
Modelling Catastrophic Loss Events Using Extreme Value Theory - Fundamental Results and Applications in Finance and Insurance*

Dr. John W. Dalle Molle, Visiting Fellow, Department of Finance, Banking and Property,
Massey University, Private Bag 11 222, Palmerston North, New Zealand 5301
Email - J.W.DalleMolle@massey.ac.nz

Date of Document - 6 November 2002

* Paper prepared for the 2002 New Zealand Society of Actuaries Conference, 14 November, 2002, Rotorua, New Zealand.

Table of Contents

1	Introduction	3
2	An Overview of Modeling with Extreme Value Theory.....	4
1.1	Overview of the Limitations Traditional Risk Measurement Approaches	4
2.1	An Overview of the Applicability of EVT in Finance and Insurance.....	7
3	Overview of Extreme Value Theory	8
4	Defining Heavy-Tailed Distribution	10
4.1	Summary of Heavy-Tailed Distributions	10
4.2	Defining Fat-Tailed Distributions	10
4.3	Power-law Distributions.....	11
4.4	Generating Heavy-Tails from Ratios of Independent Random Variables	12
4.5	Fat Tails and Returns Distributions.....	13
4.6	Fat-Tailed Distributions and Risk Measurement	14
5	Extreme Value Theory - Background	15
5.1	Overview of Extreme Value Theory Literature	15
5.2	Historical EVT Methodological Development Time Line.....	15
5.3	A Landmark EVT Application.....	18
6	Generalized Extreme Value and Generalized Pareto Distributions	18
6.1	Defining an Extreme Value and Tail Distributions.....	18
6.2	Generalized Extreme Value Distributions.....	19
6.3	Summary of Extreme Value Distributions	20
6.4	Generalized Pareto Distributions	22
6.5	Relationship between Extreme Value and Generalized Pareto Distributions	23
7	Excess Distribution Function and the Generalized Pareto Distribution.....	24
7.1	The Distribution of Exceedances	25
7.2	Limiting Distributions of Exceedances	27
8	Fundamental Results of Extreme Value Theory	28
8.1	Fundamental Mathematical EVT Results	28
8.2	Additional Probabilistic EVT Tools.....	30
8.3	Statistical Models and Inferences Using EVT	31
8.4	Summary of the Mathematical Applications of Extreme Value Theory.....	32
8.5	Summary of the Statistical Applications of Extreme Value Theory.....	32
8.6	Distributional Models Based on Extreme Value Theory	33

9	Modeling High Quantiles and the Tail Index Parameter.....	35
9.1	Modeling of the High Quantiles using Tail Index Estimator and Regular Variation Assumption for the Tail Behavior.....	35
9.2	The Generalized Pareto Distribution Functions, the Peaks-Over-Threshold Method and the Modeling of Exceedances.....	38
9.3	Applicability of the GPD to the Modeling of Exceedances.....	38
9.4	Overview of the Peaks-Over-Threshold Method.....	39
9.5	The Peaks-Over-Threshold Method.....	40
10	Explanatory Data Analysis and Extreme Value Theory.....	43
11	Estimation of the Tail Index Parameter.....	44
11.1	Overview of the Estimation of the Tail Index Parameter.....	44
11.2	EVT-based Techniques and the Estimation of High Quantiles.....	45
11.3	Overview of Tail Index Estimation Procedures.....	48
11.4	Tail Index Estimation using Maximum Likelihood Methods.....	49
11.5	Linear Regression and Tail Index Estimation.....	50
11.6	Non-parametric Tail Index Estimators.....	51
11.7	Comparison of Maximum Likelihood Estimator with Non-Parametric Estimators.....	52
12	Applications of Extreme Value Theory.....	52
12.1	Extreme Value Theory and Value-at-Risk.....	53
12.1.1	Extreme Events and Value-at-Risk.....	53
12.1.2	The EVT-based Value-at-Risk Extension and the Expected Shortfall.....	55
12.2	Extreme Events, Extreme Value Distributions and Tails of Financial Time Series.....	57
12.3	EVT, Risk Management and Risk Measurement.....	61
12.4	EVT Applications in Actuarial Science and Insurance.....	64
13	Issues Inherent to Extreme Value Theory.....	69
13.1	Difficulties Inherent to the Application of Extreme Value Theory.....	70
13.2	Advantages to the Use of EVT in Applications.....	71
13.3	Drawbacks to the Use of EVT in Applications.....	73
13.4	Summarizing the Advantages and Drawbacks of EVT.....	74
14	References.....	74
15	Figures.....	82

1 Introduction

Extreme event risk affects all aspects of risk assessment, modeling and management, especially in the contexts of credit, insurance market, and operational risks. Over the last twenty-five years there have been an increasingly large number of extreme events in the financial and insurance markets including market crashes and natural disasters that have led to extremely large losses and claims. Examples of such events include Barings, Metallgesellschaft, Long-Term Capital Management, and Orange County financial derivatives debacles, and Hurricane Andrew, the Northridge earthquake, the Kobe earthquake, and the September 11th terrorist's attacks on New York City and Washington DC in the insurance arena. Traditional statistical methods that focus on the center of the loss or claims distribution tend to severely underestimate the likelihood and magnitude of such events. *Extreme value theory (EVT)* or *extreme analysis* is a statistical methodology that is well suited for assessing, modeling and managing such catastrophic events. Traditional statistical methods focus on the probability laws governing averages of sums, whereas *EVT* concerns the properties of the largest (smallest) observation in a sample and the probability laws governing these extremal values. *EVT* can also be used to model the upper and/or lower tails of the underlying loss or claims distribution functions. *EVT* offers a framework for assessing the uncertainties inherent to describing the loss or claims distributions associated with these rare or extreme events.

Many *EVT* procedures were developed in the context of modeling hydrological extreme events starting in early 1950s. Since that time *EVT* has found frequent usage in the areas of financial risk management and insurance. Earlier investigations using *EVT* in insurance occurred in the beginning of the 1970s, and concerned the modeling of fat-tailed non-life insurance claim sizes. Early *EVT* applications in financial risk management occurred during late 1990s, and were extensions of Value-at-Risk (*VaR*) models, where *VaR* is a statistical estimate of a portfolio loss, such that there is a given (small) probability, that a loss of that size or greater will occur over a given holding period.

