has focused on the appropriateness of fatter-tailed distributions, mainly the Student t distribution and non-normal stable laws. The main problem is that the two distributions are not nested, so that the choice between one or the other is not immediate.

In this context EVT offers a valuable contribution, providing a direct comparison between the two families and thus allowing to test which of the two better suites to the data. This is possible since both families have the same type of asymptotic distribution of threshold exceedances, but with values of the shape parameter that range in two non-overlapping intervals ($\xi \in]0, 0.5[$ for the Student t and $\xi > 0.5$ for stable laws). Then, fitting a GPD to threshold exceedances from the given data, the value of the estimated shape parameter provides evidence in favour of one or the other of the two families of distributions. Overall, findings on this topic seem to support the assumption that many financial data follow a Student t distribution and have finite second moment.

Section 7 – Market Risk: VaR and ES Estimation

The main concern in employing EVT techniques in finance is related to risk management. The interest in accurately modelling the lower tail of the unknown distribution underlying observed returns of a financial activity is deeply rooted in the very essence of risk management, which aims to prevent one's own position to yield huge unexpected losses. Some key features of EVT make it an ideal tool to assess risk, especially when using quantile-based measures of risk like Value-at-Risk (VaR) or Expected Shortfall (ES).

(a) Standard parametric methods for calculating VaR fit a parametric distribution to the whole amount of available data, thus incurring the risk of obtaining unreliable estimates, biased because of the predominance of central observations over extreme ones (while VaR, by definition, does not take into account central observations). EVT, on the contrary, only considers the tail of the unknown distribution F, neglecting its central part. Therefore it allows higher accuracy in modelling the occurrence of events belonging to the tail of the distribution, compared to parametric approaches.

(b) On the same lines, not trying to fit a parametric model to the whole distribution reduces the model risk related to parametric approaches to VaR calculation.

(c) Anyway, at the same time EVT still offers a parametric device (though restricted to the tails of the distribution), thus providing the possibility of out-of-sample projections, even for extreme quantiles (a possibility which is in turn denied to historical simulation methods, if not enough data are available).

(d) Finally, EVT allows separate analyses of the lower and upper tails of the distribution. This may be important in many applications, since several financial time series are asymmetric (usually big losses are more likely to occur than big positive returns).

A remarkable amount of work has been done comparing EVT calculations of VaR to those obtained via standard methods. A fairly agreed conclusion is that EVT computation of VaR provides more accurate estimates than traditional methods, especially when we are interested in very high quantiles, i.e. when we need to compute VaR_α for $\alpha \geq 0.99$. Furthermore, when calculating conditional VaR, namely values of VaR based on past information, a GARCH prefiltering of data, as described in section 3, proves to be a valuable procedure.

Section 8 – Asset Allocation

The effectiveness of EVT to VaR calculation can be exploited when solving portfolio selection problems on the basis of a mean-VaR trade-off (instead of the usual mean-variance one). That is to say, the problem of choosing an asset allocation yielding maximum returns, subject to a constraint that limits the upper value of VaR that is tolerated; or, the other way around, the problem of finding the portfolio to which minimum risk is associated, among those guaranteeing a given lower bound for the expected return.

This problem, strictly related to the notion of a safety-first investor studied by Roy (1952) and Arzac and Bawa (1977), can be solved in a more accurate way when employing EVT to calculate VaR, according to the evidence collected in section 7.

Univariate EVT can be used when studying portfolio returns (see Bensalah, 2002, for instance). An important result in this context is that EVT solutions to mean-VaR problems are quantile-based (in the sense that the optimal allocation changes as the confidence level α of VaR_α varies), at odds with solutions based on estimation of VaR via normal distribution, which are independent of the value of α.

Multivariate EVT, on the other hand, offers the possibility of taking into account extremal dependence among different assets in the portfolio, which can diminish the benefits of diversification. Anyway, the findings of Bradley and Taqqu (2004a, 2004b) provide evidence that many financial markets are asymptotically independent (so that diversification can effectively reduce portfolio extreme risk) and that univariate EVT methods yield acceptable results for ordinary quantiles (e.g., $\text{VaR}_{0.95}$), thus being a viable approach, as they are by far simpler than multivariate EVT methods.

Section 9 – Dependence Across Markets: Correlation and Contagion

A natural field of application of multivariate EVT is the study of dependence and contagion among markets, either considered as dependence between different countries for a given financial sector, or as dependence between different financial markets of the same country.

In the former sense, the very influential paper Longin and Solnik (2001) uses multivariate EVT to test the common belief that correlation between stock markets increases in volatile periods and concludes that, in fact, correlation is related to the market trend (increasing in bear markets) rather than to volatility *per se*. EVT has been also applied to study contagion risk in banking sector and during currency crises.

In the latter sense, different analyses have been conducted, concerning for instance dependence between stock and government bond markets, or stock and exchange markets, or contagion between banking and real estate sectors (motivated by the credit crisis).

Section 10 – Further Applications

This section simply mentions other possible financial applications of EVT that have been investigated in the literature, but that we have not presented here. Important areas of application that have been analyzed quite extensively are the study of currency crises (detecting and dating crises) and operational risk measurement.

Part I

Basic Theoretical Tools and Methodological Issues

In this first part of the paper we review the foundations of extreme value theory (EVT) as a part of probability theory and the main statistical methods developed for it. The exposition is organized in such a way as to point out some issues that are relevant to the application of EVT to empirical analysis (this part will be fairly general, though we are only concerned with financial applications, to which the second part is devoted). For a systematic and detailed account of EVT and its applications many books are now available, namely: Beirlant *et al.* (2004); Coles (2001); de Haan and Ferreira (2006); Embrechts *et al.* (1997); Falk *et al.* (2004); Finkenstädt and Rootzén (2004); Galambos (1987); Kotz and Nadarajah (2000); Leadbetter *et al.* (1983); Reiss and Thomas (1997); Resnick (1987). In our presentation of EVT, we will principally follow Embrechts *et al.* (1997) and McNeil *et al.* (2005, Chapter 7) and we will always consider univariate EVT, except in 4.

1 Different Approaches to EVT

The first issue one is faced with when dealing with EVT is the choice of what approach to follow. Indeed EVT provides several different techniques upon which one can rely in order to make estimates of the quantity of interest. The main option is related to the possibility of either directly estimating the shape parameter (*non-parametric approach*; see 1.1 below), that characterizes the behaviour of the tail of the distribution, or fitting a parametric distribution to which the extremes of the underlying distribution eventually obey (*parametric approach*). The latter approach, in turn, can be pursued in a two-fold manner, according to the definition of "extreme value" one chooses: either taking as extremes the maxima of iid samples of the same dimension (*block maxima*, 1.2), or setting a given high threshold u and considering as extreme any observation exceeding u (*threshold exceedances*, 1.3).

The non-parametric approach, especially the one based on the Hill estimator, was probably the most used in the first applications of EVT to finance (roughly, during the Nineties). Later on, also the parametric approaches have been successfully employed and there is evidence that, in some applications, they can work even better[10].

1.1 Non-parametric Approaches: Tail Index Estimators

Comprehensive reviews of different non-parametric approaches can be found in Pictet *et al.* (1996) and Panaretos and Tsourti (2003) and in the above mentioned books on EVT.

[10] A list of pros of using the GPD approach (1.3) instead of the Hill estimator (1.1) is provided by McNeil and Frey (2000, p. 287), 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.

■ ***Fisher–Tippett and Gnedenko Theorems***

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

More precisely, given n iid random variables[11] $X_1, \ldots, X_n$, let M_n denote the maximum of this collection, i.e.

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

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

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

i.e. a degenerate distribution[12].

In order to obtain a non-degenerate limit distribution for the maxima, Fisher and Tippett (1928) standardized the random variable M_n by means of an affine transformation with a *location parameter* d_n and a positive *scale parameter* $c_n > 0$, thus reducing themselves to study the asymptotic distribution of

$$\frac{M_n - d_n}{c_n}.$$

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

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

[11] In sections 1 and 2 we will always assume that the data be drawn from iid random variables. We will discuss to what extent one can relax this assumption in section 3.

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

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

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

Notice that the existence of the sequence of norming constants[13] (location and scale parameters) is not always guaranteed, though for virtually any textbook distribution it has been proved.

Furthermore, when this sequence exists, it is not unique. Anyway, the value of the shape parameter does not depend on the particular sequence chosen to standardize the maxima, that is, ξ is determined by F in a univocal manner.

According to this remark, the following definition is justified.

Definition 1.1 *If there exists a sequence of norming constants such that the asymptotic distribution of the standardized maxima is of Fréchet (or Gumbel, or Weibull) type with shape parameter ξ, we will say that F is in the* maximum domain of attraction *of the Fréchet (Gumbel, or Weibull, respectively) distribution H_ξ.*

In symbols: $F \in MDA(H_\xi)$.

■ *Fréchet Family and the Tail Index*

We first recall the following definition.

Definition 1.2 *A positive, Lebesgue-measurable function L on $]0, +\infty[$ is* slowly varying at infinity *if*

$$\lim_{x \to +\infty} \frac{L(tx)}{L(x)} = 1, \qquad \forall t > 0.$$

[13] Adopting the notation $F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$, the norming constants can be chosen as follows (Embrechts *et al.*, 1997): for the Fréchet distribution, $d_n = 0$ and $c_n = F^{\leftarrow}\left(1 - \frac{1}{n}\right) = n^{\xi} L_1(n)$, with L_1 a slowly varying function (see the main text below for the definition); for the Gumbel distribution, $d_n = F^{\leftarrow}\left(1 - \frac{1}{n}\right)$ and $c_n = a(d_n)$, where the function a can be chosen as $a(x) = \int_x^{+\infty} \frac{\overline{F}(y)}{\overline{F}(x)} dy$; for the Weibull distribution, $d_n = x_F$, where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$, and $c_n = x_F - F^{\leftarrow}\left(1 - \frac{1}{n}\right) = \frac{1}{n^{\xi}} L_1(n)$, with L_1 a slowly varying function.

The prototypical example of a slowly varying function is $L(x) = \ln x$.

One can prove (Embrechts *et al.* 1997, Theorem 3.3.7) that:

Theorem 1.3 *F is in the maximum domain of attraction of a Fréchet distribution with shape parameter ξ if and only if*

$$\overline{F}(x) = 1 - F(x) = x^{-1/\xi} L(x), \quad x > 0, \tag{1}$$

where L is a slowly varying function.

Condition (1) means that the upper tail of F decays as a power function, multiplied by a slowly varying function.

The Fréchet extreme value distribution is particularly important to finance, since most financial series are fat-tailed, thus displaying an asymptotic distribution of extremes that is of Fréchet type.

The parameter $\alpha := \frac{1}{\xi} > 0$ is known as the *tail index* and is directly related to the tail behaviour of the distribution. For instance, one can prove that:

- stable laws with index less than 2 satisfy (1) with α equal to their characteristic index, so that $\alpha \in]0, 2[$;
- Student t distribution satisfies it with α equal to its degrees of freedom.

The interest in the applications is often related to the estimation of the tail index, given a sample $X_1, \ldots, X_n$ of iid data drawn from a distribution F satisfying (1).

Several estimators of the tail index have been studied in the literature. We present here the main ones, among which the Hill estimator is by far the most used[14].

The Hill estimator often outperforms other estimators as far as we are concerned in the estimation of the tail index, i.e. in the case when our cumulative distribution function is in the MDA of a Fréchet distribution. Otherwise the Hill estimator does not work and one has to employ some other estimator. As already noticed, this is not usually a major issue, since many financial time series have an extreme value distribution of Fréchet type[15] and the strong connection between the tail index and the parameters of stable laws and the Student t distribution explains why the earliest applications of EVT to finance were especially focused on estimating the tail index.

■ *Hill Estimator*

After considering the order statistics $X_{1,n} \geq \ldots \geq X_{n,n}$ of the data[16], one can write the Hill estimator, first proposed by Hill (1975), as

[14] Some recent applied papers which use the Hill estimator are, for instance, Haile and Pozo (2006), Cotter (2007b), Lestano and Jacobs (2007).

[15] Anyway, Longin and Solnik (1997), for instance, working on data from international equity markets, find that not only Fréchet, but also Gumbel and Weibull distributions can be obtained.

[16] We are assuming to deal with positive losses, so that we can take the logarithms of the data.

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

i.e. the inverse of the average log-exceedance above the threshold[17] $\ln X_{k,n}$.

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

The dependence of the Hill estimator on k is a critical issue for the application of the method to empirical studies and we will discuss it in section 2.

On the other hand, the Hill estimator has undergone both deep theoretical study and intensive application, displaying very good performance, competitive (and in some cases even superior) with respect to other EVT approaches. From a theoretical viewpoint, the favourable consideration[18] towards the Hill estimator is justified by its asymptotic properties, which are summarized in Embrechts *et al.* (1997, Theorem 6.4.6):

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

Remark 1.4 We have written the Hill estimator as an estimator of the tail index α, since this was the quantity of interest in most applications at the beginnings of the employment of EVT in financial studies. Anyway, $\hat{\xi}_{k,n}^{H} = \frac{1}{\hat{\alpha}_{k,n}^{H}}$ can be used as an estimator of the shape parameter ξ, so that we will talk indifferently of the Hill estimator as estimating either the tail index or the shape parameter.

■ *Pickands Estimator*

Studied by Pickands (1975), this estimator[19], unlike the Hill estimator, can be used to estimate the shape parameter of any of the three extreme value distributions and it reads:

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

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

[17] The Hill estimator can be derived in several different ways (see Embrechts *et al.*, 1997, or McNeil *et al.*, 2005).

[18] Simulation studies have reported evidence in favour of the Hill estimator. For instance, Kearns and Pagan (1997), comparing the estimators proposed by Hill, Pickands, and de Haan and Resnick, conclude that the Hill estimator is overall the best one.

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

- *Weak consistency* – if $k \to +\infty$ and $k/n \to 0$ for $n \to +\infty$, then $\hat{\xi}^P_{k,n} \xrightarrow{P} \xi$;
- *Strong consistency* – if $k/n \to 0$ and $k/\ln\ln n \to +\infty$ for $n \to +\infty$, then $\hat{\xi}^P_{k,n} \xrightarrow{a.s.} \xi$;
- *Asymptotic normality* – under additional hypotheses, $\sqrt{k}\left(\hat{\xi}^P_{k,n} - \xi\right) \xrightarrow{d} N\left(0, \nu\left(\xi\right)\right)$, with $\nu\left(\xi\right)$ depending on ξ in a highly non linear way[20].

A major flaw of the Pickands estimator is its high volatility, as underscored by Kearns and Pagan (1997) in their simulation study.

■ *Other Estimators*

For the sake of completeness, we recall the definition of two more estimators, though they are not as commonly employed in financial applications as the previous ones.

The Dekkers–Einmahl–de Haan (DEdH) estimator, or moment estimator, proposed by Dekkers *et al.* (1989), is an extension of the Hill estimator intended to enable it to deal with any type of extreme value distribution, not only the Fréchet family. The estimator is of the form

$$\hat{\xi}^{DEdH}_{k,n} = 1 + H^{(1)}_n + \frac{1}{2}\left[\frac{\left(H^{(1)}_n\right)^2}{H^{(2)}_n} - 1\right]^{-1},$$

where $H^{(1)}_n$ and $H^{(2)}_n$ represent empirical moments[21].

The estimator proposed by de Haan and Resnick (1980), on the contrary, has a very simple definition, given by

$$\hat{\xi}^{dHR}_{k,n} = \frac{1}{\ln k}\left(\ln X_{1,n} - \ln X_{k,n}\right).$$

■ *A Tail Index Estimator for Small Samples*

The nice asymptotic properties of the Hill estimator fail when dealing with small samples, since the estimator is biased in this case. A modification of the Hill estimator which results in an unbiased estimator for the tail index in the case of small samples[22] is provided by Huisman *et al.* (2001).

The procedure proposed consists of estimating the shape parameter $\hat{\xi}^H_k = \frac{1}{\hat{\alpha}^H_k}$ for $k = 1, \ldots, \kappa$, for a given κ, and then running the regression

$$\hat{\xi}^H_k = \beta_0 + \beta_1 k + \varepsilon_k, \qquad k = 1, \ldots, \kappa.$$

[20] The exact value of $\nu\left(\xi\right)$ is $\frac{\xi^2\left(2^{2\xi+1}+1\right)}{\left(2\left(2^{\xi}-1\right)\ln 2\right)^2}$.

[21] Precisely, $H^{(1)}_n = \frac{\sum_{j=1}^{k}\left(\ln X_{j,n} - \ln X_{k+1,n}\right)}{k}$ and $H^{(2)}_n = \frac{\sum_{j=1}^{k}\left(\ln X_{j,n} - \ln X_{k+1,n}\right)^2}{k}$.

[22] For instance, Pontines and Siregar (2008) employ this estimator with monthly data of exchange rates (for an overall period of less than twenty years).

The previous equation, for k approaching 0, provides an unbiased estimate of ξ equal to β_0. Thus, according to the authors, "applying this procedure solves the bias-variance trade-off by using the information from a whole range of conventional Hill estimates for different values of k to obtain an estimate for the tail index"[23]. Huisman *et al.* (2001) also show that the estimator resulting from this procedure is a weighted average of the different Hill estimators $\hat{\xi}_k^H$, with the weights themselves depending on k.

1.2 Parametric Approach I: GEV and the Block Maxima Method

Jenkinson (1955) proposed a unitary framework for the extreme value distribution of maxima of iid random variables. The *generalized extreme value* (GEV) *distribution*[24] (also called *Jenkinson–Von Mises representation* of extreme value distributions) with location parameter μ and scale parameter σ reads

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, & \xi \neq 0 \quad and \quad 1+\xi\frac{x-\mu}{\sigma} > 0 \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, & \xi = 0 \end{cases}$$

and nests the three limiting distributions distinguished by Gnedenko, since:

- for $\xi > 0$, delivers the Fréchet family;
- for $\xi = 0$, the Gumbel EV distribution;
- for $\xi < 0$, the Weibull family.

Therefore, we can restate the results of Fisher–Tippett and Gnedenko[25] as follows (McNeil *et al.* 2005, Theorem 7.3).

Theorem 1.5 *If $F \in MDA(H)$ for some non-degenerate distribution function H, then H must be a GEV distribution.*

In the following, we will use the shorthand $H_\xi(x) = H_{\xi,0,0}(x)$.

The first parametric approach to EVT we are going to consider, the so-called *block-maxima method*, consists of fitting the GEV distribution to a particular set of "maxima" chosen in a given sample of data.

Given a sample $X_1, \ldots, X_N$ of iid data drawn from an unknown distribution F:

[23] Huisman *et al.* (2001, p. 210).

[24] Notice that the GEV is continuous in ξ , since $\lim_{\xi\to 0} H_{\xi,\mu,\sigma}(x) = H_{0,\mu,\sigma}(x)$.

[25] As already mentioned, the norming constants for which one can conclude that $F \in MDA(H)$, for a fully parametric GEV distribution H, are not unique. Anyway, different choices for these constants yield the same family of extreme value distributions (namely, the same shape parameter ξ), the only difference being in the corresponding location and scale parameters (μ and σ , respectively).

(a) divide the sample in m non-overlapping subsamples of n observations each (*n-blocks*), for given integers m, n ($0 < m < n < N$), and denote by $M_{n,j}$ the maximum of the jth subsample;

(b) assuming that $F \in MDA(H_{\xi,\mu,\sigma})$ for some $\xi, \mu, \sigma \in \mathbb{R}$, with $\sigma > 0$, fit the GEV distribution to the sequence of block maxima $M_{n,1}, \ldots, M_{n,m}$, determining estimates $\hat{\xi}$, $\hat{\mu}$ and $\hat{\sigma}$ of the parameters ξ, μ and σ.

The estimation can be done by means of ML, though the maximization of the likelihood function is subject to the constraints

$$\sigma > 0 \qquad \text{and} \qquad 1 + \xi \frac{M_{n,j} - \mu}{\sigma} > 0, \quad \text{for} \quad j = 1, \ldots, m,$$

dependent on the parameters to be estimated. Consistency and asymptotic efficiency of the resulting non-regular MLEs can be proved for $\xi > -\frac{1}{2}$ (Smith, 1985), thus virtually covering all the cases of interest for financial applications.

Remark 1.6 Analogous results and methods can be employed to study the asymptotic distribution of minima, taking the relation $\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}$ into account.

A major pro of the block maxima method is its natural way of interpreting the problem one is trying to solve, especially when it is intrinsically structured in blocks of observations (e.g., daily observations of returns of a financial activity can be naturally split into quarterly or annual blocks).

On the contrary, the main flaw of this method is represented by its "waste of data", since many data in a single block could be regarded as extreme values, though they are not block maxima.

1.3 Parametric Approach II: GPD and Threshold Exceedances

An even more natural way to define extremes in a given sample is to set a "high" threshold u and to consider as extreme any observation exceeding u. This approach allows in principle for a more parsimonious use of data and hinges on theoretical foundations as solid as those of the block maxima method, as shown by the results of Balkema and de Haan (1974) and Pickands (1975).

In this case the distribution of interest is that of the exceedances over the threshold u, conditional on the fact that u is exceeded. We call it the *excess distribution over the threshold u*. By definition, the excess distribution over the threshold u corresponding to a random variable X with cumulative distribution function F is

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \qquad 0 \leq x < x_F - u,$$

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

The asymptotic distribution of F_u for most textbook probability distributions is the *generalized Pareto distribution* (GPD)

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi\frac{x}{\beta}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - e^{-\frac{x}{\beta}}, & \xi = 0, \end{cases} \quad \text{where} \quad \begin{cases} x \geq 0, & if \quad \xi \geq 0 \\ 0 \leq x \leq -\beta/\xi, & if \quad \xi < 0 \end{cases}$$

and $\beta > 0$. We will call ξ the *shape parameter*, as in the GEV distribution, and β the *scale parameter*.

A justification of the previous statement and connections with GEV distributions for maxima are provided by the following fundamental theorem.

Theorem 1.7 *For any $\xi \in \mathbb{R}$, there exists a positive measurable function $\beta(u)$ such that*

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

if and only if $F \in MDA(H_{\xi,\mu,\sigma})$, where $H_{\xi,\mu,\sigma}$ is a GEV distribution with the same shape parameter ξ as $G_{\xi,\beta(u)}$.

Thus, those cumulative distribution functions whose excess distribution over the threshold u converges (as u grows to the right end-point x_F) to a GPD with shape parameter ξ are exactly the same ones which lie in the maximum domain of attraction of a GEV distribution with the same shape parameter ξ.

The parametric approach to EVT provided by GPDs, analogously to the block-maxima method, consists of two steps:

- given a sample $X_1, \ldots, X_n$ of iid data, choose a threshold u and set $Y_j = \widetilde{X}_j - u$, where $j = 1, \ldots, N_u$ and $\widetilde{X}_1, \ldots, \widetilde{X}_{N_u}$ denote the data exceeding u;
- assuming that $F_u(y) = G_{\xi,\beta}(y)$ for $0 < y < x_F - u$ and some $\xi \in \mathbb{R}$, $\beta > 0$, fit the GPD to the sequence $Y_1, \ldots, Y_{N_u}$ of exceedances, determining estimates $\hat{\xi}$, $\hat{\beta}$ of the parameters ξ,β.

Again, we can employ a non-regular ML estimation procedure, performing a constrained maximization with constraints

$$\beta > 0 \quad \text{and} \quad 1 + \xi\frac{Y_j}{\beta} > 0, \quad \text{for} \quad j = 1, \ldots, N_u.$$

Remark 1.8 *(a)* Note that, in both parametric approaches to EVT, instead of using ML, the estimation can be performed via *probability-weighted moments* (PWM) [26], a method proposed by Hosking *et al.* (1985).

(b) Though the Hill estimator is acknowledged to be the most efficient estimator of ξ, McNeil and Frey (2000), on the basis of simulation studies, argue that it does not provide the most efficient estimates of quantiles (and consequently of VaR) and conclude that, especially for very high quantiles, GPD provides better estimates.

1.4 Some Generalizations

■ *The Point Process Approach*

Both parametric approaches presented in the previous sections can be subsumed in a unified framework, where the occurrence of exceedances over thresholds is modelled by means of point processes.

Recall that, under some technical assumptions, the counter

$$N(A) = \sum_{j=1}^{n} I_{\{Z_j \in A\}}$$

of the number of elements Z_j, $j = 1, \ldots, n$, belonging to a region A of some state space (e.g. $\mathbb{R}^d$, $d \geq 1$) defines a point process.

The reference point process model for iid data is the *peaks-over-threshold (POT) model*[27], essentially based on the assumption that, given regularly spaced random data $X_1, \ldots, X_n$ and a threshold u, the point process defined on the state space $\Xi =]0, 1] \times]u, +\infty[$ by the counter

$$N(A) = \sum_{j=1}^{n} I_{\{(j/n, X_j) \in A\}}, \qquad A \subseteq \Xi,$$

is a Poisson process with intensity function

$$\lambda(t, x) = \frac{1}{\sigma}\left(1 + \xi\frac{x - \mu}{\sigma}\right)^{-1/\xi - 1}.$$

In this way, the threshold exceedances follow a point process that jointly considers both the magnitude of each exceedance, by means of the quantity X_j, and the "normalized time" j/n at which it takes place, where the magnitude and time windows are given in the second and in the first component of the set A, respectively.

The hypothesis that the point process be a Poisson process entails that:

[26] Most financial applications use MLEs. Anyway, Tolikas *et al.* (2007), for instance, use the PWM method, on the basis that it is considered to be more efficient than the ML estimation procedure.

[27] We follow here the intuitive presentation given by McNeil *et al.* (2005), where the details are fully covered.

(a) for any $A = [t_1, t_2] \times [a, b]$, the probability of having exactly k exceedances falling in the strip $[a, b]$ during the time interval $[t_1, t_2]$ is Poisson distributed with parameter $\Lambda(A) = \int_A \lambda(t, x)\, dtdx$;

(b) the numbers of exceedances occurring in disjoint time-magnitude windows are assumed to be independent;

Finally, the assumption on the particular expression for the intensity function λ allows to recover a GPD as the limiting distribution for the excess distribution over a threshold u.

As usual, the estimation of the model can be pursued via ML[28]. Moreover, a reparameterization of the model can be accomplished in order to split the log-likelihood function in the sum of two terms involving different sets of parameters and allow drawing separate inferences.

One main advantage of this approach is that, unlike in the GPD approach, when modelling threshold exceedances the parameters of the POT model do not show theoretical dependence on the chosen threshold u.

■ ***The Box-Cox-GEV Distribution***

Bali (2003b) introduces a new generalization of extreme value distributions, obtained by applying to the GEV of Jenkinson (1955) the transformation proposed by Box and Cox (1964). The resulting Box-Cox-GEV has cumulative distribution function

$$B_{\xi,\mu,\sigma,\lambda}(x) = \frac{\exp\left\{-\lambda\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\} - 1}{\lambda} + 1.$$

Notice that the Box-Cox-GEV nests the GEV distribution and the GPD, since:

- for $\lambda = 1$, we immediately recover the GEV distribution, being $B_{\xi,\mu,\sigma,1}(x) = H_{\xi,\mu,\sigma}(x)$;
- for $\lambda \to 0$, we obtain[29] $B_{\xi,\mu,\sigma,\lambda}(x) \to G_{\xi,\mu,\sigma}(x) = G_{\xi,\sigma}(x-\mu)$, i.e. the GPD ($x - \mu$ are exceedances over the threshold μ).

Empirical studies on financial data in Bali (2003b, 2007) conclude that the Box-Cox-GEV with some $\lambda \in]0, 1[$ fits the distribution of extreme values better than both the GEV distribution and the GPD.

2 Setting the Cut-off

Both parametric and non-parametric extreme value methods share a common drawback: the estimates they provide are sensitive to the choice of the cut-off parameter, that is to say the

[28] See Coles (2001) for the details.

[29] Using the limit $\lim_{t\to 0} \frac{a^t-1}{t} = \ln a$.

number m of blocks in the block-maxima method, the threshold u in the threshold-exceedance method, or the index k of the order statistic marking the frontier between the centre and the tail of the distribution in non-parametric approaches.

The choice of the cut-off parameter is thus a central issue to any application of EVT, since in principle different estimates of the shape parameter can lead to significantly different estimates of the VaR, for example, or to different conclusions concerning the distributional assumption that better fits the data under consideration[30].

The determination of a "good" cut-off is not easy, since it involves a trade-off between bias and variance. Reasoning for instance about the cut-off for the Hill estimator, note that:

- if k is set too high, many data are included in the tail of the distribution, even if not extreme, yielding biased parameter estimates;
- on the other hand, low values of k give fewer observations of extreme data, resulting in inefficient estimates of the parameters, with huge standard errors.

Several methods have been adopted in the literature to cope with this issue. We present here the main alternatives, reasoning on the case of the Hill estimator, just to fix ideas.

■ *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, usually based on one of the methods that we are presenting in the following paragraphs.

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*

When using the Hill estimator to estimate the tail index α of a given distribution, a very common way to determine a good choice for the cut-off k is given by the so-called *Hill plots*. They are simply the graphical representation, in a coordinate system, of the set

$$\left\{\left(k, \hat{\alpha}^H_{k,n}\right) : k = 2, \ldots, n\right\},$$

i.e., they plot a graph of the estimates of the tail index as a function of the cut-off k. Regions of the plot that are approximately close to be horizontal lines indicate values of k for which the estimate $\hat{\alpha}^H_{k,n}$ is essentially stable with respect to the choice of the cut-off.

Analogously, *Pickands plots* can be employed when dealing with the Pickands estimator, where the set $\left\{\left(k, \hat{\alpha}^P_{k,n}\right) : k = 2, \ldots, n\right\}$ is now to be plotted[31].

[30] Lux (2000), for instance, argues that contrasting conclusions in the literature concerning the behaviour of German stock returns are due to different a priori choices of the cut-off.

[31] On the same lines, one can also construct Jenkinson plots, or Pareto plots.

Another useful instrument is provided by *mean excess plots*. Defining the sample mean excess function as

$$e_n(u) = \frac{\sum_{i=1}^{n} (X_i - u) \cdot I_{\{X_i > u\}}}{\sum_{i=1}^{n} I_{\{X_i > u\}}},$$

one can plot the set $\{(X_{k,n}, e(X_{k,n})) : k = 2, \ldots, n\}$ in a coordinate system. 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 $\xi < 0$).

These graphical tools can be very helpful in the choice of k, 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. In practice, they are often used in combination with other methods, in order to obtain a first delimitation of the region to sift for determining the optimal k. In any case, exploratory data analysis (usually based on QQ-plots or the mean excess function) is always in case before running EVT estimates.

■ ***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 characterizing the above mentioned trade-off. Indeed, the MSE of an estimator $\hat{\vartheta}$ is

$$MSE\left(\hat{\vartheta}\right) = E\left[\left(\hat{\vartheta} - \vartheta\right)^2\right] = \left[E\left(\hat{\vartheta} - \vartheta\right)\right]^2 + Var\left(\vartheta\right),$$

so that minimizing MSE simultaneously implies the minimization of both bias and variance.

In practice, the selection of the optimal cut-off is based on the following algorithm:

(a) simulate data via Monte Carlo from a known distribution F in the domain of attraction of a Fréchet distribution with tail index α;

(b) compute the Hill estimator of α for different choices of k;

(c) choose the value of k for which the MSE is minimal.

The simulation from a known distribution is required in order to know the exact value of the tail index to use in the evaluation of the MSE. The usual choice for F is that of a Student t, which allows for different degrees of tail fatness and is standard when modelling financial data.

This method has two major flaws. The first one is the distributional assumption in step (a), since it looks like somehow denying one of the most appealing features of EVT, that is the possibility of modelling the tails without knowing the exact underlying distribution (in fact, one of the main applications of EVT is precisely as a tool to determine the correct distributional choice for the data; in this case the Monte Carlo method that we are considering would be faced with a circularity issue). In any case, there is no guarantee that the optimal choice for k when dealing with simulated data coincides with the optimal choice induced by real data, as we don't know the distribution of the latter.

The second flaw is related to the very election of MSE minimization as an optimality criterion (and therefore affects data-driven methods as well). The reason is two-fold. On the one hand, Manganelli and Engle (2004) point out that the evaluation of the optimal cut-off based on this criterion yields a biased estimator and no theoretical inquiry has been done to quantify the resulting bias when, for instance, estimating VaR (the EVT approach to VaR calculation involves indeed a non-linear transformation of the estimated shape parameter). On the other hand, if the iid 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 iid, an MSE based optimality criterion is questionable.

■ ***Data Driven Algorithms***

To overcome the flaw of the requirement 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 at hand.

A review and a comparative analysis of the main ones is provided by Lux (2000), to whom we refer the reader. His analysis focuses on five methods, among which we bring to the attention the method proposed by Hall (1990) and Daníelsson and de Vries (1997a), which employs a bootstrap approach to estimating MSE, and the method by Beirlant *et al.* (1996), which is based on an iterative regression approach.

3 Dependence in the Data

In the previous sections we assumed that our data were iid. However, since we are interested in applying EVT to financial data, in many cases this assumption turns out to be highly unrealistic.

Two main ways of taking the dependence structure of the data into account have been pursued in the literature. The first one relies on a pertinent extension of the theoretical framework of EVT to strictly stationary time series that satisfy a certain hypothesis, including ARMA and ARCH/GARCH processes. The second approach is a two-step procedure that first models the correlation structure of the observations by fitting, e.g., a GARCH model to the data and then performs the estimation of the GPD distribution (or the GEV) on the residuals of the first regression, as they can be considered as roughly iid.

Finally, we will mention the approach proposed by Chavez-Demoulin *et al.* (2005), which uses point processes to model the in-cluster behaviour of extremes.