This paper presents overviews of the fundamental results of *EVT*, the implementation of *EVT* procedures and the applications in finance and insurance. The paper begins with definitions of thin, normal and heavy-tailed distributions, and a discussion of the differences between the tail and extreme value distributions. Next is a discussion of the limiting distributional families for extremal phenomena and the observations at high quantiles from the tail regions of a distribution. These families of distributions include the generalized extreme value (*GEV*) distributions, which are used in the modeling of the

limiting behavior of the maximum and minimum, and the generalized Pareto family of distributions, which are related to the *GEV* family and under certain conditions, can be used in approximating for the conditional distribution of exceedances over a given high threshold. Then there is a summary of some of the fundamental results of *EVT* including the convergence-of-types theorem, the maximum domains of attraction theorem and the importance of the tail index parameter, which defines the shape of the tail region. From this follows a section on *EVT* explanatory data analysis including sampling procedures for both the extremal values and the conditional excesses over a given high threshold. Next is a section on the modeling of *EVT* distributions, which requires the selection of a high threshold and the estimation of the tail index parameter. Parametric and nonparametric estimation procedures and goodness-of fit tests for the tail index parameter are detailed in this section. The two basic approaches to the determination of the tail index parameter, which are the peaks over thresholds method and direct estimation of the tail index parameter, are then covered. A description of the methods for determining the choice of the high threshold, which determines where the tail region begins, is given. These methods include the use of quantile-quantile ($Q-Q$) plots and plots of the estimates of the mean excess function or the tail index versus the value of the threshold. Before numerous *EVT* applications are presented, the advantages and limitation of the *EVT* in practical applications will be summarized. Finally *EVT* modeling applications of financial loss and insurance claims distributions will be overviewed. These include the following financial applications: the modeling of returns and volatility in commodities, equity and foreign exchange markets; interest rates; futures margins; extensions of value-at-risk, stress testing and scenario analysis; capital adequacy requirements; and credit, market, and operational risk management. The applications in the insurance markets that will be discussed include the modeling of ruin probabilities, claim size and claims times for heavy-tailed damage claim distributions arising in the aftermath of extremely severe earthquake, fire, flood, and windstorm events and the modeling of catastrophic reinsurance treaties.

2 An Overview of Modeling with Extreme Value Theory

1.1 Overview of the Limitations Traditional Risk Measurement Approaches

It can be said that the first generation *VaR* models have been discredited by the end of the 1990s. The comments of the Bank for International Settlement after they had interviewed a number of market practitioners during 1999 (see page 182 of DeRosa (2001)):

“There was a range of responses among the interviewees as to whether the magnitude of mature market turbulence was within or above their VaR limits. A large majority of the interviewees admitted that last autumn’s [1998] events were in the “tails” of distribution and that therefore their VaR models were useless for measuring and monitoring market risk. On the other hand, other respondents (though a very small number) judged that their VaR models remained adequate ... The deficiencies of the VaR methodology in conditions of contagion were unanimously emphasized by market participants. The surge in VaR levels above predefined limits during the crisis compelled market participants to unwind positions in the assets for which VaR limits were exceeded. Because of the widespread use of similar models, similar behavior was adopted by numerous investors. The resulting simultaneous pressure to unwind positions dried up the liquidity of the markets and therefore exacerbated price volatility.”

David DeRosa, a part-time academic and international financial markets trader added the following comments (see pages 182-183 of DeRosa (2001)):

“Clearly the world has a pressing need for a robust new model that can produce dependable estimates of portfolio risk. What is also clear is that simplistic models of risk that depend naively on historical estimates of variance and correlation are flawed in this application. What was missing in the first generation of risk models was an application of the nature of the risk associated with macroeconomic dislocations.”

Traditional parametric statistical and econometric techniques are based on the estimation of the entire distribution usually focusing on the center region of the distribution. These techniques are not designed to model extreme quantiles and tail event probabilities. Estimates of quantiles and probabilities are required not only from near the boundary of the range of the observed data, but also at points outside the range in many risk assessment and management applications. *EVT* techniques are designed specifically to estimate the extreme quantiles and their corresponding probabilities through the use of models that fit the empirical survival function of a set of observations using only the samples from the extreme regions of the range of the data as opposed to using the entire data set.

“Ninety percent of the events cause ten percent of the losses, and the last ten percent of the events cause ninety percent of the losses”

This is a variation on the ninety percent rule of project scheduling (see Section 2 of Rootzén and Klüppelberg (May 27, 1999)).

More often than not, catastrophic financial events happen that were considered extremely unlikely, but possible events. These are rare, but extreme events and usually cannot be handled using standard risk management techniques. This underlies the basis for the use of extreme value theory in risk management since even very unlikely events will happen, given enough time. That is, sooner or later an event will occur that causes a loss exceeding any past loss. These assertions can be stated mathematically using *Lévy's zero-one law*, which can be stated in a risk management context as follows:

“If the likelihood of a daily exposure to an extremely catastrophic event is greater than zero, i.e. even less than 0.1 percent, then this event will eventually happen.”

Lévy's zero-one law basically implies that sooner or later large loss events exceeding the magnitude of any past realized loss event will occur. Since it is inevitable that an extreme loss will eventually happen, it would be useful to understand its consequences and costs. The *compound Poisson limit theorem* can be used to provide guidelines to describe the consequences, which can be stated in a risk management context as follows:

“Let there be a p % chance that on a given day, a large loss be realized the next day. Also assume that large losses occur in small clusters of mean length L . Then the likelihood that at least one large loss occurs during a period of T days is given by

$$\text{Pr ob(That there is a large loss in a period of } T \text{ days)} = 1 - \exp\left(-\frac{Tp}{L}\right).”$$

Another mathematical aspect of modeling extreme losses is the observation of the *Law of Large Numbers*, which can be stated as follows:

“The observed averages from an experiment that is repeated a large number of times will be close to its population average.”

This implies that the mean values are not particularly useful in the modeling of extreme risk exposures. In the context of extreme losses, this is not what happens since once an extreme loss occurs it is usually not repeated.