3.1 EVT for Strictly Stationary Time Series

The basic idea on which the extension of EVT to strictly stationary time series[32] relies is that of studying the associated iid process with the usual tools of standard EVT and check if a link can be established between the asymptotic behaviour of the extremes of the time series and those of the associated random process.

Consider then a strictly stationary time series (X_i), with stationary distribution F, and a strict white noise process $\left(\widetilde{X}_i\right)$ having the same cumulative distribution function. Denoting by M_n and $\widetilde{M}_n$ the maxima of the two blocks $(X_1, \ldots, X_n)$ and $\left(\widetilde{X}_1, \ldots, \widetilde{X}_n\right)$, respectively, we are interested in studying those strictly stationary time series for which there exists $\vartheta \in]0, 1]$ such that, for some non-degenerate distribution H,

$$\lim_{n \to +\infty} P\left(\frac{\widetilde{M}_n - d_n}{c_n} \leq x\right) = H(x) \quad \Leftrightarrow \quad \lim_{n \to +\infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = H^{\vartheta}(x), \tag{2}$$

where d_n and $c_n > 0$ are norming constants, as usual. The real number ϑ is called the *extremal index* of the given process[33].

Since, as it can be easily verified, for any GEV distribution H_ξ, its power H_ξ^ϑ is still a GEV distribution with the same shape parameter ξ, equivalence (2) means that the asymptotic distribution of the block maxima of the iid process is a GEV distribution with shape parameter ξ if and only if the block maxima of the time series are asymptotically distributed according to a GEV with the same shape parameter ξ (though, in general, with different location and scale parameters). Therefore, for a sufficiently high threshold u, the probability $P(M_n \leq u)$ can be approximated by $\left[P\left(\widetilde{M}_n \leq u\right)\right]^\vartheta$, which is known to be equal to $F^{n\vartheta}(u)$, since the process is iid. Thus, in this case the probability $P(M_n \leq u)$ will not be equal to $F^n(u)$, as in the iid case, but will be roughly approximated by $F^{n\vartheta}(u)$, where $n\vartheta \leq n$, given that $\vartheta \in]0, 1]$, as a consequence of our definition of the extremal index. When $\vartheta < 1$, extreme values in the time series tend to cluster and the quantity $n\vartheta$ stands for the number of independent clusters, while ϑ^{-1} can be interpreted as the mean size of the clusters.

This theoretical development of EVT is relevant to financial applications, since it has been proved[34] that ARMA, ARCH and GARCH processes, which are frequently employed in modelling financial time series, satisfy (2). Thus we can extend to these processes all the useful tools provided by EVT. Anyway, when fitting the model to the data, we have one more parameter to estimate, the extremal index ϑ. Moreover, since in general $\vartheta < 1$ (with the exception of ARMA processes with Gaussian strict white noise innovations, for which $\vartheta = 1$), one has to take into account that convergence will be slower than expected in the iid case and a greater

[32] Recall that a time series (X_t) is strictly stationary if, for all $n \in N \setminus \{0\}$ and $t, t_1, \ldots, t_n \in \mathbb{R}$, the random vectors $(X_{t_1}, \ldots, X_{t_n})$ and $(X_{t_1+t}, \ldots, X_{t_n+t})$ have the same joint distribution, i.e. if the distribution of $(X_{t_1}, \ldots, X_{t_n})$ is invariant under time shift.

[33] See Leadbetter *et al.* (1983) and Embrechts *et al.* (1997) for a mathematically sound definition and study of the extremal index. A recent work employing this approach when applying EVT to financial issues is Ho (2008).

[34] De Haan *et al.* (1989) and Mikosch and Stărică (2000) for ARCH(1) and GARCH(1,1) processes, respectively.

number of data will be needed, since in any sample of n observations only $n\vartheta$ clusters of them will actually matter to the purpose of estimating the behaviour of extremes. Therefore, one can still use EVT methods from the iid framework, but the estimators obtained may be less accurate and neglecting this fact could lead to inadequate resolutions in order to cope with the risk of occurrence of extreme events.

In practice, a common way to proceed is to decluster the data (there exist several methods, but the issue of a precise definition of the boundaries of a cluster makes things quite fuzzy) and then apply standard EVT methods.

3.2 Data Filtering and the Two-step Procedure

Another approach to manage dependence in the time series that has been successfully employed in the literature on financial applications of EVT is the one suggested by Diebold *et al.* (1998) and implemented by McNeil and Frey (2000) [35].

The main intuition on which the method hinges is that, though financial data are not iid, the standardized residuals obtained as a by-product of modelling the dependence structure of the time series are roughly iid. To the extent to which this is true, one could hope to avoid the problem of dependence in the data by applying EVT to the residuals instead of directly applying it to the raw data.

This idea of a filtering, or a pre-whitening, of the data yields then a two-step procedure that acts in the following way. Given a strictly stationary time series (X_t), whose dynamics is driven by $X_t = \mu_t + \sigma_t Z_t$, with Z_t a strict white noise process of unknown distribution:

(a) fit an AR(1)-GARCH(1,1) process to model the conditional mean and volatility; evaluate forecasts for μ_{t+1} and σ_{t+1}, based on the estimated model, and compute the standardized residuals;

(b) model the tails of the unknown distribution of the standardized residuals by means of EVT.

Notice that estimation of the GARCH model at step (a) is performed by means of pseudo-maximum likelihood (PML). This means that a GARCH(1,1) model with normal innovations is estimated, though we do not really commit our belief to this distributional assumption. Anyway, the use of PML is a sensible choice, since it yields a consistent and asymptotically normal estimator.

The crucial point for the two-step approach to be a viable way to the solution of the issue of dependence is the property that the residuals of the GARCH model fitted at the first step be iid, in order to apply EVT at the second step.

However, a recent paper by Jalal and Rockinger (2008) shows that, even when the GARCH model is misspecified, GARCH filtering followed by EVT modelling of the standardized residuals delivers good results, thus proving to be quite a robust approach (precisely, they check the

[35] The main reference regarding the two-step approach is McNeil and Frey (2000). An anticipation of this approach is given in an example by Embrechts *et al.* (1997, Fig. 5.5.4., p. 270). Further applications can be found in Bali (2007), Bali and Neftci (2003), Byström (2004, 2005), Cotter (2007a, 2007b), Krehbiel and Adkins (2005), Kuester *et al.* (2006), Lestano and Jacobs (2007). See Dias and Embrechts (2003) for an application of the method to the multivariate setting.

consequences of applying it to either GARCH data with non-normal innovations, or data simulated by non-GARCH models). Anyway, a somehow contrasting result is found by Kuester *et al.* (2006), who observe that better results in VaR estimation can be obtained with non-normal innovations, thus suggesting that "distributionally nonparametric models do indeed depend on the distribution assumed in the filtering stage"[36].

Possible variations and generalizations of the two-step procedure concern the following features.

- The choice of the filtering model. McNeil and Frey (2000) justify the choice of AR(1)-GARCH(1,1) on the grounds of its parsimoniousness and effectiveness. Dias and Embrechts (2003) use an ARMA-GARCH model, while, as just mentioned, Kuester *et al.* (2006) criticize the AR(1)-GARCH(1,1) model as incapable of correctly specifying financial data and improve on this point by considering skewed t innovations. Chan and Gray (2006) use an AR-EGARCH model with t innovations to take into account weekly seasonality in conditional volatility and leverage effects, while Bali (2007) uses a GJR-GARCH model with skewed t innovations.
- Different EVT techniques can be employed at step (b). While McNeil and Frey use GPD and threshold exceedances, Cotter (2007b), for instance, uses the Hill estimator.
- Extension to a multivariate framework. The above cited paper Dias and Embrechts (2003) applies the two-step procedure to model the marginal distributions of a multivariate time series.

3.3 A Point Process Approach to In-cluster Behaviour Modelling

The two methods for dealing with the dependence structure of time series we described above can work fairly well in practice, but they suffer from some drawbacks as well. The two-step procedure crucially depends on the results of the fitting of a GARCH model in the first step and could deliver inaccurate estimates if the model were misspecified, while the extremal index approach followed by declustering dismisses the task of modelling the short-run behaviour of extremes and is faced by the difficulties related to practical identification and treatment of clusters.

This task, on the contrary, is explicitly undertaken by a point process approach introduced by Chavez-Demoulin *et al.* (2005). The method is a modification of the peaks-over-threshold model we described in section 1.4, in order to relax the iid hypothesis and allow to deal with stationary dependent time series. In this case, the occurrence of threshold exceedances is not modelled by a homogeneous Poisson process; rather, a marked point process with a self-exciting structure for the times of occurrence of excesses is employed, i.e. a point process in which the amplitude of each event $w_{\mathrm{j}} = X_j - u$ (*mark*) is considered and the intensity function depends on the time elapsed since the last exceedances took place (*self-exciting process*). Typically, this dependence is modelled by means of a function of both time and w_{j}, that is monotonically non-decreasing in time, so that the more recent an exceedance is, the more it contributes to the current intensity.

[36] Kuester *et al.* (2006, p. 84).

The basic assumption under which the model is derived is that times and marks are independent, conditional on the information available at the time of the previous exceedance[37].

4 The Multivariate Setting

One major issue of EVT concerns its essentially univariate nature. The concept of a maximum, as well as that of a threshold excess, is based on the existence of an order relation. When moving to a multivariate setting, order relations generally provide partial orderings only and many of them can be legitimate candidates, thus somehow yielding even more arbitrariness in the construction of a multivariate extreme value theory (MEVT) than can be found in the univariate case. However, the most difficult problem is related to the dependence structure of extremes in a multivariate series and the choice of a proper copula to model it. As in the previous section we were interested in modelling the dependence structure of a univariate series along its temporal development, analogously we will now focus on dependence among extreme realizations of different components of a multivariate process.

Despite its pitfalls, the importance of a multivariate extension of EVT techniques is considerable, since in many applications we are faced with time series that are multivariate in nature and the dependence structure of their extreme values has to be appropriately modelled and taken into account. For instance, multivariate extreme value theory can be successfully employed when performing portfolio selection or when testing the existence of contagion across different markets[38].

In the following we will only reason on bivariate distributions. The extension to general multivariate distributions can be theoretically conceivable, but it is often demanding from a computational viewpoint (see section 8 below) and it is not always completely straightforward, so that, in practice, most applications restrict to d-variate EVT with $d = 2$ or $d = 3$.

We first introduce the notions of asymptotic dependence and independence (4.1), then consider the multivariate generalization of Fisher-Tippett and Gnedenko theorems and the analogue of the block maxima method (4.2); finally, we present the extension of the threshold exceedance approach to the multivariate setting (4.3).

4.1 Tail Dependence

When considering the extremal behaviour of a multivariate distribution, it is important to study the strength of dependence between the tails of different components of the distribution itself, in order to appropriately model the dependence structure. A useful measure of extremal dependence can be defined as follows.

[37] We don't enter here the mathematical details of the theory, for which we refer the reader to Chavez-Demoulin *et al.* (2005), or to the presentation of the method provided by McNeil *et al.* (2005).

[38] Ané (2006) makes an interesting and non-standard use of multivariate EVT, studying univariate Asian-Pacific stock indices via a two-component extreme value distribution: one component accounts for "ordinary" extremes, while the other one stands for the outliers of the distribution, thus enhancing the accuracy of EVT in modelling really extreme events.

Definition 4.1 *Let X_1, X_2 be random variables with cumulative distribution functions F_1, F_2, respectively. The* coefficient of upper tail dependence *of X_1 and X_2 is*

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

provided this limit exists and is finite.

Thus the coefficient of upper tail dependence[39] measures the conditional probability that X_2 exceed the qth quantile, given that X_1 does, as q tends to 1. Loosely speaking, it measures to what extent the occurrence of extreme realizations of the first random variable affects the probability of observing extreme realizations of the second one.

If $\lambda_u = 0$, X_1 and X_2 are said to be *asymptotically independent* (in the upper tail); otherwise, i.e. if $\lambda_u \in]0, 1]$, X_1 and X_2 are *asymptotically dependent*.

The extremal dependence structure is typically different from the dependence we find at the centre of the distribution, given that, for instance, asymptotical independence can be achieved when the components of the distribution are not independent. For example, a bivariate normal variable with correlation $\rho \neq 1$ is asymptotically independent, though it is not linearly independent, unless $\rho = 0$.

Taking extremal dependence into account is very important when using multivariate EVT. For instance, Bradley and Taqqu (2004b) find that applying multivariate extreme value models for asymptotically dependent variables to asymptotically independent ones can overstate portfolio risk, thus leading to an excessively conservative position (see 8 below).

4.2 Multivariate Block Maxima

In the multivariate setting, first of all, we are interested in defining multivariate maxima and finding suitable normalizations under which they converge to some non-degenerate distribution, exactly as in the univariate case (1.2).

Thus, given $X_1, \ldots, X_n$, iid random vectors in $\mathbb{R}^k$ with joint cumulative distribution function F, denote by $M_{n,j} = \max\{X_{1,j}, \ldots, X_{n,j}\}$, $j = 1, \ldots, k$, the maximum of the jth component of $X_1, \ldots, X_n$. The multivariate block maxima method looks for a limiting distribution of standardized componentwise block maxima $M_n = (M_{n,1}, \ldots, M_{n,k})^T$. In particular, analogously to the univariate case, if there exist random vectors c_n, d_n in $\mathbb{R}^k$, where c_n has positive components, and a non-degenerate joint cumulative distribution function H such that

$$\begin{aligned} & \lim_{n \to +\infty} P\left(\tfrac{M_{n,1} - d_{n,1}}{c_{n,1}} \leq x_1, \ldots, \tfrac{M_{n,k} - d_{n,k}}{c_{n,k}} \leq x_k\right) \\ = & \lim_{n \to +\infty} F^n\left(c_{n,1}x_1 + d_{n,1}, \ldots, c_{n,k}x_k + d_{n,k}\right) \\ = & H\left(x_1, \ldots, x_k\right), \end{aligned}$$

[39] A version of this measure that is invariant with respect to the marginal distributions of the random variables considered can be obtained standardizing the marginal distributions themselves, e.g. replacing X_1, X_2 by $-1/\ln F_{X_1}(X_1)$ and $-1/\ln F_{X_2}(X_2)$, respectively. Moreover, a complementary measure was introduced by Ledford and Tawn (1996) to evaluate extremal dependence in the case of variables that are asymptotically independent. See e.g. Poon *et al.* (2004) for a presentation of both measures and an application to dependence between financial markets.

then we say that F is in the *maximum domain of attraction* of H, in symbols $F \in MDA(H)$, and call H a *multivariate extreme value* (MEV) distribution and its copula C_0 an *extreme value copula*.

This construction is closely similar to that of the univariate block maxima method, since it originates as an extension of that approach to the multivariate setting, but the relation between multivariate and univariate EVT can be expressed in an even more precise way. To do this we need one more definition, namely that of copula domain of attraction.

Given two copula functions C,C_0, we say that C is in the *copula domain of attraction* of C_0, in symbols $C \in CDA(C_0)$, if

$$\lim_{t \to +\infty} C^t \left(u_1^{1/t}, \ldots, u_k^{1/t} \right) = C_0 (u_1, \ldots, u_k),$$

for all $u \in [0,1]^k$.

The relation between the limiting joint distribution of multivariate block maxima and the limiting distributions of its margins is provided by the following theorem[40].

Theorem 4.2 *Let $F(x) = C(F_1(x_1), \ldots, F_k(x_k))$, where the marginal distribution functions $F_1, \ldots, F_k$ are continuous, and let $H(x) = C_0(H_1(x_1), \ldots, H_k(x_k))$ be a multivariate extreme value distribution. Then $F \in MDA(H)$ if and only if $C \in CDA(C_0)$ and $F_j \in MDA(H_j)$, for all $j = 1, \ldots, k$.*

Therefore, in particular, the marginal distributions of F determine the margins of H, but leave the extreme value copula unaffected; the latter, indeed, is only determined by the copula of F.

When coming to estimation, the procedure is again similar to the univariate case, based on two main steps. Given a sample $X_1, \ldots, X_N$ of iid vectors drawn from an unknown distribution F:

(a) divide the sample in m blocks of n observations each, where $0 < m < n < N$, and denote by M_n^i the componentwise maximum of the ith block;

(b) assuming that $F \in MDA(H)$ for some multivariate extreme value distribution

$$H_{\xi,\mu,\sigma,\eta}(x) = C_\eta (H_{\xi_1,\mu_1,\sigma_1}(x_1), \ldots, H_{\xi_k,\mu_k,\sigma_k}(x_k)),$$

fit $H_{\xi,\mu,\sigma,\eta}$ to the sequence of block maxima $M_n^1, \ldots, M_n^m$, determining estimates $\hat{\xi}$, $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\eta}$ of the parameters.

As usual, the estimation can be done via ML. One can choose whether to perform a one-step estimation or a two-step estimation, first fitting the marginal distributions and then estimating the copula model.

[40] See Galambos (1987) and McNeil *et al.* (2005).