2.1 An Overview of the Applicability of EVT in Finance and Insurance

The branch of statistics referred to as extreme value theory (*EVT*) can be used to address the problem of modeling large rare market movements. *EVT* provides a family of natural models for extremal phenomena, i.e. the assessment of catastrophic or extreme events. *EVT* concerns the analysis of the extreme events from the tail of a distribution and their associated tail probabilities. It uses statistical laws that concern rare events whose sample frequencies are extremely small. *EVT* is a methodology that can be used to characterize the rare, but not impossible, occurrences of extreme changes. *EVT* offers a natural model for defining a worst case that can be tolerated no matter how infrequent is the occurrence of the scenario.

EVT allows for the reconciliation in one modeling approach consideration of normal events and extreme events. Extreme value theory (*EVT*) or extreme analysis is concerned with the properties of the maximum (or the minimum) in large samples, which also allows for the determination of the shape of the tails of a distribution. *EVT* can be thought of as a natural complement to the central limit theorem: The former addresses the properties of the sample fluctuations of normalized and centered cumulative sums, whereas the latter addresses the properties of the sample fluctuations of normalized and centered maxima. That is, *EVT* can be viewed as a law of large numbers for extremes as opposed to a limiting law of scaled and centered random sums. Extreme value distributions represent the possible limit distributions of centered and scaled maxima (and minima) of *IID* random variables. That is, according to the theory of extreme values, the largest or smallest value from a collection of *IID* random variables tends to an asymptotic distribution that only depends on the tail of the distribution of the underlying (or basic) random variable.

EVT offers a framework for assessing the uncertainty that surrounds rare or extreme market events such as market crashes. *EVT* is one well-established methodology that can be used to predict the likelihood of extreme market events based on historical prices. Conventional confidence (or tolerance) interval techniques may be suitable for the estimation of 90, 95, and possibly even the 99 percent levels of thin-tailed distributions such as the normal or Gaussian distribution, which is the basis for more risk measurement techniques. The Gaussian distribution, which performs reasonably well for conventional profit/loss symmetric distributions with normal tails, will usually underestimate at the higher confidence levels for the potential losses when there are heavy-tails present. *EVT* is specifically designed to address the problem of modeling extreme market moves by using well-established statistical laws about rare events that allow for the likelihood of

events that are extremely rare. *EVT* attempts to fit density functions to historical price series that exhibit a wide variety of shapes, and it places greater weight of the tails than the normal distribution. *EVT* can be thought as a law of large numbers for the extremes rather than the average of a distribution. *EVT* forms the basis of procedures that are applicable in non-normal extreme market situations. It allows the risk manager to implement risk management models that include rare but damaging events and offers procedures to measure the risk inherent to their consequences (see Embrechts (2000A), Embrechts (2000B), Embrechts (2001), McNeil (1998), McNeil (2000) and Smith (2000)).

EVT is also well suited to estimate the potential extreme scenarios for the risk factors for stress testing. *VaR* is a high quantile on the distribution of losses and provides an approximate upper bounded for a loss that is exceeded only a small fraction of the time. The *EVT* extension of *VaR* is very useful when the tail probabilities of the extreme events are required in the risk measurement process. The *EVT* approach is based on the assumption that the extreme values of the profit/loss distribution are what the firm is most concerned about. Then, the statistical theory of *EVT* is used to model the expected extreme loss above a specified high quantile that may be outside the range of the observed returns data. It must be noted that *EVT* extrapolation is only as good the given sample information.

3 Overview of Extreme Value Theory

Traditional statistical methods usually concern the probability laws governing averages of sums, whereas extreme value theory (*EVT*) concerns the properties of the largest (smallest) observation in a sample and the probability laws governing these extremal values (see Section 1.5 of Beirlant, Teugels, and Vyncker (1996)). An *extremal value* or *extreme event* can be either the largest or smallest observation in a finite sample. *EVT* is based on the stochastic behavior of the maximum and minimum of *IID* random variables from some underlying distribution function. The largest and the smallest values in a sample, and other statistics related to them are generally named *extreme statistics*. *EVT* can be used to model the upper and lower tails of an underlying distribution function. That is, *EVT* models specifically focus on the tail regions including the distributional properties of the following extreme statistics:

- 1) The extremes, i.e. the maximum and minimum of a sample.
- 2) The extreme and intermediate order statistics

3) The exceedances over (below) high (low) thresholds.

Conversely, the tail regions of the underlying distribution function, or functional parameter related to it, may be evaluated by means of statistical procedures that are based on extreme and intermediate order statistics or exceedances over a given threshold.

The following remarks of Alan Greenspan indicate the recent interest in modeling extreme statistics and events:

“Work that characterizes the statistical distribution of extreme events would be useful, as well.”

These are the remarks that Alan Greenspan, chairman of the Federal Reserve Board of Governors made at a Research Conference on Risk Management and Systemic Risk in Washington, DC on November 16, 1996. Extreme value theory is one possible methodology that implements the wishes of Mr. Greenspan and will be discussed in detail. Some of the major concerns of risk managers give way to the following needs:

- 1) Widely accepted methods supported by regulatory authorities and other officials, for selecting risk factor models and performing reliable assessment of risk management systems.
- 2) Quantifying the uncertainties, i.e. model risk, associated with the lack of data and the extrapolation methods used to model upper tail region of the loss distribution beyond the maximal sample observation.
- 3) Avoiding extremely costly financial disasters or other systemic risk exposures arising due the use of inappropriate models.
- 4) The availability of new and more reliable estimates of extremely rare events.
- 5) Making the new advances in univariate and multivariate extreme value theory accessible in a comprehensible form to risk managers.

Traditional parametric statistical and econometric techniques are based on the estimation of the entire distribution and usually focus on the center of the distribution. These techniques are ill-suited for modeling extreme quantiles and tail event probabilities. For example, the unsatisfactory performance of a portfolio is often associated with catastrophic losses due to extreme events such as the collapse of a large bank or systemic problems associated with an unexpected and volatile currency devaluation. Non-parametric density estimation

techniques such as kernel smoothing perform poorly with respect to modeling the tail regions of the distributions. Extreme value theory and the corresponding three maximum domains of attractions distributions for the largest and smallest values are a well suited methodology for the analysis of extreme events.

4 Defining Heavy-Tailed Distribution

4.1 Summary of Heavy-Tailed Distributions

The *tails* or *tail regions* of a probability density distribution are the *extreme values* located far from the mean value. If a probability distribution exhibits a relatively large mass in its extreme ends, it is said to have “*fat tails*” or “*heavy tails*” or “*broad tails*.” Fat tails are “*fatter*” than normal tails and have unbounded second moments. Thin tails are “*thinner*” than normal tails and are bounded in the sense that they have zero probability density above (below) some extreme threshold level.

4.2 Defining Fat-Tailed Distributions

A distribution function $F_X(x)$ of the random variable X is said to have a “*heavy tail*” (or “*thick tail*” or “*fat tail*”) if its moment generating function is such that $M_X(s) = \infty$ for $s > 0$. A distribution function $F_X(x)$ is said to have a “*light tail*” if it has an exponentially bounded tail, where an exponentially bounded tail is defined for constants $a, b > 0$ such that

$$\bar{F}_X(x) < a \exp(-bx) \quad (4.2.1)$$

for all $x > 0$ and $\bar{F}_X(x) = 1 - F_X(x)$, which is the tail function. Common examples of heavy-tail distributions include the lognormal, Pareto and Weibull distributions, all with shape parameters less than one (See Sections 2.2.3 and 2.5 of Rolski, Schmidli, Schmidt, and Teugels (1999)).

Depending on the various properties of the distributions, there are three possible types of tails for returns distribution:

- 1) *Normal-tailed distribution* for which moments of all orders exist, implying that the distribution function has tails that decline at an exponential rate. The Gumbel distribution is the limiting distributions for normal-tailed distributions.

2) *Fat-tailed distributions*, which are unstable in the sense that not all moments exist, implying that the distribution function has tails that decline at a polynomial rate. The Fréchet distribution is the limiting distributions for fat-tailed distributions.

3) *Thin-tailed distributions*, which can be truncated, i.e. fixed endpoint(s), and such that not all moments exist. The Weibull distribution is the limiting distributions for thin-tailed distributions.

Koedijk, Schafgans and de Vries (1990) use the term “*tail-fatness*” to refer to the probability weight in the tails of a distribution and the term “*fat-tailed*” distribution to imply a distribution that has fatter tails than the normal distribution. Tail fatness can be measure using the tail index parameter and fat-tailed can be assessed using the excess kurtosis measure.

4.3 Power-law Distributions

Power law distributions, which can be used to describe fat-tailed distributions, will be discussed in this section. The power-law distribution function is the cornerstone of univariate *EVT*. Specifically it is required for most of the statistical procedures used in the estimation of the tail event probability, extreme quantiles and upper order statistics. The Paris-born Italian social economist *Vilfredo Pareto* (1848-1923) was the first to study power-law distributions, and this is why they are commonly referred to as Pareto-type distributions. Pareto investigated the statistical characteristics of the wealth of individuals in a stable economy.

Define $\bar{F}_x(X > x)$ as the *survival function* or *survivor function* or *tail probability function* or *tail function* of distribution function $F_x(X < x)$ of the random variable X :

$$\begin{aligned}\bar{F}_x(X > x) &= \text{Pr ob}(X > x) \\ &= 1 - F_x(X < x)\end{aligned}\tag{4.3.1}$$

If the distribution function F_x is in the domain of attraction of the Fréchet distribution, which will be defined in a section (6.3), then \bar{F}_x is a power-law function times a slowly varying function $k(x)$:

$$\text{Pr ob}(X > x) = k(x)x^{-\alpha}\tag{4.3.2}$$

for all $x > 0$ and where $\alpha > 0$ is a constants that must be estimated. α is referred to as the *tail index* or *tail exponent* parameter. A slow varying function is defined by the following relationship

$$\lim_{t \rightarrow \infty} \frac{k(tx)}{k(t)} = 1 \quad (4.3.3)$$

Note that the power law assumption is assumed to hold only in the extreme tail region. The power law distribution for $k(x) = k$ where k is constant (to be estimated) is defined as

$$\text{Pr ob}(X > x) = k x^{-\alpha}. \quad (4.3.4)$$

The term Pareto-type tail implies that the tail region of a distribution tends to zero at a polynomial rate:

$$1 - F_x(x) = x^{-\alpha} \rightarrow 0 \quad (4.3.5)$$

as $\alpha \rightarrow \infty$. Usually the random variable X is written in standard form, i.e. adjusted for location shift b and a scale transformation a :

$$\text{Pr ob}((aX + b) > x) = k x^{-\alpha}. \quad (4.3.6)$$

Typically the choice $b = 0$ is used since for heavy-tailed distributions the maxima will approach infinity at a fast enough rate that corrections for location do not really assist in obtaining a non-degenerate limit distribution. The power-law distribution is counterintuitive since it may lack a characteristics scale in that its variance and other moments may not be bounded (see section 2.1 of Beirlant, Teugels, and Vyncker (1996), page 52 of Diebold, Schuermann and Stroughair (2000) and pages 171-173 of Starica (2000)).

4.4 Generating Heavy-Tails from Ratios of Independent Random Variables

The following ratio can generate fat-, thick- or heavy-tailed distributions:

$$Y = \frac{X_1}{X_2} \quad (4.4.1)$$

where X_1 and X_2 are independent random variables. The right tail of the skewed distribution described by Y is much “*thicker*” than the tails of either of the original

independent random variables X_1 and X_2 . Mixtures of centered normal or Gaussian distributions can be constructed such that they exhibit leptokurtic behavior. If X_1 and X_2 are both normally distributed, then Y has a Cauchy distribution.