The same critiques that could be made to the univariate block maxima method hold in the multivariate setting as well. Moreover, a new issue arises, since we do not know the actual copula C of the distribution F and we cannot know exactly which extreme value copula contains C in its copula domain of attraction. Therefore, in practice one has to work with some suitable parametric copula C_η solving a trade-off between two requirements: on the one hand, that C_η be a sensible choice for the data at hand (for instance, that it be a parametric family allowing for asymptotic independence, if the data are asymptotically independent) and, on the other hand, that C_η be easy to implement. A standard choice in the financial literature is the logistic model[41], that is quite easy to implement and for which there is some empirical evidence that, even though it may be a theoretically incorrect model for the data at hand, it can yield a sufficiently accurate description of the observed behaviour of the data[42] (in the case of tail dependence).

4.3 Multivariate Threshold Exceedances

In this section we present an approach to multivariate EVT that extends the univariate threshold exceedance method[43] (see 1.3 above).

The main idea is to fit a multivariate distribution $\widetilde{F}(x) = C_\theta\left(\widetilde{F}_1(x_1), \ldots, \widetilde{F}_k(x_k)\right)$ to the data, where the margins $\widetilde{F}_j$, $j = 1, \ldots, k$, are GPDs of the form

$$\widetilde{F}_j(x_j) = 1 - \lambda_j\left(1 + \xi_j \frac{x_j - u_j}{\beta_j}\right)^{-1/\xi_j},$$

C_θ is a suitable parameterized family of extreme value copulas and the maximum likelihood estimates are calculated with censored data. This means that, when we set a high vector threshold $u = (u_1, \ldots, u_k)^T$, given a vector $X_i = (X_{i,1}, \ldots, X_{i,k})^T$, in general it will hold neither $X_i \leq u$ nor $X_i \geq u$, since some components of X_i will be greater than the corresponding components of u, while some others will be smaller. For the sake of simplicity, let exactly the first h $(1 \leq h \leq k)$ components of X_i be those for which $X_{i,j} > u_j$ holds. The information carried by the other components when studying threshold exceedances reduces to the fact that they are not above the threshold, irrespectively of their actual value. Censoring of the data consists of taking this fact into account. Precisely, it means that the contribution L_i of the vector X_i to the likelihood function is only determined by the first h components, namely

$$L_i = \left.\frac{\partial^h \widetilde{F}(x_1, \ldots, x_k)}{\partial x_1 \ldots \partial x_h}\right|_{(x_1,\ldots,x_k)=\left(X_{i,1},\ldots,X_{i,h},u_{h+1},\ldots,u_k\right)} \tag{3}$$

[41] The logistic (or Gumbel) copula model is $C_{\gamma,\alpha,\delta}(u_1, u_2) = u_1^{1-\alpha} u_2^{1-\delta} \exp\left\{-\left[(-\alpha \ln u_1)^\gamma + (-\delta \ln u_2)^\gamma\right]^{1/\gamma}\right\}$.

[42] For a critical review of the main advantages and drawbacks of copula-based extreme behaviour modelling, occasioned by the general disarray caused by the credit crisis, see Embrechts (2009). In particular, the intrinsically static nature of the copula is highlighted as a prominent flaw, as "the subprime crisis made clear the shortcomings of copula-based modeling with respect to sudden widening of credit spreads" (Embrechts, 2009, p. 36).

[43] The approach is substantially based on Ledford and Tawn (1996). We follow the presentation given by McNeil *et al.* (2005). See also Balkema and Embrechts (2007) for a geometric approach to multivariate EVT in the spirit of the Peaks over Thresholds (POT) method.

In practice, for a bivariate distribution[44], this amounts to partitioning $\mathbb{R}^2$ in four regions

$$A_{0,0} =]-\infty, u_1[\times]-\infty, u_2[\,, \qquad A_{1,0} = [u_1, +\infty[\times]-\infty, u_2[\,,$$

$$A_{0,1} =]-\infty, u_1[\times [u_2, +\infty[\,, \qquad A_{1,1} = [u_1, +\infty[\times [u_2, +\infty[$$

and computing the value of L_i as

$$L_i = \begin{cases} \widetilde{F}\left(u_1, u_2\right), & for \quad X_i \in A_{0,0} \\ \left.\frac{\partial \widetilde{F}(x_1,x_2)}{\partial x_1}\right|_{(X_{i,1},u_2)}, & for \quad X_i \in A_{1,0} \\ \left.\frac{\partial \widetilde{F}(x_1,x_2)}{\partial x_2}\right|_{(u_1,X_{i,2})}, & for \quad X_i \in A_{0,1} \\ \left.\frac{\partial^2 \widetilde{F}(x_1,x_2)}{\partial x_1 \partial x_2}\right|_{(X_{i,1},X_{i,2})}, & for \quad X_i \in A_{1,1}, \end{cases}$$

according to (3).

5 The Choice of the Data-Set

In this section we briefly summarize some problems concerning the choice and the preparation of the data-set to which apply EVT methods. Though this choice can be ascribed to the analysis of applications, we include it here since it represents a relevant methodological issue when using EVT.

■ *Raw vs Log Data*

Instead of considering either raw data or simple returns, it is customary in finance to consider either log prices or log returns on those prices, due to considerable advantages, such as nicer statistical properties of the time series of log returns, compared to those of the time series of raw data, and the additivity property for multiperiod returns induced by the logarithmic transformation (see e.g. Tsay, 2005)[45].

Statistical properties of the data and of their transformations are especially crucial to a methodology, like that supplied by EVT, which is based on asymptotic properties but is necessarily faced with scarcity of data (given that it deals with extreme values, which are rare events, by definition).

[44] In addition to the above mentioned references, see also Longin and Solnik (2001, Appendix 2) and Bradley and Taqqu (2004b) for a detailed derivation of the maximum likelihood function in the bivariate case.

[45] For instance, Longin (1999), approaching the problem of setting optimal margin levels for silver futures contracts on COMEX via EVT, instead of "dollars per contract" (the unit measure adopted by margin committees) uses percentage log returns, arguing that this definition is valuable, since: "It provides the econometrician with a stationary time series; it is independent of the unit of measurement; and it is stable under time-aggregation." (Longin, 1999, p. 142).

As the dichotomy between raw and log data is concerned, no general recipe is available, but the decision has to be taken according to the specific data-set at hand. We only mention two opposite situations that one could encounter. On the one hand, there are data which do not clearly exhibit an asymptotic GPD of threshold exceedances, like the daily changes in the LIBOR studied by Krehbiel and Adkins (2008). As shown by the authors, considering log changes makes the data suitable to the use of EVT methods. On the other hand, when the data themselves already display huge variations and many extreme values of considerable magnitude, like the electricity prices on Nord Pool analyzed by Byström (2005), it is advisable not to work with log prices[46].

■ ***Frequency of the Data***

The literature on high frequency financial data has grown considerably in the recent past. In principle, the issue of the frequency of data could matter significantly to EVT, given the above mentioned antinomy between the asymptotic nature of that theory and the actual scantness of data. Anyway, in financial applications of EVT there is a great variety in this respect. Most papers use daily data (especially when stock returns are concerned), but relevant work has been done also with higher frequency data. For instance, Lux (2001) uses minute-to-minute changes during trading hours at the Frankfurt Stock Exchange over a period of seven years (thus collecting almost 300,000 observations), Werner and Upper (2004) study tick data (approximately covering five years) of German Bund futures for a total amount of 13.4 million trades, while Hauksson *et al.* (2001) consider 10 minute to biweekly returns of the foreign exchange market and conclude that high frequency data can significantly improve extreme value estimates.

Indeed, the choice of high frequency data provides a huge bulk of observations, thus enhancing the possibility of performing accurate estimates of the parameters governing the tail behaviour of the distribution, but put some problem forth as well. In particular, high frequency time series usually exhibit seasonality as a considerable component. Thus, some preparation of the data is often required.

In any case, the choice of the frequency should be coherent with the kind of data at hand and with the purpose of the analysis. For instance, Ho (2008), applying EVT in order to date currency crises, uses monthly data, since a component entering the index of exchange market pressure (EMP) is only available at a monthly frequency, while Longin (2000), performing EVT based estimates of VaR, relates the choice of the frequency to the liquid or illiquid nature of the position one holds[47].

[46] Obviously the question of the choice of suitable data does not restrict to that of raw versus log data. For example, studying exchange rates, Koedijk *et al.* (1990) suggest that, in order to achieve stationarity, it is a more sensible option to deal with exchange rate returns rather than exchange rates themselves. Hsieh (2001), on the contrary, pursues stationarity by using log exchange rates. In some cases, another possibility to obtain better data on a particular financial instrument is to consider futures contracts on that instrument, as proposed by Werner and Upper (2004) studying the bond market.

[47] "The choice of the frequency should be related to the degree of liquidity and risk of the position. For a liquid position,high frequency returns such as daily returns can be selected as the assets can be sold rapidly in good market conditions. The frequency should be quite high as extreme price changes in financial markets tend to occur during very short time-periods as shown by Kindleberger (1978). Moreover, low frequency returns may not be relevant for a liquid position as the risk profile could change rapidly. For an illiquid position, low frequency returns such as weekly or monthly returns could be a better choice since the time to liquidate the assets in good market conditions may be longer. However, the choice of a low frequency implies a limited number of (extreme) observations, which could impact adversely upon the analysis as extreme value theory is asymptotic by nature." (Longin, 2000, pp. 1104-1105).

Finally, notice that there are at least three more ways in which one can enrich the data-set in order to make EVT a viable approach.

- Expanding the time window as much as possible. For example, Bali (2007) uses daily index levels of the Dow Jones Industrial Average (DJIA) from May 26, 1896 through December 29, 2000 (more than 28 thousand observations). This is the most obvious solution, but not always a sensible one, since for forecasting purposes very old data could deteriorate estimates rather than improve them.

- Jointly modelling both lower and upper extreme values. Though dismissing a significant feature of EVT (namely its capability of separately modelling the two tails of the distribution), this can be a sensible approach if the empirical distribution is approximately symmetric.

- Pooling different data series in a single one. This blend is done for example by Byström (2007), who, studying the iTraxx Europe CDS index market and having only daily observations covering a period of one year and a half, merges all subindices into a single data series.

The previous remarks suggest that the abundance of data *per se* is not always a sufficient condition for having good EVT estimates. This is in agreement with the findings of Huisman *et al.* (2001), who model approximately the same (not scanty, in principle) data-set of Loretan and Phillips (1994) but obtain different estimates, due to the use of a small sample adjusted Hill estimator (see 1.1 above) instead of the standard one[48].

[48] Huisman *et al.* (2001, p. 213).

Part II

Financial Applications

Having discussed in the first part the main features of EVT and the main issues that can arise when applying it to empirical studies, in this second part we review the main financial applications of EVT.

6 Testing for Different Distributional Assumptions

Pioneering studies by Mandelbrot (1963, 1967) and Fama (1963, 1965) questioned the adequateness of the assumption of normality for financial time series, thus being either regarded as heterodox (given the successful employment of the normality assumption in the theories of Markowitz and Sharpe) or ignored (as in the Black and Scholes model for option pricing). There is now a lot of evidence that the normal distribution is too thin-tailed to adequately fit financial data from many different markets. That is to say, extreme observations are much more likely to occur than one would predict according to the normality assumption. This hints at the possibility that distributions with power-law tails be better suited to model real data than the Gaussian exponentially decaying tail is.

A deep study of fat-tailed distributions[49] has then been carried out, aiming to find appropriate distributions to fit financial data. The main candidates considered in the literature are stable laws and the Student t distribution. A considerable debate on the advantages of one or the other has taken place:

- Student t distribution combines fat tails (observed in practice) with the existence of variance (assumed in many economic and financial models); skewed versions of the t distribution have also been proposed to take into account empirical evidence of skewness in several financial time series;
- stable laws, first suggested by Mandelbrot as an appropriate modelling tool, allow for both heavy tails and skewness; furthermore, for any fixed characteristic index α, they are a closed family with respect to addition, but, on the other hand, non-normal stable laws have infinite variance.

Notice that, as a consequence, another way to state the problem is to determine the existence of the moments of the distribution underlying the data. The issue is not merely of theoretical concern, since different distributional assumptions can lead to dramatically different estimates of expected risk[50].

[49] A reference book on this topic is Rachev (ed.) (2003).

[50] Lux (1996) compares the expected frequency of daily returns exceeding 20% (in absolute value) for different estimates of the tail index α. For $\alpha = 1.7$ (corresponding to a stable law), the expected frequency is of once within seven years, while $\alpha = 3.0$ (Student t) yields once within 75 years.

Anyway, a comparative study of the two distributional assumptions is difficult, since stable and t distributions are not nested and thus a direct test rejecting one of them in favour of the other seems difficult to conceive.

Since these different distributional assumptions are strictly related to the amount of heaviness in the tails, EVT looks very promising. Indeed, EVT offers a remarkably elegant and effective solution to the problem of a direct comparison of the two families, as EVT nests both models, given that both stable laws and the Student t are in the maximum domain of attraction of a Fréchet type extreme value distribution, with the tail index α equal to the characteristic index of the stable law and to the degrees of freedom of the t distribution, respectively (1.1).

Therefore, in practice:

(a) assuming that the distribution of the data belongs to the domain of attraction of a Fréchet type extreme value distribution, estimate the tail index α of the distribution (or estimate the shape parameter, which immediately yields the tail index, as its inverse);

(b) if α is significantly less than two, then conclude in favour of the stable law assumption; on the contrary, if α is greater than two, the Student t assumption is more appropriate.

This EVT approach was first employed by Koedijk *et al.* (1990), trying to evaluate how heavy-tailed are bilateral EMS foreign exchange rates. Using weekly data on spot exchange rates of 8 different currencies quoted against the US dollar, approximately during the period 1971-1987, the authors find point estimates for α below two for most exchange rates and, at a 5% significance level, the hypothesis $\alpha < 2$ is never rejected, while the hypothesis $\alpha > 2$ is rejected on few occasions. In any case, distributional assumptions of normal and of mixture of normal distributions are definitely rejected. Moreover, they undertake a study of the stability of α with respect to the creation of the EMS and of the effects of aggregation over time.

Lux (2000) uses this approach, based on the estimation of the tail index, in order to obtain a definite conclusion about the finiteness of the second moment of the distribution of German stock returns, since previous papers recorded diverging results, and concludes that, when using algorithms for endogenous selection of the optimal cut-off, there is evidence for heavy-tails with finite variance.

Vilasuso and Katz (2000) use daily aggregate stock-market index prices for several countries ranging from 1980 to 1997 in order to assess the hypothesis that the returns follow a stable distribution and they conclude that there is scant support for this assumption, while Student t and ARCH processes seem to be more suitable models.

An analogous conclusion is drawn by Longin (2005), who applies the method to logarithmic daily percentage returns on the S&P500 index for the period 1954-2003 and concludes that both the normal and the stable law hypotheses are rejected, while the Student t distribution and ARCH processes are not, thus providing reasonable tools for unconditional and conditional, respectively, modelling of the US stock market[51].

[51] These conclusions corroborate previous findings from Jansen and de Vries (1991) and Loretan and Phillips (1994), who, analyzing US stock market returns, concluded for the existence of finite second moments (possibly finite third and fourth moments as well, but not higher then the fourth one, in any case).

7 Market Risk: VaR and ES Estimation

While testing for different distributional assumptions has been the main concern of financial applications of EVT during the first decade since EVT was recognized as a valuable tool for financial studies (roughly, the Nineties of the past century), the new millennium definitely saw Value-at-Risk (VaR) calculations as the pre-eminent topic in financial applications of EVT, due to the combination of two main factors: on the one hand, the growing interest in risk management and the emergence of VaR as the standard measure for risk, both confirmed and strengthened by the Basel Committee resolutions; on the other hand, the fact that both VaR and EVT are concerned with the tails of the distribution, regardless of its central part, so that EVT seems a very natural approach to VaR estimation.

In this section, we first review the definitions and main properties of VaR and Expected Shortfall (ES), the main alternative to VaR (section 7.1), and recall how VaR and ES estimates can be obtained via EVT (7.2). Then we concentrate on the comparison between EVT and standard approaches to VaR calculation (7.3).

7.1 VaR and ES: Definition and Main Properties

Standard references on VaR are the reviews by Duffie and Pan (1997) and Pearson and Smithson (2002) and the books by Dowd (1998, 2002) and Jorion (1997). Significant critiques to VaR as an effective risk measure are moved by Daníelsson (2002) and by Artzner *et al.* (1999) and Acerbi (2004), who propose more suitable measures (ES and coherent risk measures in general).

After the 1987 crash, in several companies risk measures similar to VaR were introduced, until JP Morgan brought Value-at-Risk to the attention of a wide audience in the mid-Nineties and the Basel II Accord established it as the basis for market risk measurement in financial institutions.

VaR is by definition (the opposite of) *the minimum loss that can occur at a given (high) confidence level, for a predefined time horizon.* Regulatory norms set at ten days the time horizon (the period in which a bank is supposed to be able to liquidate its position) and at 99% the confidence level.

In symbols, the VaR at the $\alpha \cdot 100\%$ confidence level, where $\alpha \in]0,1[$, for a time horizon of 1 time-unit (the same unit that determines the frequency of the data employed) is given by

$$VaR_{\alpha}(X) = -\inf\{x \in \mathbb{R} : P(X \leq x) \geq 1-\alpha\} = -F_X^{\leftarrow}(1-\alpha),$$

where X is a random variable standing for some random return and $F_X^{\leftarrow}$ is the generalized inverse[52] of the cumulative distribution function of X. Notice that, due to the minus sign in its definition, VaR represents losses as positive amounts. When directly considering a time series of positive losses, the previous definition is replaced by

$$VaR_{\alpha}(Y) = \inf\{y \in \mathbb{R} : P(Y > y) \leq 1-\alpha\} = \inf\{y \in \mathbb{R} : F_Y(y) \geq \alpha\} = F_Y^{\leftarrow}(\alpha).$$

[52] $F_X^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}$.

The main advantages of VaR are the following:

- it is easy to understand and interpret from a financial viewpoint, thus providing an effective tool for management purposes;
- it focuses on the tail of the distribution only, thus capturing the occurrence of huge rare losses.

Anyway, VaR has at least two considerable drawbacks:

- it deals more with the cut-off between the centre and the tail of the distribution, rather than with the tail itself, since it provides information about the minimum loss that will occur with a certain low frequency $1 - \alpha$, completely disregarding what happens when going farther in the tail;
- it does not seem to thoroughly behave as a sensible risk measure, given that examples can be provided in which the VaR of a portfolio of investments is greater than the sum of the VaRs of the single investments, thus contradicting the acknowledged role of diversification in lowering the level of risk (mathematically, this means that VaR is not a convex function, i.e. it can happen that $VaR(w_1X + w_2Y) > w_1VaR(X) + w_2VaR(Y)$ for some portfolio weights w_1 and w_2).

A huge variety of approaches to VaR computation have been proposed in the literature. The main alternatives can be roughly divided into three families[53]:

- non-parametric methods, the most outstanding being historical simulation;
- semi-parametric methods, like EVT and CAViaR;
- parametric methods, like RiskMetrics approach and GARCH models.

Historical simulation (HS) is based on the empirical distribution as obtained from the data: the $\alpha \cdot 100\%$ VaR is identified with the (opposite of) the $1-\alpha$ empirical quantile. While avoiding the problem of choosing a distribution for the time series we are considering, HS suffers from three heavy shortcomings[54]: it is unable to predict the occurrence of future extreme events, unless they have some ancestor in the time series; it relies on an iid assumption, which is unrealistic for financial data; it is unable to provide reliable estimates of high confidence level VaR, since it would require too many data.

On the contrary, fully parametric methods are based on an explicit distributional assumption to model the time series. The model is fitted to the data and, by means of the estimated parameters, allows one to calculate the VaR at any specified confidence level.

The main problem with this approach is that we do not know the actual distribution of the data at hand (this is often a major problem we are in need to solve, as seen in the previous

[53] This classification, up to minor taxonomical details, is quite standard. See for instance Manganelli and Engle (2004) and Gençay and Selçuk (2004).

[54] Rachev *et al.* (2008, p. 188).

section). Therefore, parametric methods can result in inaccurate estimates of VaR, potentially yielding either severe losses (if the actual risk level is underestimated), or unfruitful conservative positions (in the case of overestimation).

Semi-parametric methods try to address this issue, merging advantages of both non-parametric and parametric methods. EVT, in this sense, can be regarded as a semi-parametric approach to VaR, since it avoids imposing a given distribution to the whole set of data (thus reducing model risk), but it focuses on the tail, trying to model its asymptotic behaviour (thus not incurring in the main difficulties of non-parametric methods either).

Irrespectively of the approach chosen to calculate VaR, we cannot overcome the two flaws of VaR mentioned above, i.e. the lack of control on what happens far away in the tail and the lack of convexity (more precisely, the lack of the subadditivity property). To tackle these difficulties, further risk measures have been proposed, the best known being *Expected Shortfall* (ES).

Expected Shortfall, also known as *Conditional Value-at-Risk* (CVaR), or *Average Value-at-Risk* (AVaR), at the confidence level $\alpha \cdot 100\%$, is defined[55] as *the average of the VaRs which are greater than* VaR_α, i.e.

$$ES_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_t(X)\,dt.$$

Therefore, by definition, ES at the confidence level $\alpha \cdot 100\%$ takes into account what happens in the tail, beyond the correspondent VaR_α. Moreover, it can be shown that ES is subadditive, thus standing as an interesting candidate for a sensible risk measure[56].

7.2 An EVT Approach to VaR and ES Calculation

The definitions of VaR and ES are essentially concerned with extreme quantiles of the distribution. Therefore, EVT provides a natural solution to the problem of their estimation.

A simple EVT estimate of VaR can be obtained[57] applying the following algorithm:

(a) assuming that the data are in the maximum domain of attraction of a GEV distribution, fix a high threshold u and fit the GPD to the exceedances over u, obtaining estimates $\hat{\xi}$ and $\hat{\beta}$ of the parameters ξ and β, respectively;

(b) estimate the tail probability by means of $\hat{\bar{F}}(x) = \frac{N_u}{N}\left(1+\hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}$, where $\frac{N_u}{N}$ is the ratio between the number N_u of exceedances observed in the given sample and the size N of the sample itself and stands for an estimator of $\bar{F}(u)$;

(c) invert the previous formula to obtain an estimate of the α quantile, i.e.

[55] Expected Shortfall is well defined only for random variables X with finite mean $E|X| < +\infty$ (Rachev *et al.*, 2008).

[56] Indeed, ES is a coherent risk measure in the sense of Artzner *et al.* (1999), i.e. it satisfies the following four axioms: monotonicity, positive homogeneity, subadditivity and translation invariance.

[57] See for instance McNeil *et al.* (2005, p. 283), or Gençay and Selçuk (2004, p. 298).

$$VaR_{\alpha}(X) = u + \frac{\hat{\beta}}{\hat{\xi}}\left[\left(\frac{N}{N_u}\cdot(1-\alpha)\right)^{-\hat{\xi}} - 1\right].$$

The estimator for the tail probability at step (b) can be justified decomposing $\bar{F}(x) = 1 - F(x) = P(X > x)$ in the product $P(X > u)\,P(X > x|X > u)$, according to the definition of conditional probability, and exploiting the equalities $P(X > u) = \bar{F}(u)$ and $P(X > x|X > u) = P(X - u > x - u|X > u) = \bar{F}_u(x - u)$.

From the given estimate of $VaR_{\alpha}(X)$, one can obtain an estimate of $ES_{\alpha}(X)$ as well, namely

$$ES_{\alpha}(X) = \frac{1}{1-\alpha}\int_{\alpha}^{1} VaR_t(X)\,dt = \frac{VaR_{\alpha}}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}}.$$

The previous arguments rely on the application of a GPD-based approach, modelling threshold exceedances. If, in turn, the other main parametric method is employed (the block maxima method), one can estimate $VaR_{\alpha}(X)$ by means of the $\alpha^n \cdot 100\%$ confidence level Value-at-Risk for the block maxima M_n, taking into account the asymptotic equality[58] $VaR_{\alpha}(X) = VaR_{\alpha^n}(M_n)$. Otherwise, if a non-parametric approach is pursued, using the Hill estimator, Value-at-Risk can be estimated by means of[59]

$$VaR_{\alpha}(X) = X_k \cdot \left[\frac{N}{N_u}\cdot(1-\alpha)\right]^{-\hat{\xi}_k^H},$$

where X_k denotes the kth order statistic (the same one chosen as a cut-off when estimating the shape parameter ξ with the Hill estimator) and α indicates here the confidence level at which Value-at-Risk is calculated (not the tail index, unlike in section 1.1).

The previous EVT approaches to VaR and ES estimation are unconditional, as they don't take into account the dependence structure of the time series under consideration. When conditional VaR has to be estimated, i.e. Value-at-Risk conditional on past information, considering the possible heteroskedasticity of the data, the techniques considered in section 3 have to be integrated in the previous scheme. For instance, a common and effective approach is based on the two-step procedure applied by McNeil and Frey (2000)[60], which first fits a GARCH-type model to the data and then applies EVT to the standardized residuals (see 3.2 above). Value-at-Risk is then computed for the standardized residuals according to the unconditional scheme we have just presented and the conditional estimate of VaR is obtained by substituting the unconditional quantile Z_t in the equation $X_t = \mu_t + \sigma_t Z_t$ governing the dynamics of returns.

[58] See Longin (2000), where the explicit derivation of a GEV-based estimator of Value-at-Risk is provided, using this equality.

[59] See Daníelsson and de Vries (1997a).

[60] See also Kuester *et al.* (2006).

7.3 Comparative Studies

The first influential studies applying EVT to VaR computation are probably those of Pownall and Koedijk (1999), Daníelsson and de Vries (1997b, 2000), Longin (2000), McNeil and Frey (2000) and Neftci (2000).

Pownall and Koedijk (1999), studying the crisis of Asian markets, provide a conditional approach to VaR calculation employing EVT and find that it yields an improvement in VaR estimation, compared to the technique employed by RiskMetrics.

Longin (2000), estimating with data from the S&P 500 index, makes a comparison between EVT and four standard methods for calculating VaR, namely historical simulation, modelling with normal distribution, modelling with GARCH processes and the exponentially weighted moving average (EWMA) process for the variance employed by RiskMetrics. This paper first considers, in the computation of VaR in an EVT setting, both long and short positions, to which the lower and the upper tail of the distribution, respectively, are of interest [61].

McNeil and Frey (2000) apply their two-step procedure[62] to obtain conditional VaR and ES estimates for S&P 500 and DAX indices, BMW share price, the US dollar/British pound exchange rate and the price of gold. Comparing the estimates provided by the two-step method with those coming from unconditional EVT, GARCH modelling with conditional normal innovations and GARCH modelling with conditional t innovations, they conclude[63] that, on the whole, the conditional approach to VaR provided by EVT outperforms the others.

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. He concludes that "the results, applied to eight major risk factors, show that the statistical theory of extremes and the implied tail estimation are indeed useful for VaR calculations. The implied VaR would be 20% to 30% greater if one used the extreme tails rather than following the standard approach"[64].

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. See for instance Bao *et al.* (2006), Bali (2007), Bekiros and Georgoutsos (2005), Brooks *et al.* (2005), Gençay and Selçuk (2004), Ho *et al.* (2000), Kuester *et al.* (2006), Lee and Saltoğlu (2002), Manganelli and Engle (2004), Tolikas *et al.* (2007).

The overall impression emerging by an analysis of the literature is that consensus has been reached on the following main conclusions:

- EVT based estimates of VaR outperform estimates obtained with other methodologies for very high quantiles, namely for $\alpha \geq 0.99$;
- the farther one moves into the tails, i.e. the greater α, the better EVT estimates are;

[61] The calculation of VaR for both long and short positions has been performed later on also by Ané (2006), Byström (2007), Marimoutou *et al.* (2009) and Ho *et al.* (2000), among others.

[62] See 3.2 above.

[63] "In 11 out of 15 cases our approach is closest to the mark. On two occasions GARCH with conditional t innovations is best and on one occasion GARCH with conditional normal innovations is best. In one further case our approach and the conditional t approach are joint best. On no occasion does our approach fail (lead to rejection of the null hypothesis), whereas the conditional normal approach fails 11 times; unconditional EVT fails three times" (McNeil and Frey, 2000, p. 290).

[64] Neftci (2000, p. 14).

- when $\alpha < 0.99$ ($\alpha = 0.95$, typically), evidence is mixed and EVT can perform significantly worse than other techniques.

These conclusions invite us to somehow circumstantiate the absolute confidence that some preliminary studies bestowed on EVT methods and recognize both the actual advantages of EVT and its limitations.

A significant result in this sense was already established by Daníelsson and de Vries (2000), who, comparing two different EVT estimators with historical simulation and RiskMetrics on data from U.S. stocks, stated that "for the 5th percentile, RiskMetrics performs best. The reason for this is that at the 5 % level we are sufficiently inside the sample so that the conditional prediction performs better than unconditional prediction. However, as we move to the tails, RiskMetrics consistently underpredicts the tail, with ever larger biases as we move farther into the tails"[65]. In other words, when estimating $\text{VaR}_{0.95}$ we are not gone sufficiently far in the tails yet in order for EVT to work efficiently, while from the 99% confidence level upward EVT is more appropriately employed, since in this case we are really concerned with extreme events.

To a similar conclusion point also the results of Gençay and Selçuk (2004), who compare EVT to historical simulation and normal and Student t distribution modelling, using daily stock market data from several emerging economies around the world[66]. When studying the DAX index also Tolikas *et al.* (2007) reach the same conclusion, though finding that, when a sufficiently large amount of data is available, historical simulation can yield comparable results (thus somehow at variance with the current opinion on HS, as seen above).

An important refinement to the previous conclusions is reported by Bekiros and Georgoutsos (2008a). When studying market data as summarized by the Cyprus Stock Exchange general index, their findings agree with the ones just mentioned. On the contrary, when turning to daily returns on the US dollar/Cyprus pound exchange, the performance of EVT methods is considerably worse than in other studies, definitely ruling out EVT as a viable candidate for estimating "low" confidence level ($\alpha < 0.98$) VaR (even for higher confidence levels, EVT is comparable to other methods, not outperforming them). The authors explain this fact with the relatively scant tail fatness of exchange returns, compared to stock market data.

The previous remark is interesting, as it stresses an important feature that should be underscored in any comparative study, namely the (possible) dependence of the outcome of a comparison on the particular kind of data employed. The vast majority of papers applying EVT to finance employ time series from the stock market. The second most relevant source of data is probably represented by exchange rates, but there are studies dealing with almost any sort of data, from equity returns and interest rates[67] to energy and commodity market data[68], up to

[65] Daníelsson and de Vries (2000, p. 254).

[66] "Again, at moderate levels of both tails, there is no clear winner. However, as we move towards the higher quantiles, the GPD model is clearly better. [...] Especially at 99th and higher quantiles, the GPD model clearly dominates others in terms of VaR forecasting in nine emerging markets we cover in this study" (Gençay and Selçuk, 2004, p. 300).

[67] Bali (2003a) analyzes daily observations for the annualized yields on 3-month, 6-month, 1-year and 10-year U.S. Treasury securities; Neftci (2000), as seen above, includes in his data-set three-month, two-year, five-year and seven-year interest rates from the U.S.; while Werner and Upper (2004) study Bund futures returns.

[68] See e.g. Byström (2005) and Chan and Gray (2006) for electricity markets, Krehbiel and Adkins (2005) for an application to the NYMEX energy complex and Marimoutou *et al.* (2009) for an application to the oil market.

credit derivatives data[69].

Notice that the previous conclusions roughly hold for both conditional and unconditional VaR estimates. Kuester *et al.* (2006) state that unconditional models for VaR can actually perform quite well if $\alpha = 0.99$, while yielding unacceptable estimates for lower values of α (0.975 and 0.95). This is the case not only for EVT, but also for historical simulation and a skewed t-based parametric approach, while normal and (symmetric) t distributions perform even worse. These findings hint at the necessity of a conditional approach to VaR estimation. Comparing thirteen conditional models on data from the NASDAQ Composite Index, the authors conclude that the two-step EVT procedure and filtered historical simulation (FHS)[70], both with normal and skewed t innovations for the GARCH filter, "are always among the best performers among all conditional models and for all values of $0 < \lambda < 0.1$ [$0.9 < \alpha < 1$, in our notation]. Moreover, the ST-EVT delivers virtually exact results for all $0 < \lambda < 0.1$ [$0.9 < \alpha < 1$], while N-EVT is competitive for $0 < \lambda < 0.025$ [$0.975 < \alpha < 1$] and then weakens somewhat as λ increases toward 0.1 [decreases toward 0.9]"[71].

■ *Appendix: Backtesting*

In this section, in order to keep things as straightforward as possible and to highlight common paths and conclusions, we have omitted to address a relevant issue[72], that we briefly develop now on its own. That is, studies comparing different techniques for VaR estimation have to choose some performance criterion on the basis of which they rank those techniques.

The most common criterion employed when using methods for unconditional VaR estimation is the *violation ratio*[73]. A violation is said to occur at time $t+1$ if the observed realization x_{t+1} is greater[74] than the estimated Value-at-Risk $\hat{VaR}_\alpha^{t+1}$. A detailed description of the method is provided by Kupiec (1995), Christoffersen (1998), McNeil and Frey (2000) and Bali (2007). Tests related to the violation ratio can be carried out for both in-sample and out-of-sample performance of Value-at-Risk estimates. In the former case, one looks for violations in the same time window used to estimate the parameters of the model, while in the latter case one focuses on a subsequent period for backtesting. In practice, testing for instance for out-of-sample performance:

(a) consider a time window of n observations $\{x_{t-n+1}, \ldots, x_t\}$ and calculate $\hat{VaR}_\alpha^{t+1}$ on the basis of this sequence;

[69] Byström (2007) uses data from the iTraxx Europe CDS index market in order to infer margin calculations for an hypothetical exchange-traded CDS index futures market.