4.5 Fat Tails and Returns Distributions

The log-returns distributions of financial times are typically leptokurtic with heavy and highly non-normal tails, where the log-returns are defined as follows: Let $\{r_1, r_2, \dots, r_n\}$ denote the realized log-returns of the collection of prices $\{S_0, S_1, S_2, \dots, S_n\}$

$$r_t = \ln(S_t / S_{t-1}) = \ln(S_t) - \ln(S_{t-1}) \quad (4.5.1)$$

for $t = 1, 2, \dots, n$. That is, the tails of financial returns distributions typically exhibit fat-tails or in other words, a large number of large price changes. This implies that when the returns are standardized through scaling, i.e. normalizing by the volatility, exhibit more mass in the tails than the standard Gaussian distribution. More mass in the tails implies that very large and very small values of the observed time series $\{S_t\}$ occur more often than in a Gaussian white noise sequence. This means that the distribution of the log-returns is heavy-tailed, i.e. the tail probabilities $\text{Prob}(S_t < -z_q)$ and $\text{Prob}(S_t > z_q)$ of the observed log-returns series are greater than the tails at the same high quantile z_q of the standard Gaussian distribution.

The distribution of the log-returns process is fat-tailed, implying it is also non-normal. In a risk management context, this characteristic indicates that the losses and gains can be much more severe than those anticipated by the Gaussian distribution. Heavy tails implies that the tail region or the excess probability as function of the corresponding high quantile can be approximately described using a power law function. This implies that it makes sense to focus the analysis on the tail region of the distribution instead of the center of the distribution. Extreme value theory is useful in discriminating between alternative probability models for stock-market returns. Some recent financial markets examples of so-called fat-tailed events include:

- 1) The 1987 Stock market crash
- 2) The bond market collapse of 1958
- 3) The bond market collapse of 1994
- 4) The sharp fluctuations in the Japanese yen/US dollar exchange rate in mid-1995

- 5) The aftermath of the Mexican Crisis starting in late December 1994 and into 1995
- 6) The events surrounding the Southeast Asian crisis starting in July 1997 and lasting until the end of the year.
- 7) The events surrounding the Russian debt moratorium in August 1998, leading into the subsequent difficulties and bailout of Long-Term Capital Management.

4.6 Fat-Tailed Distributions and Risk Measurement

Historically, most risk measurement techniques assumed that log-returns of the individual components of the portfolio were jointly normally distributed, where the log-returns are the natural logarithm of the price relatives. The normal distribution for log-returns is a more than suitable theoretical choice since it has very tractable properties and it can be completely described by its mean and variance. Unfortunately, the normality assumption has been continually criticized as inadequate in empirical studies beginning with Benoit Mandelbrot in 1963 when he showed empirically that the distribution of asset returns had fatter or heavier tails than the normal distribution (see Fama (1963) and Mandelbrot (1963)). The presence of fat tails causes problems for the normal distribution since the probability of an extreme large log-returns of three standard deviations and greater for the normal distribution is extremely close to zero, which contradicts the empirical evidence on the log-returns of financial time series. Basically, fat tails implies that the tail region of a fat-tailed distribution approaches zero at a slower rate than the normal distributions. Specifically fat-tailed distributions approach zero at polynomial rates whereas the tails of the normal distribution approach zero at an exponential rate. See Figure 1 for an illustration of two log-returns distributions such that one distribution exhibits fatter tails than the other. Fat-tailed distributions can be generated using the ratio $R_t = S_t/S_{t-1}$, which is a price relative and the $r_t = \ln(R_t)$ is the commonly used log-returns measure; where S_t is the price of an asset at time t . The right tail of the skewed distribution described by R_t is much “fatter” than the tails of S_t . This inability of standard risk measurement techniques based on the normal distribution to account for the empirically-observed stylized fact of fat tails has led researchers to investigate alternatives to the normal distribution theory.

Fat-tailed returns distributions cause additional complexities in the risk measurement process in the sense of more parameters to estimate and the absence of large number of observations from the tail regions. The challenge in modeling extreme losses is that they are very infrequent and usually not observable. When there are significant deviations from the measure of central tendency, which could be either the mode or the mean loss, the

modeling of the losses becomes very difficult and in many cases unfeasible. Fat tails cause a small change in the tolerance level to result in a significant change in the upper bound of the losses. Also the issue of where the tail region begins must be addressed. For the risk manager, the presence of fat-tails implies that the extreme losses occur more frequently than would if the returns distribution were normal. This implies that a risk measurement such as a *VaR* forecast, which based on fat-tailed returns distributions such as a student-t distribution with a small number of degrees of freedom, will be greater than a *VaR* forecast based on a normality assumption. The degrees of freedom for a student-t distribution correspond to the number of observations. For example, a *VaR* forecast at the 99 percent level for a Student-t distribution returns distribution with 5 degrees of freedom is equal to 3.365 standard deviations, compared to 2.326 standard deviations for a normal distribution.

5 Extreme Value Theory - Background

5.1 Overview of Extreme Value Theory Literature

Reasonably accessible references to the theory and practice of *EVT* include: Beirlant, Teugels, and Vyncker (1996) where the focus is on nonparametric tail index estimation; Castillo (1988), where the focus is on the *EVT* properties of extreme order statistics and high quantiles; De Haan (1994) where the focus is more statistical; Embrechts, Kluppelberg, and Mikosch (1997), which is a little more advanced text from the viewpoint of actuarial mathematicians, but it is a leading text for quantitative risk managers; Reiss and Thomas (2000), which is a more practical text with numerous applications in insurance, finance, hydrology and other engineering fields.

The investigations of the extreme returns of equity markets by Jansen and De Vries (1991), Longin (1996) and Loretan and Phillips (1994) were some of the earliest financial investigations using *EVT*. Early investigations of *EVT* in risk measurement include Longin (1996) and Bassi, Embrechts, and Kafetzaki (1998). They considered determining capital requirements based on high quantiles on the profit/loss distributions, which is related to *VaR* forecasting.