[70] Proposed by Barone-Adesi *et al.* (1998), FHS shares with the two-step procedure of McNeil and Frey (2000) the first step, i.e. the GARCH filtering of the data, but at the second step employs historical simulation rather than EVT.

[71] Kuester *et al.* (2006, p. 79). The authors perform their analysis for different time window sizes (1000, 500, 250) obtaining in any case evidence in favour of skewed t EVT and FHS approaches outperforming the other ones (in particular for a time window of 250 observations), or being at least comparable with them.

[72] Another interesting path that can be followed in the literature on EVT-based estimation of VaR is the problem of time aggregation and the square root scaling rule, suggested by the Bank of International Settlements but criticized by some academicians. On this topic, see for instance McNeil and Frey (2000), who propose a Monte Carlo simulation approach, and Jansen *et al.* (2000), who in turn discuss the use of a "tail index-root" rule, rescaling through a root of the time horizon as in the square root rule, but with the index of the root given by the tail index of the distribution. See also Diebold *et al.* (1998) and Daníelsson and de Vries (1997b).

[73] Interesting remarks on the main issues related to backtesting of VaR estimates can be found in Pearson and Smithson (2002). Another comparison criterion that we do not present here, but which is sometimes used, is that proposed by White (2000).

[74] For simplicity, we are assuming a sample of positive losses.

(b) consider then the following p observations, $\{x_{t+1}, \ldots, x_{t+p}\}$, and define a counter of overall violations occurred in this sample by $V = \sum_{j=1}^{p} I_{\left\{x_{t+j} > V\hat{a}R_{\alpha}^{t+j}\right\}}$; the violation ratio is then defined as $\kappa = V/p$ (number of violations occurred in a sample of p observations);

(c) test the null hypothesis that $\kappa = 1 - \alpha$, exploiting the fact that, under this hypothesis, the probability of observing V violations in a sample of p realizations can be computed by means of a binomial distribution with parameters p and $1 - \alpha$.

A suitable test, proposed by Kupiec (1995) to check if the equality $\kappa = 1 - \alpha$ is statistically tenable[75], is the likelihood ratio statistic

$$LR_{uc} = 2\left\{\ln\left[\kappa^{V}\left(1-\kappa\right)^{p-V}\right] - \ln\left[\left(1-\alpha\right)^{V}\alpha^{p-V}\right]\right\},$$

which is known as *test for unconditional coverage* and is asymptotically $\chi^2(1)$ distributed.

Since the unconditional coverage test is not sufficiently powerful when dealing with small samples and, more importantly, it does not take into account the volatility clustering characterizing financial time series, it can be a misleading criterion for VaR estimate accuracy, if considered on its own.

Another test, the *test for conditional coverage* developed by Christoffersen (1998), considers both unconditional accuracy and independence of violations (as unconditional VaR estimates tend to exhibit a violation pattern which is flawed by clustering effects), since it reads

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

where LR_{ind} stands for a second likelihood ratio statistic, devoted to test for the null hypothesis of serial independence against the alternative hypothesis of Markov dependence of the first order. Since LR_{ind} converges in distribution to a $\chi^2(1)$ as well, then LR_{cc} converges to a $\chi^2(2)$.

8 Asset Allocation

A portfolio selection problem based on the maximization of returns subject to risk constraints can be formally written as

$$\begin{array}{lll} \max\limits_{w\in\mathbb{R}^n} & \mu^T w & \\ s.t. & \rho_j(w) \leq R_j, & j = 1, \ldots, J \\ & \sum_{i=1}^{n} w_i = 1 & \\ & w_i \geq 0, & i = 1, \ldots, n, \end{array}$$

[75] Gençay and Selçuk (2004) argue that a violation ratio κ less than the expected value $1 - \alpha$ is not necessarily better than a violation ratio exceeding that value, if we are not committed to policy purposes. To risk managers the closeness to the value $1 - \alpha$ should matter above all, since, if on the one hand $\kappa > 1 - \alpha$ is not a good case (given that it indicates that we are underestimating the actual riskiness of our position), on the other hand $\kappa < 1 - \alpha$ is not a good scenario either, as it hints at an excessively conservative position.

where w_i are the unknown weights of the portfolio, $\mu^T w$ is the expected value of the portfolio, ρ_j are risk measures and R_j are upper bounds on risk. When $j = 1$ and $R_1(w) = w^T \Sigma w$, we recover the classical mean-variance framework of Markowitz (1952).

Anyway, as stressed by Bensalah (2002, p. 5), "there is no final answer or optimal measure of risk", the optimality of a given allocation being conditional on the actual correspondence of its underlying assumptions to the risk preference of the investor. A plausible profile of risk preference, alternative to that based on variance, is represented by the safety-first criterion, a concept introduced by Roy (1952) and developed by Arzac and Bawa (1977), which is based on a constraint limiting downside risk. As Jansen *et al.* (2000) argue, a similar profile is both relevant in practice, given that practical circumstances can impose an asymmetric treatment of upside and downside risk, and psychologically sensible, since a lot of experimental evidence for loss aversion is available.

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 may provide suitable tools for accurate calculations. Indeed, it has been used in the context of portfolio selection with limited downside risk by Jansen *et al.* (2000). Facing the problem of choosing between investing in a mutual fund of bonds or a mutual fund of stocks, they 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. On the same lines, Susmel (2001) uses EVT to model the tails of Latin American emerging markets and studies the effects of diversification for a US investor including those markets in his portfolio, based on a safety-first criterion.

This framework extends the use of EVT for VaR calculation to the problem of portfolio selection[76], i.e. problem (4) with risk measure ρ given by VaR. In this mean-VaR setting, Consigli (2002), for instance, solves the asset allocation problem with data comprising the Argentinean crisis of July 2001, 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 analyzed; 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.

A detailed analysis of the mean-VaR portfolio selection problem is given by Bensalah (2002), who considers both the direct problem of maximizing return, subject to a constraint on the VaR, that is problem (4) with $j = 1$ and $R_1(w) = VaR_\alpha(w)$, and its dual, i.e. the problem of minimizing Value-at-Risk, subject to a constraint which imposes a lower bound to the expected return. Moreover, Bensalah makes a comparison between different ways of calculating VaR_α, namely historical simulation, normal VaR 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 level α (0.99, 0.999, 0.9999), he concludes that historical simulation and normal VaR yield the same allocation irrespectively of the given α, investing the whole capital in the short-term security, both in the maximum return and in the minimum risk problem. This is no more true when we consider EVT based calculations of VaR. Indeed, in this case two important features appear:

- portfolio composition changes as the confidence level α grows;

[76] The importance of VaR for asset allocation is detailed by Bali (2007), who derives a modified version of the CAPM, relating the downside risk of a portfolio, as measured by VaR, to the expected return of the portfolio itself.

- the riskier (in terms of duration) asset is given non-zero weight.

Precisely, the weight acknowledged to this asset is an increasing function of α. Moreover, its weight is greater in the minimum risk than in the maximum return problem.

Thus Bensalah can conclude that an EVT approach, being tailored on extreme events and differently responding to different confidence levels, seems to be better suited to regulatory purposes and to cope with periods of market stress. In the latter circumstances, EVT could offer an alternative strategy to the "flight to quality" strategy, thus reducing the behaviour risk (the crisis could be exacerbated if everybody reacted the same way).

Finally, to make effective computation possible when dealing with multiple assets, Bensalah (2002) proposes an algorithm. The problem that motivates its introduction is the relevance of the dimensionality issue in portfolio selection, since, as the number of assets grows, a brute force approach to the calculation of the optimal portfolio becomes soon intractable, even numerically (for n assets and a precision of 1%, 100^{n-1} calculations are needed). The algorithm proposed allows one to fix in advance the number m of portfolios among which to search and is based on m iterations of the following step: randomly generate the weight w_1 of the first asset from a uniform distribution with support on the interval [0,1] and set[77] $w_2 = 1 - w_1$; then, estimate the extreme value distribution of the resulting portfolio and the corresponding risk and return. After m iterations, one is endowed with m portfolios[78] and can choose the optimal one in this sample.

An improvement on this algorithm is provided by Bradley and Taqqu (2004a), who propose a two-step procedure that, in the first step, uses the sampling algorithm of Bensalah to generate a sensible starting point for the application of an incremental trade algorithm at the second step. On the same lines of Bensalah (2002), Bradley and Taqqu (2004a) study the risk-minimization problem comparing normal and extreme value distributions in the computation of VaR and ES as risk measures. Making both simulation and empirical analyses, they find that:

(a) the optimal allocation is quantile-based, i.e. depends on α, confirming the findings of Bensalah (2002);

(b) at standard confidence levels ($\alpha = 0.975, 0.99$), the optimal allocation can differ from that obtained under the hypothesis of normal distribution, but the extra risk taken under the normality assumption is not particularly large, thus making it a viable alternative to EVT, especially given its easiness of implementation; however, when moving to extreme quantiles ($\alpha = 0.999, 0.9999$), the difference between the two approaches cannot be ignored anymore;

(c) in the simple case of two assets, as α changes, the weight given to each asset strictly follows the ratio of marginal risks between the two corresponding markets; specifically, under asymptotical independence, the optimal portfolio diversifies more into the riskier asset for very high confidence levels α.

The latter point is related to the problem of diversification and the issue of asymptotic dependence, stating that the driving factor is rather the ratio of marginal risks than dependence.

[77] When considering n different assets, w_2 is randomly generated from a uniform distribution with support $[0, 1 - w_1]$, w_3 from a uniform distribution with support $[0, 1 - w_1 - w_2]$, and so on. Finally, one sets $w_n = 1 - w_1 - w_2 - \ldots - w_{n-1}$.

[78] It is easy to program the algorithm in such a way that the portfolios generated be different.

Diversification is widely acknowledged as a leading rule of portfolio selection (several hedge funds successfully employ it), but the benefit of diversification can be seriously hampered when the markets used to diversify risk away exhibit some form of dependence. Anyway, in accord with Poon *et al.* (2004), the findings of Bradley and Taqqu (2004a) for twelve international equity markets are that most of them are asymptotically independent and the few cases of asymptotic dependence are related to each other by geographic proximity.

This explains why one can reasonably study a portfolio selection problem, which in principle is a multivariate problem, by means of univariate techniques (applied to some functional of the allocation vector, typically the portfolio return).

A direct comparison of the performance of univariate and multivariate techniques for an asset allocation problem is the aim of Bradley and Taqqu (2004b), who conclude that multivariate EVT yields accurate results if the copula employed allows for asymptotic independence, thus reflecting a chief characteristic of the data at hand, while multivariate EVT models with a dependence function[79] that prescribes asymptotic dependence overstate the risk of a properly diversified portfolio[80]. On the contrary, the use of a univariate approach, which is easier to implement, seems to be advisable when we are not interested in particularly extreme quantiles.

9 Dependence Across Markets: Correlation and Contagion

The issue of dependence among financial time series we have mentioned when treating the problem of portfolio selection can be specifically addressed when studying correlation and contagion among markets. Two approaches have been adopted in this respect: on the one hand, dependence between different countries for a given financial sector has been studied; on the other hand, dependence among different financial markets of the same country is of interest as well.

A highly relevant example of the former approach is given by Longin and Solnik (2001), who employ multivariate EVT to test the validity of the common belief that correlation between stock markets increases during volatile periods. They use multivariate threshold exceedances, modelling dependence by means of 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. Fitting a bivariate model to couples of monthly equity index returns for the U.S. and each of the other four countries of the G-5 (United Kingdom, France, Germany and Japan), the authors find that extreme correlation is statistically different from the correlation estimated by means of a multivariate normal model. Furthermore, the correlation pattern of threshold exceedances is asymmetric, since it depends on the chosen threshold both in size and in sign. Indeed, correlation of exceedances is increasing in the absolute value of the threshold, if the threshold itself is negative; otherwise, it is decreasing. Therefore, Longin and Solnik can conclude that "[...] the probability of having large losses simultaneously on two markets is much

[79] Properly speaking, the copula and the dependence function of a multivariate extreme value distribution do not coincide, though they are closely related. The definition of the latter can be found in most of the reference books on EVT, e.g. Reiss and Thomas (1997).

[80] On the other hand, Poon *et al.* (2004, p. 601) conclude that "if asset returns are asymptotic dependent, the risk estimation is not sensitive to the wrong assumption for the tail dependence and the methods used to estimate large event risk". This is a further justification for the employment of simple but effective models such as, for instance, the logistic one, when asymptotic dependence has been ascertained (see 4.2 above).

larger than would be suggested under the assumption of multivariate normality. It appears that it is a bear market, rather than volatility per se, that is the driving force in increasing international correlation"[81].

In a complementary way, Bekiros and Georgoutsos (2008b) study the correlation of extreme returns between seven Asia-Pacific stock markets and the U.S. They use a bivariate logistic dependence function to model threshold exceedances, as Longin and Solnik (2001), in order to produce a ranking of Asia-Pacific countries in three broad categories of risk, to check whether these countries form a distinct block, with respect to, for instance, Europe and the U.S. The conclusion is in the negative, and U.S. investors can benefit from diversifying their portfolios with assets from Asian countries. This remains true even during crisis periods, given that a sensitivity analysis conducted by estimating the model with two separated sets of data, one including the 1987 crash and the other subsequent to it, yields similar results. In particular, 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 crises of the Eighties and of the Nineties.

Anyway, the choice of the previous articles to measure extreme dependence by means of the correlation coefficient is questionable. Poon *et al.* (2004), instead, use the coefficient of upper tail dependence (see 4.1 above) and a complementary measure developed by Ledford and Tawn (1996) to assess dependence across markets. Precisely, they analyze daily returns on stock indices of the G-5 countries, by means of bivariate EVT. Since nonzero estimates of the coefficient are obtained only for 13 out of 84 possible pairs of countries, the assumption of asymptotic dependence would be inappropriate in the vast majority of cases, resulting in an overestimation of systemic risk in international markets.

For policy purposes, a particular interest is attached to studies regarding currency crises contagion and cross-country dependence in the banking sector.

The former issue is considered for instance by Haile and Pozo (2008), who apply EVT to set appropriate thresholds that enable to discriminate between crisis and normal periods and conclude that currency crises, as experienced in the Nineties are indeed contagious and that the main way for contagion to occur is the trade channel. Moreover, regional proximity plays a significant role as well (neighbourhood effects channel). Similar results on the contagious nature of currency crises are obtained by Garita and Zhou (2009), who indicate EVT as a suitable tool to detect contagion. Using an exchange market pressure (EMP) index to measure currency crises, as it is standard in the literature, they choose as a dependence measure a quantity introduced by Embrechts *et al.* (2000) and first applied by Hartmann *et al.* (2004) (see below): namely, the expected value of the number k of extreme events (the EMP index of k countries exceeding a high Value-at-Risk threshold, in the case of Garita and Zhou), conditional on the fact that at least one such event has occurred ($k \geq 1$). The findings thus obtained rule out currency crises contagion as a global phenomenon, confirming it, on the contrary, as a regional one. Moreover, the authors conclude that financial openness and a monetary policy aiming at price stability can reduce the probability of currency crises.

As the latter issue is concerned, i.e. that of banking sector dependence, an interesting reference is represented by Hartmann *et al.* (2005), who study banking sector stability separately considering daily returns on stock prices (during the period 1992-2004) of two groups of 25 major

[81] Longin and Solnik (2001, p. 670).

banks of the U.S. and of the Euro area. The main tool of analysis is multivariate EVT (reduced, with appropriate manipulation, to a univariate setting) and dependence is studied by means of two different measures: one that accounts for bank contagion risk, considering multivariate extreme spillovers; the other measuring aggregate banking systemic risk, with reference to a benchmark (e.g., stock market indices). The former indicator provides evidence of bank spillover risk being lower in the Euro area than in the U.S., probably due to weak cross-boarder linkages of European countries. The second measure, on the other hand, yields similar effects of macro shocks on both European and American banking sector stability. Moreover, structural stability tests detect an increase (very gradual, actually) in systemic risk, both for Europe and the U.S., in the second half of the Nineties.

Increasing cross-country interconnections during the first years of the new Millennium have been detected by Chan-Lau *et al.* (2007) in their study on contagion risk related to the role of London as a hub of worldly banking sector, though "home bias" still plays a crucial role, i.e. the risk of contagion among local banks is high compared to cross-border risk. Based on a variant of multivariate EVT, the authors provide a detailed account of contagion risk among the major U.K. banks and with respect to foreign banks (e.g. Barclays is the most prone to contagion from foreign banks, while HSBC provides the highest contagion risk towards them).

Still related to contagion in the banking sector is the work of Pais and Stork (2009), who, motivated by the credit crisis, make an inquiry on contagion risk between Australian-New Zealand banking and real estate sectors. Using EVT, they find that, as a consequence of the credit crisis, the probability of extreme negative returns has increased in both sectors, as well as the probability of inter-sector contagion.

We have thus come to consider the second aspect of dependence, namely dependence across different kinds of financial activities (cross-asset or inter-sector dependence). In this field, an interesting insight is provided by Hartmann *et al.* (2004), who study contagion risk between stock markets and government bond markets of the G-5 countries. Instead of estimating full-sample correlation, which is biased towards normality and is not the actual quantity of interest, they directly focus on the expected number of crashes in a given market (stocks or bonds), conditional on the event that one crash has occurred. Defining crash levels on a historical basis (20% and 8% losses for stocks and bonds, respectively; weekly data over the period 1987-1999 are employed), the authors use a non-parametric approach to multivariate EVT to estimate the conditional expected value they are considering. Extreme cross-border linkages within the same asset class turn out to be stronger for stock markets than for bond markets, while bivariate stock-bond cocrashes, though not displaying particularly high probabilities *per se*, are considerable when compared to the unconditional univariate probabilities of crashes in a single market and, in any case, they are higher than the probabilities of extreme comovements estimated with a multivariate normal distribution.

Finally, a more restricted study concerning stock market index-foreign exchange rate dependence is conducted by Bekiros and Georgoutsos (2008a), who concentrate on the particular case of Cyprus. Using EVT to estimate extreme returns and correlation as an indicator for dependence, they conclude that low correlation can be found between stock market daily returns and exchange rates with U.S. dollar, even during crisis periods. Neither the bear market conditions characterizing the period covered by the data employed for backtesting make any difference to the estimated extreme correlation, thus suggesting that the results of Longin and Solnik (2001) do not apply to this context.

10 Further Applications

In the previous sections we have described four main applications of EVT to finance. Though we will not pursue our investigation farther, we want at least to record some more possible applications.

We have mentioned in section 9 the topic of contagion concerning currency crises, though indeed detecting and dating currency crises is itself a relevant issue. EVT has been employed to this purpose by Pozo and Amuedo-Dorantes (2003) and Ho (2008), for instance. See also Pontines and Siregar (2007, 2008) for relevant critical observations and Haile and Pozo (2006), who study the connection between the exchange rate regime and the probability of currency crises to occur, concluding that the announced regime has an impact, while the actually observed one does not.

Another important field of application to which growing attention is converging is that of measuring operational risk, mainly due to its explicit consideration in the Basel II Accord. A major pro of EVT in this respect is represented by its capability to model extreme events, such as big unexpected losses due to human errors. On the other hand, a serious limitation is imposed by the scarcity of data, that typically hinders asymptotical theories like EVT. Work in this area can be found in Moscadelli (2004), Dutta and Perry (2006), Allen and Bali (2007), Jobst (2007), Abbate *et al.* (2008), Chapelle *et al.* (2008) and Serguieva *et al.* (2009). A critique to employment of EVT in quantifying operational risk is circumstantiated by Sundmacher and Ford (2007).

Finally, we mention a few other analyses which cannot be subsumed in the previous taxonomy. For instance, several articles are devoted to set optimal margin levels for futures contracts with different underlying, e.g. Longin (1999) (silver futures contracts traded on COMEX), Dewachter and Gielens (1999) (NYSE composite futures), Cotter (2001) (stock index futures selected from European exchanges), Byström (2007) (CDS index futures, as we have already seen). Byström (2006) uses EVT to calculate the likelihood of failure in the banking sector, comparing his outcomes with Moody's and Fitch ratings. Markose and Alentorn (2005) use the GEV distribution to model risk neutral probability and obtain a closed form solution for the price of a European option.

Conclusions

We have presented extreme value theory from a double perspective: on the one hand, the main elements of the probabilistic theory and the statistical methods related to it; on the other hand, their applications to finance.

The first part was intended as a critical resume of both the foundations of the theory and its scope and limitations. From a theoretical viewpoint, EVT shows some considerable pros:

(a) it offers tools, with strong theoretical underpinnings, to model extreme events, which are of great interest in many applications, pertaining to several different fields (in finance, in particular, EVT is especially useful in the context of risk measurement, given the importance of extreme events to the overall profitability of a portfolio);

(b) it provides a variety of such tools, ranging from non-parametric methods to point processes, thus guaranteeing a flexible approach to the modelling of extreme events, that can be adjusted to the particular features of the problem at hand;

(c) the fact that the vast majority of standard distributions, even though displaying considerably different tail behaviour, can be equally modelled by EVT also increases flexibility;

(d) furthermore, the flexibility and the accuracy of modelling are enhanced by the fundamental characteristic of EVT, namely its exclusive consideration of the tail of the distribution of the data, disregarding ordinary observations (the centre of the distribution);

(e) and they are also enhanced by the capability of EVT of independently modelling each tail of the distribution;

(f) finally, the availability of parametric approaches allows for projections and forecasting of extreme events.

Some drawbacks emerged as well:

(a) the most problematic one is probably the dependence of the parameters on the choice of the so-called cut-off (i.e., the delimitation of the subsample employed to estimate the extreme quantiles), given that there is not yet complete agreement on how such a choice should be made;

(b) moreover, the basic theory of extreme values assumes that the data are not serially correlated; when this assumption is violated, we have some alternative approaches at hand, but there is no agreement on which of them is the most suitable one;

(c) multivariate EVT is admittedly not as straightforward as its univariate counterpart and can still encounter severe computational limitations, in some applications;

(d) EVT is characterized by an unavoidable trade-off between its asymptotical nature (always in need for a huge amount of data) and its interest in extreme events (rare, by definition); therefore, the choice and preparation of the data-set can be a crucial step in applying EVT.

Coming to applications, we have only considered financial applications and mainly focused on some of them. The most important one, both for its role in financial regulation and for the amount of contributions to the research concentrating on it, is the employment of EVT for the estimation of quantile-based risk measures, such as Value-at-Risk and Expected Shortfall. Many papers deliver comparative analyses of the accuracy of different methods for VaR calculation and they agree in indicating EVT as a considerably valuable candidate when calculating VaR at high confidence levels (namely greater than or equal to 99%). The great degree of accuracy displayed by EVT-based estimates of VaR for several different markets probably makes the employment of EVT in risk measurement one of the most relevant and better acknowledged contributions of extreme value theory to finance.

Related to the use of EVT for risk management is the role of EVT in asset allocation problems, associated to the concept of "safety first investor". The awareness of the importance of taking into account the risk profile of the investor is permeating the financial practice and, for investors who are particularly interested in avoiding extreme shocks, i.e. huge and rare losses, EVT provides a suitable tool, given its accuracy in modelling such shocks.

Finally, portfolio selection naturally (though not necessarily) entails the consideration of a multivariate setting. In this setting, another important problem is that of systemic risk and the issue of contagion across markets in presence of extreme events. This topic, highlighted by the credit crisis, deserves a particular attention, since the dependence pattern in a multivariate time series can be different in normal times and under stress conditions, i.e. extremal dependence can differ from ordinary correlation. This fact has an impact on diversification effects and has to be explicitly modelled and taken into account. Multivariate EVT offers statistical tools suited to this aim.

The list of possible applications of EVT to finance is longer, of course, but these examples may suffice to show the essence of the contributions that this theory brings to the financial literature (and practice). Those contributions are based on the very definition of extreme value theory, namely on its capability to accurately model the distribution of extreme events, which are the main concern of modern risk management. Thus, in the end, we come back to the widely quoted motto of DuMouchel we began with, which is key to EVT: "Let the tails speak for themselves".

References

[1] Abbate, D., Farkas, W., Gourier, E. (2008). "Operational risk quantification using extreme value theory and copulas: from theory to practice". Working Paper.

[2] Acerbi, C. (2004). "Coherent representations of subjective risk-aversion". In G. Szegö (ed.). *Risk Measures for the 21st Century*. New York: Wiley, pp. 147–207.

[3] Allen, L., and T. G. Bali (2007). "Cyclicality in catastrophic and operational risk measurements". *Journal of Banking & Finance* 31:1191–1235.

[4] Ané, T. (2006). "Two-component extreme value distribution for Asia-Pacific stock index returns". *International Journal of Theoretical and Applied Finance* 9(5):643–671.

[5] Angelidis, T., A. Benos and S. Degiannakis (2007). "A robust VaR model under different time periods and weighting schemes". *Review of Quantitative Finance and Accounting* 28(2):187–201.

[6] Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999). "Coherent measures of risk". *Mathematical Finance* 9:203–228.

[7] Arzac, E. R., and V. S. Bawa (1977). "Portfolio choice and equilibrium in capital markets with safety-first investors". *Journal of Financial Economics* 4:277–288.

[8] Bacmann, J.-F., and G. Gawron (2004). "Fat tail risk in portfolios of hedge funds and traditional investments". Working Paper, RMF Investment Management.

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

[10] Bali, T. G. (2003b). "The generalized extreme value distribution". *Economics Letters* 79:423–427.

[11] Bali, T. G. (2007). "A generalized extreme value approach to financial risk measurement". *Journal of Money, Credit and Banking* 39(7):1613–1649.

[12] Bali, T. G., and S. N. Neftci (2003). "Disturbing extremal behavior of spot rate dynamics". *Journal of Empirical Finance* 10:455–477.

[13] Balkema, G., and P. Embrechts (2007). *High Risk Scenarios and Extremes. A Geometric Approach*. Zurich: European Mathematical Society.

[14] Balkema, A. A., and L. de Haan (1974). "Residual life at great age". *Ann. Probab.* 2:792–804.

[15] Bao, Y., T.-H. Lee and B. Saltoğlu (2006). "Evaluating predictive performance of value-at-risk models in emerging markets: A reality check". *Journal of Forecasting* 25(2):101–128.

[16] Barão, M. I., and J. A. Tawn (1999). "Extremal analysis of short series with outliers: sea-levels and athletic records". *App. Statist.* 48(4):469–487.

[17] Barone-Adesi, G., F. Bourgoin and K. Giannopoulos (1998). "Don't look back". *Risk* 11:100–103.

[18] Basle Committee on Banking Supervision (1995). "An internal model-based approach to market risk capital requirements". Available at http://www. bis.org.

[19] Basle Committee on Banking Supervision (1996a). "Overview of the Amendment to the Capital Accord to incorporate market risks". Available at http://www. bis.org.

[20] Basle Committee on Banking Supervision (1996b). "Supervisory framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements". Available at http://www.bis.org.

[21] Beirlant, J., Y. Goegebeur, J. Segers and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Chichester (England): Wiley.

[22] Beirlant, J., W. Schoutens and J.J.J. Segers (2004). "Mandelbrot's extremism". Discussion Paper 2004-125, Tilburg university, Center for Economic Research.

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

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

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

[26] Bekiros, S. D., and D. A. Georgoutsos (2008b). "The extreme-value dependence of Asia-Pacific equity markets". *J. of Multi. Fin. Manag.* 18:197–208.

[27] Bensalah, Y. (2000). "Steps in applying extreme value theory to finance: A review". Bank of Canada Working Paper 2000–20.

[28] Bensalah, Y. (2002). "Asset allocation using extreme value theory". Bank of Canada Working Paper 2002–2.

[29] Bollerslev, T., and V. Todorov (2009). "Tails, fears and risk premia". CREATES Research Papers 2009–26, School of Economics and Management, University of Aarhus.

[30] Box, G., and D. Cox (1964). "An analysis of transformations". *Journal of the Royal Statistical Society B* 26(2):211–264.

[31] Bradley, B. O., and M. S. Taqqu (2004a). "An extreme value theory approach to the allocation of multiple assets". *International Journal of Theoretical and Applied Finance* 7(8):1031–1068.

[32] Bradley, B. O., and M. S. Taqqu (2004b). "Asset allocation when guarding against catastrophic losses: a comparison between the structure variable and joint probability methods". *Quantitative Finance* 4:619–636.

[33] Brooks, C., A. D. Clare, J. W. Dalle Molle and G. Persand (2005). "A comparison of extreme value theory approaches for determining Value-at-Risk". *Journal of Empirical Finance* 12:339–352.

[34] Byström, H. N. E. (2004). "Managing extreme risks in tranquil and volatile markets using conditional extreme value theory". *International Review of Financial Analysis* 13:133–152.

[35] Byström, H. N. E. (2005). "Extreme value theory and extremely large electricity price changes". *International Review of Economics and Finance* 14:41–55.

[36] Byström, H. N. E. (2006). "Using extreme value theory to estimate the likelihood of banking sector failure". *The European Journal of Finance* 12(4):303–312.

[37] Byström, H. N. E. (2007). "Back to the future: futures margins in a future credit default swap index futures market". *The Journal of Futures Markets* 27(1):85–104.

[38] Chan, K. F., and P. Gray (2006). "Using extreme value theory to measure value-at-risk for daily electricity spot prices". *International Journal of Forecasting* 22:283–300.

[39] Chan-Lau, J. A., S. Mitra, and L. L. Ong (2007). "Contagion risk in the international banking system and implications for London as a global financial center". IMF Working Paper WP/07/74.

[40] Chapelle, A., Y. Crama, G. Hübner and J.-P. Peters (2008). "Practical methods for measuring and managing operational risk in the financial sector: A clinical study". *Journal of Banking & Finance* 32:1049–1061.

[41] Chavez-Demoulin, V., A. C. Davison and A. J. McNeil (2005). "Estimating value-at-risk: A point process approach". *Quantitative Finance* 5(2):227–234.

[42] Chavez-Demoulin, V., and P. Embrechts (2009). An EVT primer for credit risk. To appear in Lipton, A., and A. Rennie (eds.) (2010). *Handbook of Credit Derivatives*. Oxford University Press.

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

[44] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

[45] Consigli, G. (2002). "Tail estimation and mean–VaR portfolio selection in markets subject to financial instability". *Journal of Banking & Finance* 26:1355–1382.

[46] Cotter, J. (2001). "Margin exceedances for European stock index futures using extreme value theory". *Journal of Banking & Finance* 25:1475–1502.

[47] Cotter, J. (2004). "Downside risk for European equity markets". *Applied Financial Economics* 14:707–716.

[48] Cotter, J. (2005). "Extreme risk in futures contracts". *Applied Economics Letters* 12:489–492.

[49] Cotter, J. (2006). "Extreme value estimation of boom and crash statistics". *The European Journal of Finance* 12(6–7):553–566.

[50] Cotter, J. (2007a). "Extreme risk in Asian equity markets". MPRA Paper 3536, University Library of Munich, Germany.

[51] Cotter, J. (2007b)."Varying the VaR for unconditional and conditional environments". *Journal of International Money and Finance* 26:1338–1354.

[52] Cotter, J., and K. Dowd (2006). "Extreme spectral risk measures: An application to futures clearinghouse margin requirements". *Journal of Banking & Finance* 30:3469–3485.

[53] Cotter, J., and K. Dowd (2007). "The tail risks of FX return distributions: A comparison of the returns associated with limit orders and market orders". *Finance Research Letters* 4(3):146–154.

[54] Cotter, J., and F. M. Longin (2006). "Margin setting with high-frequency data". MPRA Paper 3528, University Library of Munich, Germany.

[55] Dacorogna, M. M., U. A. Müller, O. V. Pictet and C. G. de Vries (2001). "Extremal forex returns in extremely large data sets". *Extremes* 4:105–127.

[56] Daníelsson, J. (2002). "The emperor has no clothes: Limits to risk modelling". *Journal of Banking & Finance* 26:1273–1296.

[57] Daníelsson, J., and C. G. de Vries (1997a). "Tail index and quantile estimation with very high frequency data". *Journal of Empirical Finance* 4:241–257.

[58] Daníelsson, J., and C. G. de Vries (1997b). "Value-at-Risk and extreme returns". In P. Embrechts (ed.). *Extremes and Integrated Risk Management.* London: Risk Waters Group, pp. 85–106.

[59] Daníelsson, J., and C. G. de Vries (2000). "Value-at-Risk and extreme returns". Annales d'Économie et de Statistique 60:239–270.

[60] Dekkers, A. L. M., J. H. J. Einmahl and L. de Haan (1989). "On the estimation of the extreme-value index and large quantile estimation". *The Annals of Statistics* 17:1833–1855.

[61] Dekkers, A. L. M., and L. de Haan (1989). "A moment estimator for the index of an extreme-value distribution". *The Annals of Statistics* 17:1795–1832.

[62] Dell'Aquila, R., and P. Embrechts (2006). "Extremes and robustness: a contradiction?". *Financial Markets and Portfolio Management* 20:103–118.

[63] Dewachter, H., and G. Gielens (1999). "Setting futures margins: the extremes approach". *Applied Financial Economics* 9:173–181.

[64] Dias, A., and P. Embrechts (2003). "Dynamic copula models for multivariate high-frequency data in finance". Discussion paper, Department of Mathematics ETH Zurich.

[65] Diebold, F. X., T. Schuermann and J. D. Stroughair (1998). "Pitfalls and opportunities in the use of extreme value theory in risk management". In A.-P. N. Refenes, J.D. Moody and A.N. Burgess (eds.). *Advances in Computational Finance*, 3–12. Amsterdam: Kluwer. Reprinted in *Journal of Risk Finance* 1:30–36.

[66] Dowd, K. (1998). *Beyond Value-at-Risk: The new science of risk management*. Chichester: Wiley.