5.2 Historical EVT Methodological Development Time Line

The first investigations into the area of extreme value theory were by statisticians. Then the hydrologists came along in the early 1970s, actuaries in the 1980s, financial time series analysts in the mid-1990s, and finally the quantitative risk manager in the late 1990s. Some of the important historical events in extreme value theory are given below:

- 1) The standard Pareto distribution is named after the Paris-born Italian, social economist *Vilfredo Pareto* (1848-1923). Pareto was the first to study power-law distributions, and this is why these distributions are commonly referred to as Pareto-type distributions. Pareto investigated the statistical characteristics of the wealth of individuals in a stable economy.
- 2) Early pioneering investigations concerning extreme value theory were that of Fréchet (1927) and Fisher and Tippett (1928) and concerned the limiting behavior of the distribution functions of the sample maxima and sample minima. Fisher and Tippett (1928) were the first to heuristically derive the fundamental three types of extreme value limit laws and concerns limit laws for maxima. The Fisher-Tippett theorem describes the limiting behavior of the sample distribution of the maxima, and is analogous to the central limit theorem for the sample distribution of the normalized sums.
- 3) The *Weibull distribution* is named after the Swedish physicist *Waloddi Weibull* who first proposed this distribution in 1939 in the context of investigations of the strength of materials (See Weibull (1939)).
- 3) von Mises (1936) first proposed combining the three types of *EVT* distributions into a single representation called the *generalized extreme value distribution*.
- 4) The sufficient conditions for a distribution to belong to the maximum domain of attraction of an extreme value distribution were originally given in Von Mises (1936) and are frequently called the *Von Mises conditions*.
- 5) Gnedenko (1943) was the first to derive the fundamental three limiting types of extreme value laws from a rigorous mathematical standpoint. The necessary and sufficient conditions that a collection of random variable belongs to each of the three domains of attraction were first given by Gnedenko (1943) and then in a simpler form by De Haan (1976).
 - 5.1) Gnedenko (1943) showed that if the tail regions of distribution functions decay like power function, then the distribution function belongs to the domain of attraction of the Fréchet distribution.
- 6) The applicability of extreme value theory was extensively developed by Gumbel (1958)
 - 6.1) The Gumbel distribution was first proposed by Gumbel (see Gumbel (1958)).

- 6.2) The block maxima or per-period method is the oldest of the models of classical extreme value analysis, and was proposed in Gumbel (1958). It is also referred to as Gumbel's method of exceedances.
- 6.3) The distributions of the maxima and minima are related by the principle of symmetry property that was introduced by Gumbel (1958). The principle of symmetry implies that after some technical details are taken care of, the results for the maxima can be modified to apply to the minima.
- 6.4) Gumbel (1958) also presents the details of the statistical estimation procedures and gives numerous illustrative examples of *EVT* applications in science and engineering.
- 7) The γ -parameterization of the extreme value distributions is due to von Mises (1936), and also is frequently attributed to Jenkinson (1955).
- 8) The *general extreme value distribution (GEV)* with three degrees of freedom was introduced in Jenkinson (1955) and Jenkinson (1969) to identify the frequency distributions of the largest values of meteorological data when the limiting form of the extreme value distribution is unknown.
- 9) The *excess distribution to types theorem* or the *Pickands-Balkema-de-Haan theorem*, which relates the limiting form of the excess distribution function to the generalized Pareto distribution (*GPD*) function for high thresholds was first shown in Balkema and de Haan (1974) and Pickands (1975). They showed that the limiting expression for the scaled exceedances (or excesses) over a high threshold u of the random variable X could be approximated by the GPD function.
- 10) The Peaks-over-Threshold (*POT*) method has been used by hydrologists for over 25 years. See Davison and Smith (1990), section 6.5 of Embrechts, Kluppelberg, and Mikosch (1997), and Rootzén and Tajvidi (2000) for a discussion of its development.
- 11) The first applications in risk management were the contributions of actuarial mathematicians Andrew J. McNeil and Paul Embrechts and the colleagues mainly at the Swiss University ETH in Zurich (For example; see Embrechts (2000A), Embrechts (2000B), Embrechts (2001), McNeil (1998), and McNeil (2000)).

5.3 A Landmark EVT Application

Embrechts (2001) discusses the severe westerly storm that battered the Flemish and Dutch coastal areas on February 1, 1953 where more than 1,800 people died in the Dutch Provinces of Holland and Zeeland after several dykes collapsed. The flooding represented a surge of 3.85 meters above normal sea level (or N.A.P. + 3.85 meters), where normal sea level in the Netherlands is typically measured in (N.A.P. + x) meters, where N.A.P. refers to Normaal Amsterdams Peil, which is the Dutch reference level corresponding to mean sea level and x is the excess above N.A.P. . This surge came close to the N.A.P. + 4 meter surge of November 1, 1570, which is known as the “*Allerheiligenvloed.*” In the aftermath of the 1953 disaster, the Dutch government asked a group of mathematicians and scientists under the supervision of van Dantzig to address the problem of what is the optimal minimal height for new dykes in order that they are able to withstand and once-in-10,000 year event, i.e. the 0.0001 quantile on the annual maximum distribution. The van Dantzig report estimated that the dykes should be at least N.A.P. + 5.14 meters, and their statistical analysis was based on extreme value theory (see De Haan (1990), Embrechts (2001), Example 6.1.1 on pages 284-285, and Example 6.1.8 on pages 288-289 of Embrechts, Kluppelberg, and Mikosch (1997) for additional discussions of this applications).

6 Generalized Extreme Value and Generalized Pareto Distributions

6.1 Defining an Extreme Value and Tail Distributions

Let $\{X_1, X_2, \dots, X_n\}$ denote a set of *IID* random variables from a population having a common continuous distribution function $F_X(x)$ defined on unbounded support where x indicates an observed value of X and n is the number of equispaced data points observed within a given period of time. Also let $\{x_{(j)}: j = 1, 2, \dots, n\}$ denote the order values of the observed sample $\{X_1, X_2, \dots, X_n\}$ defined as the following rearrangement in ascending order:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n-1)} \leq x_{(n)}. \quad (6.1.1)$$

The distribution function of $x_{\max} = x_{(n)} = \max(x_1, x_2, \dots, x_{n-1}, x_n)$ is given by

$$F_{x_{\max}}(x) = (F_X(x))^n, \quad (6.1.2)$$

and the distribution function of $X_{\min} = X_{(1)} = \min(x_1, x_2, \dots, x_{n-1}, x_n)$ is given by

$$F_{X_{\min}}(x) = 1 - (1 - F_X(x))^n. \quad (6.1.3)$$

There is a distinction between the tail distribution and the extreme value distribution: Assume there is a sample $\{x_1, x_2, \dots, x_n\}$ drawn from some distribution defining a *IID* random variable. Define $M_n = X_{\max} = \max(x_1, x_2, \dots, x_n)$. The asymptotic distribution of X_{\max} as $n \rightarrow \infty$ is called the *extreme distribution*. Asymptotically, the *tail distribution* represents the distribution of the upper tails of an extremal distribution.

6.2 Generalized Extreme Value Distributions

Extreme value distributions represent the possible limit distributions of centered and scaled maxima (and minima) of *IID* random variables. There are three different extreme value *GEV* distributions for continuous distributions:

- 1) The Gumbel Distribution
- 2) The Fréchet Distribution
- 3) The Weibull Distribution

The *general extreme value (GEV) distribution* with three degrees of freedom was introduced in Jenkinson (1955) and Jenkinson (1969) to identify the frequency distributions of the largest values of meteorological data when the limiting form of the extreme value distribution is unknown. It is also referred to as the Jenkinson-von-Mises representation of the extreme value distributions. The general extreme value distribution is the generalization of the three extreme value distributions (up to an affine transformation). The distribution and density functions of the generalized extreme value distribution functions can be collectively summarized as follows (See section 3.4 of Embrechts, Kluppelberg, and Mikosch (1997):

$$F_{GEV}(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}) & \text{if } \gamma \neq 0 \\ \exp(-\exp(-x)) & \text{if } \gamma = 0 \end{cases} \quad (6.2.1)$$

and

$$f_{GEV}(x) = \begin{cases} \exp\left(-\left(1 + \gamma x\right)^{-1/\gamma}\right) \left(1 + \gamma x\right)^{-(1+1/\gamma)} & \text{if } \gamma \neq 0 \\ \exp(-x - \exp(-x)) & \text{if } \gamma = 0 \end{cases} \quad (6.2.2)$$

respectively, for $1 + \gamma x > 0$. The shape parameter $\alpha = 1/\gamma$ determines the shape of the tail region of the distribution, and in turn the type of extreme value distribution:

- 1) The Type I extreme value or Gumbel distribution is the limiting case as $\gamma = 1/\alpha \rightarrow 0$. These distributions have tails that decline exponentially and all their moments exist.
- 2) The Type II extreme value or Fréchet distribution is the case where $\gamma = 1/\alpha > 0$. These distributions have tails that decline as a power function (or polynomially) and only moments of orders up to α exist.
- 3) The Type III extreme value or Weibull distribution is the case where $\gamma = 1/\alpha < 0$. These distributions do not have tails since they have a finite end point.

6.3 Summary of Extreme Value Distributions

These three distributions can be described using the following general expression, which is referred to as the *GEV* distribution function:

$$F_{GEV}(y) = \begin{cases} \exp\left(-\left(1 - \frac{y}{\alpha}\right)^{-\alpha}\right) & \text{if } \alpha \neq 0 \\ \exp(-\exp(-y)) & \text{if } \alpha = 0 \end{cases} \quad (6.3.1)$$

for $1 + \gamma y > 0$, and where $y = (x - \mu)/\sigma$, μ is a location parameter, $\alpha = 1/\gamma$ is a shape parameter or Pareto index and $\sigma > 0$ is a scale parameter (see section 6.4.4 and pages 158-160 of Embrechts, Kluppelberg, and Mikosch (1997) and Chapter 1 of Reiss and Thomas (2000)).

The relationships for three different extreme value distributions for continuous distributions are given as follows:

1) The Gumbel Distribution

If the tail of the original distribution function is *unbounded* and it decreases at least as rapidly as an exponential function, the asymptotic extreme value distribution is a Type I extreme value distribution (or a Gumbel distribution or the double exponential distribution). The Gumbel distribution function is also known as the Fisher-Tippett

Type I distribution and is denoted as EV_0 , F_{EV_0} , G_0 or Λ . The Gumbel distribution function is defined as follows:

$$F_{EV_0}(y) = \exp(-\lambda \exp(-y)) \quad (6.3.2)$$

for $-\infty < y < \infty$ and where $y = (x - \mu)/\sigma$. Examples of Type I asymptotic distributional forms include: the exponential, gamma, normal, lognormal, logistic, and the Gumbel distributions. Distributions with this type of tail decay are referred to as *light-tailed distributions*.

2) The Fréchet Distribution

If the upper tail of the original distribution function is unbounded, but not all of its moments are finite, and it decreases at a power law rate; the asymptotic extreme value distribution is a Type II extreme value distribution (or a Fréchet distribution). The Fréchet distribution function is also known as the Fisher-Tippett Type II distribution and can be denoted as EV_1 , F_{EV_1} , $G_{1,\alpha}$ or Φ_α . The Fréchet distribution function is defined as follows:

$$F_{EV_{2,\alpha}}(y) = \begin{cases} \exp(-\lambda y^{-\alpha}) & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases} \quad (6.3.3)$$

where $\alpha > 0$ denote a positive constant and where $y = (x - \mu)/\sigma$. Examples of the Type II asymptotic distributional form include the Pareto, Burr, Student's t, Cauchy, Log-gamma, and Fréchet distributions. Also various mixture models are of the Type II asymptotic distributional form. Distributions with this type of tail decay are referred to as *heavy-tailed distributions* in the sense that not all of their moments are finite. This implies that if the underlying distribution is fat-tailed; the Fisher-Tippett Type II distribution is the only general family that it can belong to, in the limit. The Fréchet distribution function is the appropriate form for most financial time series.

3) The Weibull Distribution

If the original distribution function has a bounded tail, but not all of its moments are finite, the asymptotic extreme value distribution is a Type III extreme value distribution (or Weibull distribution). The Weibull distribution function is also known as the Fisher-Tippett Type III distribution and can be denoted as EV_2 , F_{EV_2} , $G_{2,\alpha}$ or Ψ_α . The Weibull distribution is defined as follows:

$$F_{EV3,\alpha}(y) = \begin{cases} \exp(-\lambda(-y)^\alpha) & \text{if } y < 0 \\ 1 & \text{if } y \geq 0 \end{cases} \quad (6.3.4)$$

where $\alpha > 0$ denotes a positive constant and where $y = (x - \mu)/\sigma$. Examples of Type III asymptotic distributional forms include the reflected-power, uniform, beta, and Weibull distributions. Distributions with this type of tail decay are referred to as *short-tailed distributions*.