[67] Dowd, K. (2002). *Measuring Market Risk*. New York: Wiley.

[68] Duffie, D., and J. Pan (1997). "An overview of Value-at-Risk". *Journal of Derivatives* 7:7–49.

[69] DuMouchel, W. H. (1983). "Estimating the stable index α in order to measure tail thickness: A critique". *The Annals of Statistics* 11(4):1019–1031.

[70] Dutta, K., and J. Perry (2006). "A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital". Working Papers 06–13, Federal Reserve Bank of Boston.

[71] Einmahl, J. H. J., and L. F. M. de Haan (2009). "Estimating extreme bivariate quantile regions". Discussion Paper 2009–29, Tilburg University, Center for Economic Research.

[72] Embrechts, P. (2009). "Linear correlation and EVT: Properties and caveats". *Journal of Financial Econometrics* 7(1):30–39.

[73] Embrechts, P., L. de Haan and X. Huang (2000). "Modelling multivariate extremes". In P. Embrechts (ed.). *Extremes and Integrated Risk Management*. London: Risk Waters Group, pp. 59–67.

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

[75] Falk, M., J. Hüsler and R.-D. Reiss (2004). *Laws of Small Numbers: Extremes and Rare Events*. Basel: Birkhäuser.

[76] Fama, E. F. (1963). "Mandelbrot and the stable Paretian hypothesis". *Journal of Business* 35:420–29.

[77] Fama, E. F. (1965). "The behavior of stock market prices". *The Journal of Business* 38:34–105.

[78] Fernandez, V. (2003). "Extreme value theory and Value-at-Risk". *Revista de Analisis Economico* 18(1):57–85.

[79] Ferro, C. A. T., and J. Segers (2003). "Inference for clusters of extreme values". *J. R. Statist. Soc.* B 65(2):545–556.

[80] Finkenstädt, B., and H. Rootzén (2004). *Extreme Values in Finance, Telecommunications and the Environment*. Chapman & hall/CRC.

[81] Fisher, R. A., and L. H.C. Tippett (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample". *Proceedings of the Cambridge Philosophical Society* 24:180–90.

[82] French, C. W., "The tail that wags the hedge fund dog". Working Paper.

[83] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Melbourne: Krieger.

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

[85] García, A., and R. Gençay (2006). "Risk-cost frontier and collateral valuation in securities settlement systems for extreme market events". Working Papers 06–17, Bank of Canada.

[86] García, A., and R. Gençay (2007). "Managing adverse dependence for portfolios of collateral in financial infrastructures". Working Papers 07-25, Bank of Canada.

[87] Garita, G., and C. Zhou (2009). "Can open capital markets help avoid currency crises?". DNB Working Papers 205, Netherlands Central Bank, Research Department.

[88] Gawron, G. (2007). *Tail risk of hedge funds: An extreme value application*. Göttingen: Cuvillier.

[89] Gençay, R., and F. Selçuk (2004). "Extreme value theory and Value-at-Risk: Relative performance in emerging markets". *International Journal of Forecasting* 20:287–303.

[90] Gençay, R., and F. Selçuk (2006). "Overnight borrowing, interest rates and extreme value theory". *European Economic Review* 50:547–563.

[91] Gençay, R., F. Selçuk and A. Ulugülyağci (2003). "High volatility, thick tails and extreme value theory in value-at-risk estimation". *Insurance: Mathematics and Economics* 33:337–356.

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

[93] Gnedenko, B. V. (1943). "Sur la distribution limite du terme maximum d'une serie aleatoire". *Annals of Mathematics* 44:423–53.

[94] Goldberg, L.R., G. Miller and J. Weinstein (2008). "Beyond Value-at-Risk: Forecasting portfolio loss at multiple horizons". *Journal of Investment Management* 6(2):1–26.

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

[96] Gupta, A., and B. Liang (2005). "Do hedge funds have enough capital? A value-at-risk approach". *Journal of Financial Economics* 77:219–253.

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

[98] Haan, L. de, and S. I. Resnick (1980). "A simple asymptotic estimate for the index of a stable distribution". *Journal of the Royal Statistical Society B* 42:83–87.

[99] Haan, L. de, S. I. Resnick, H. Rootzén and C. G. de Vries (1989). "Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes".

[100] Haile, F. D., and S. Pozo (2006). "Exchange rate regimes and currency crises: An evaluation using extreme value theory". *Review of International Economics* 14(4):554–570.

[101] Haile, F. D., and S. Pozo (2008). "Currency crisis contagion and the identification of transmission channels". *International Review of Economics and Finance* 17:572–588.

[102] 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.

[103] Hartmann, P., S. Straetmans and C. de Vries (2004). "Asset market linkages in crisis periods". *The Review of Economics and Statistics* 86(1):313–326.

[104] Hartmann, P., S. Straetmans and C. de Vries (2005). "Banking system stability: A cross-Atlantic perspective". Working Paper Series 527, European Central Bank.

[105] Hauksson, H. A., M. Dacorogna, T. Domenig, U. Müller and G. Samorodnitsky (2001). "Multivariate extremes, aggregation and risk estimation". *Quantitative Finance* 1:79–95.

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

[107] Ho, L.-C., P. Burridge, J. Cadle and M. Theobald (2000). "Value-at-risk: Applying the extreme value approach to Asian markets in the recent financial turmoil". *Pacific-Basin Finance Journal* 8:249–275.

[108] Ho, T.-K. (2008). "Extremal analysis of currency crises in Taiwan". *Applied Economics* 40:1175–1186.

[109] Holton, G. A. (2003). *Value-at-risk: Theory and practice*. San Diego: Academic Press.

[110] Hosking, J. R. M., J. R. Wallis and E. F. Wood (1985). "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments". *Technometrics* 27:251–261.

[111] Hsieh, P.-H. (2001). "Modelling the frequency and severity of extreme exchange rate returns". *J. Forecast.* 20:485–499.

[112] Huisman, R., K. G. Koedijk, C. J. M. Kool and F. Palm (2001). "Tail-index estimates in small samples". *Journal of Business & Economic Statistics* 19(2):208–216.

[113] Jalal, A., and M. Rockinger (2008). "Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data". *Journal of Empirical Finance* 15:868–877.

[114] Jansen, D. W., K. G. Koedijk and C. G. de Vries (2000). "Portfolio selection with limited downside risk". *Journal of Empirical Finance* 7:247–269.

[115] Jansen, D. W., and C. G. de Vries (1991). "On the frequency of large stock returns: Putting booms and busts into perspective". *The Review of Economics and Statistics* 73(1):18–24.

[116] Jenkinson, A. F. (1955). "The frequency distribution of the annual maximum (or minimum) values of meteorological elements". *Quarterly Journal of the Royal Meteorology Society* 87:145–58.

[117] Jobst, A. (2007). "Operational risk – The sting is still in the tail but the poison depends on the dose". IMF Working Papers 07/239, International Monetary Fund.

[118] Jondeau, E., and M. Rockinger (2003). "Testing for differences in the tails of stock-market returns". *Journal of Empirical Finance* 10:559–581.

[119] Jorion, P. (1997). *Value-at-Risk: The new benchmark for controlling market risk.* Chicago: McGraw-Hill.

[120] Juri, A., and M. V. Wüthrich (2002). "Copula convergence theorems for tail events". *Insurance: Mathematics and Economics* 30(3): 405–420.

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

[122] Koedijk, K. G., M. M. A. Schafgans and C. G. de Vries (1990). "The tail index of exchange rate returns". *Journal of International Economics* 29:93–108.

[123] Koedijk, K. G., P. A. Stork and C. G. de Vries (1992). "Differences between foreign exchange rate regimes: the view from the tails". *Journal of International Money und Finance* 11:462–473.

[124] Koh, W., and D. Paxson (2007). "Real extreme R&D discovery options". Bankers Markets & Investors 86.

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

[126] Krehbiel, T., and L. C. Adkins (2005). "Price risk in the NYMEX energy complex: An extreme value approach". *The Journal of Futures Markets* 25(4):30–9337.

[127] Krehbiel, T., and L. C. Adkins (2008). "Extreme daily changes in U.S. Dollar London inter-bank offer rates". *International Review of Economics and Finance* 17:397–411.

[128] Kuester, K., S. Mittnik and M. S. Paolella (2006). "Value-at-Risk prediction: A comparison of alternative strategies". *Journal of Financial Econometrics* 4(1):53–89.

[129] Kupiec, P. H. (1995). "Techniques for verifying the accuracy of risk measurement models". *Journal of Derivatives* 3:73–84.

[130] Leadbetter, M. R., G. Lindgren and H. Rootzén (1983). *Extremes and Relates Properties of Random Sequences and Processes.* Berlin: Springer.

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

[132] Ledford, A. W., and J. A. Tawn (2003). "Diagnostics for dependence within time series extremes". *J. R. Statist. Soc.* B 65(2):521–543.

[133] Lee, T.-H., and B. Saltoğlu (2002). "Assessing the risk forecasts for Japanese stock market". *Japan and the World Economy* 14:63–85.

[134] Lestano and J. P. A. M. Jacobs (2007). "Dating currency crises with ad hoc and extreme value-based thresholds: East Asia 1970–2002 [Dating currency crises]". *Int. J. Fin. Econ.* 12:371–388.

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

[136] Longin, F. M. (2000). "From Value-at-Risk to stress testing: The extreme value approach". *Journal of Banking & Finance* 24:1097–1130.

[137] Longin, F. M. (2005). "The choice of the distribution of asset returns: How extreme value theory can help?". *Journal of Banking & Finance* 29:1017–1035.

[138] Longin, F.M., and B. Solnik (1997). "Correlation structure of international equity markets during extremely volatile periods". Working Paper 97-039, ESSEC, Cergy-Pontoise, France.

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

[140] Loretan, M., and P. Phillips (1994). "Testing the covariance structure of heavy-tailed time series". *Journal of Empirical Finance* 1:211–248.

[141] Lux, T. (1996). "The stable Paretian hypothesis and the frequency of large returns: An examination of major German stocks". *Applied Financial Economics* 6:463–475.

[142] Lux, T. (2000). "On moment condition failure in German stock returns: An application of recent advances in extreme value statistics". *Empirical Economics* 25:641–652.

[143] Lux, T. (2001). "The limiting extremal behaviour of speculative returns: An analysis of intra-daily data from the Frankfurt Stock Exchange". *Applied Financial Economics* 11:299–315.

[144] Mancini, L., and F. Trojani (2007). "Robust Value-at-Risk prediction". Swiss Finance Institute Research Paper Series 07–31.

[145] Mandelbrot, B. B. (1963). "The variation of certain speculative prices". *Journal of Business* 36:394–419.

[146] Mandelbrot, B. B. (1967). "The variation of some other speculative prices". *Journal of Business* 40:393–413.

[147] Manganelli, S., and R. Engle (2004). "A comparison of Value-at-Risk models in finance". In G. Szegö (ed.). *Risk Measures for the 21st Century*. New York: Wiley.

[148] Marimoutou, V., B. Raggad and A. Trabelsi (2009). "Extreme value theory and Value-at-Risk: Application to oil market". *Energy Economics* 31:519–530.

[149] Markose, S. M., and A. Alentorn (2005). "The generalized extreme value (GEV) distribution, implied tail index and option pricing". Discussion Papers, University of Essex, Department of Economics.

[150] Markowitz, H. (1952). "Portfolio selection". *Journal of Finance* 7:77–91.

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

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

[153] Mikosch, T., and C. Stărică (2000). "Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process". *Annals of Statistics* 28:1427–1451.

[154] Moscadelli, M. (2004). "The modelling of operational risk: Experience with the analysis of the data collected by the Basel Committee". Temi di discussione del Servizio Studi n. 517, Banca d'Italia.

[155] Neftci, S. N. (2000). "Value-at-Risk calculations, extreme events, and tail estimation". *Journal of Derivatives* 7(3):23–38.

[156] Nielsson, U. (2009). "Measuring and regulating extreme risk". *Journal of Financial Regulation and Compliance* 17(2): 156–171.

[157] Novak, S. Y., and J. Beirlant (2006). "The magnitude of a market crash can be predicted". *Journal of Banking & Finance* 30:453–462.

[158] Pais, A., and P. A. Stork (2009). "Contagion risk across Real Estate firms and banks: An Australian and New Zealand analysis". Working Paper.

[159] Panaretos, J., and Z. Tsourti (2003). "Extreme value index estimators and smoothing alternatives: A critical review". MPRA Paper 6390, University Library of Munich, Germany.

[160] Pearson, N. D., and C. Smithson (2002). "VaR: The state of play". *Review of Financial Economics* 11:175–189.

[161] Pickands, J. III (1975). "Statistical inference using extreme order statistics". *Ann. Statist.* 3:119–131.

[162] Pictet, O. V., M. M. Dacorogna and U. A. Müller (1996). "Hill, bootstrap and jackknife estimators for heavy tails". Working Paper 1996–12–10, Olsen and Associates.

[163] Pontines, V., and R. Siregar (2007). "The Yen, the US dollar, and the trade weighted basket of currencies: Does the choice of anchor currencies matter in identifying incidences of speculative attacks?". *Japan and the World Economy* 19:214–235

[164] Pontines, V., and R. Siregar (2008). "Fundamental pitfalls of exchange market pressure-based approaches to identification of currency crises". *International Review of Economics and Finance* 17: 345–365.

[165] Poon, S.-H., M. Rockinger and J. Tawn (2004). "Extreme value dependence in financial markets: diagnostics, models, and financial implications". *The Review of Financial Studies* 17(2):581–610.

[166] Pownall, R. A. J., and K. G. Koedijk (1999). "Capturing downside risk in financial markets: the case of the Asian Crisis". *Journal of International Money and Finance* 18:853–870.

[167] Pozo, S., and C. Amuedo-Dorantes (2003). "Statistical distributions and the identification of currency crises". *Journal of International Money and Finance* 22:591–609.

[168] Quintos, C., Z. Fan and P. C. B. Phillips (2001). "Structural change tests in tail behaviour and the Asian crisis". *The Review of Economic Studies* 68(3):633–663.

[169] Rachev, S. T. (ed.) (2003). *Handbook of Heavy Tailed Distributions in Finance*. Amsterdam: Elsevier.

[170] Rachev, S. T., S. V. Stoyanov and F. J. Fabozzi (2008). *Advanced Stochastic Models, Risk Assessment and Portfolio Optimization: The Ideal Risk, Uncertainty and Performance Measures*. Hoboken (New Jersey): Wiley.

[171] Reiss, R.-D., M. Thomas (1997). *Statistical Analysis of Extreme Values with Applications in Finance, Hydrology and Other Fields*. Basle: Birkhäuser.

[172] Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Process*. New York: Springer.

[173] Resnick, S. I. (2004). "On the foundations of multivariate heavy-tail analysis". *Journal of Applied Probability* 41:191–212.

[174] Resnick, S. I. (2007). *Heavy-Tail Phenomena*. New York: Springer.

[175] Roy, A. D. (1952). "Safety first and the holding of assets". *Econometrica* 20:431–449.

[176] Serguieva, A., S. Bozhkov and K. Yu (2009). "Catastrophic and operational risk measurement for financial institutions". Preprint.

[177] Smith, R. L. (1985). "Maximum likelihood estimation in a class of non-regular cases". *Biometrika* 72:67–90.

[178] Smith, R. L. (1987). "Estimating tails of probability distributions". *Annals of Statistics* 15:1174–1207.

[179] Stărică, C. (1999). "Multivariate extremes for models with constant conditional correlations". *Journal of Empirical Finance* 6:515–553.

[180] Straetmans, S., and R. Versteeg (2009). "The effect of capital controls on exchange rate risk". Working Paper.

[181] Sundmacher, M., and G. Ford (2007). "When statistics fail: Extreme events in financial institutions". Working Paper.

[182] Susmel, R. (2001). "Extreme observations and diversification in Latin American emerging equity markets". *Journal of International Money and Finance* 20:971–986.

[183] Tolikas, K., A. Koulakiotis and R. A. Brown (2007). "Extreme risk and Value-at-Risk in the German stock market". *The European Journal of Finance* 13(4):373–395.

[184] Tsay, R. S. (2005). *Analysis of Financial Time Series*. Hoboken (New Jersey): Wiley-Interscience.

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

[186] Wagner, N. (2003). "Estimating financial risk under time-varying extremal return behavior". *OR Spectrum* 25:317–328.

[187] Wagner, N. (2005). "Autoregressive conditional tail behavior and results on Government bond yield spreads". *International Review of Financial Analysis* 14(2):247–261.

[188] Werner, T., and C. Upper (2004). "Time variation in the tail behavior of bund future returns". *The Journal of Futures Markets* 24(4):3873–98.

[189] White, H. (2000). "A reality check for data snooping". *Econometrica* 68(5):1097–1126.

[190] Yamai, Y., and T. Yoshiba (2005). "Value-at-risk versus Expected Shortfall: A practical perspective". *Journal of Banking & Finance* 29:997–1015.

[191] Zhou, C. (2009). "Dependence structure of risk factors and diversification effects". DNB Working Papers 219, Netherlands Central Bank, Research Department.