6.4 Generalized Pareto Distributions

The generalized Pareto distribution (*GPD*) function will be discussed in this section. Another important *EVT* result is the *excess distribution to types theorem* or the *Pickands-Balkema-de-Haan theorem*, which relates the limiting form of the excess distribution function to the *GPD* function for high thresholds (see Balkema and de Haan (1974) and Pickands (1975)). This theorem suggests that the *GPD* is the appropriate approximation for the excess distribution function F_x'' for a sufficiently high threshold u . The *GPD* is the only non-degenerate distribution that approximates the distribution of exceedances $X - u$ over a sufficiently high threshold u as the threshold approaches the upper endpoint of the underlying distribution F_x . This result is useful in the estimation of the tails of excess distributions F_x'' since it offers knowledge of the actual limiting distributions that govern the extreme values, i.e. the distribution of the excesses above a given threshold. This results arises from the modeling of excesses over a given threshold, the probability of an excess over a given threshold can be modeled using a Bernoulli random variable with parameter p . Using the law of small numbers, the sums of the Bernoulli random value, i.e. a binomial random value, can be approximated by the Poisson random variable, which in turn can be approximated by the generalized Pareto distribution if the threshold is high enough and the probability of exceeding the threshold is very close to zero.

The three limiting distributions in the *GPD* family include the Pareto, uniform, and standard exponential distribution functions. The Generalized Pareto distribution (*GPD*) is given by the following expression:

$$F_{GPD}(x) = \begin{cases} 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \text{if } \gamma = 0 \end{cases} \quad (6.4.1)$$

and the density function of the *GPD* is given by

$$f_{GPD}(x) = \begin{cases} \sigma^{-1/\gamma} (\sigma + \gamma x)^{-(1+1/\gamma)} & \text{if } \gamma \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right) & \text{if } \gamma = 0 \end{cases} \quad (6.4.2)$$

where $1 + \gamma x/\sigma > 0$, and $\sigma > 0$. $\alpha = 1/\gamma$ is the tail index, and is the same tail index from the generalized extreme value distributions (see section 3.4 of Embrechts, Kluppelberg, and Mikosch (1997) and Chapter 1 of Reiss and Thomas (2000)).

The following analytical relationship exists between the *GPD* functions F_{GPD} and the generalized extreme value (*GEV*) distribution functions $F_{GEV}(x)$ for γ -parameterization:

$$F_{GPD}(x) = 1 + \ln(F_{GEV}(x)) \quad (6.4.3)$$

where $\alpha = 1/\gamma$. These distributions are said to be generalized in the sense that it subsumes certain other distributions under a common parametric form. The three limiting distributions in the *GPD* family include the Pareto, uniform, and standard exponential distribution functions:

- 1) The exponential distribution function is defined for $x \geq 0$ and $\alpha = 1/\gamma \rightarrow 0$, and corresponds to the Gumbel distribution.
- 2) The Pareto distribution function is defined for $x \geq 1$ and $\alpha = 1/\gamma > 0$, and corresponds to the Fréchet distribution.
- 3) The Beta distribution functions is defined for $-1 \leq x \leq 0$ and $\alpha = 1/\gamma < 0$, and corresponds to the Weibull distribution.

6.5 Relationship between Extreme Value and Generalized Pareto Distributions

The relationships for three different *GPDs* for continuous distributions are given as follows (see Chapter 1 of Reiss and Thomas (2000)):

1) The Exponential Distribution

The exponential distribution function corresponds to the Gumbel distribution. The exponential distribution is denoted as *GPO*. The exponential distribution function corresponds to the Gumbel distribution as follows:

$$\begin{aligned}
F_{GP0}(y) &= 1 + \ln(F_{EV0}(y)) \\
&= 1 - \exp(-y)
\end{aligned} \tag{6.5.1}$$

for $y \geq 0$ and where $y = (x - \mu)/\sigma$.

2) The Pareto Distribution

The Pareto (or ordinary Pareto) distribution function corresponds to the Fréchet distribution. The Pareto distribution is denoted as $GP1, \alpha$. The Pareto distribution has a long history as in actuarial mathematics as a model for large losses. The Pareto distribution is the most relevant of the GPD functions for risk management applications since it is heavy-tailed. The heavy-tailed Pareto distribution does possess a complete set of bounded moments (check this). The *Pareto (or ordinary Pareto) distribution function* corresponds to the Fréchet distribution as follows:

$$\begin{aligned}
F_{GP1, \alpha}(y) &= 1 + \ln(F_{EV1, \alpha}(y)) \\
&= 1 - y^{-\alpha}
\end{aligned} \tag{6.5.2}$$

for $y \geq 1$, and where $\alpha > 0$ and $y = (x - \mu)/\sigma$.

3) The Beta Distribution

The Beta distribution function corresponds to the Weibull distribution function. The Beta distribution is denoted as $GP2, \alpha$. The beta distribution $F_{GP2, \alpha}(x)$ belonging to the generalized Pareto distributions constitute a subclass of the usual family of beta distributions. The *Beta distribution function* corresponds to the Weibull distribution function as follows

$$\begin{aligned}
F_{GP2, \alpha}(y) &= 1 + \ln(F_{EV2, \alpha}(y)) \\
&= 1 - (-y)^{-\alpha}
\end{aligned} \tag{6.5.3}$$

for $-1 \leq y \leq 0$, and $\alpha < 0$, and where $y = (x - \mu)/\sigma$.

7 Excess Distribution Function and the Generalized Pareto Distribution

When information from the entire tail region of the returns distribution is required and not just the distribution of a particular extremal point, then the tail region can be modeled using the conditional excess distribution function (see Balkema and de Haan (1974) and Pickands

(1975)). The conditional excess distribution function describes the conditional distribution of the exceedances (or excesses) over a given threshold level. The *GPD* is a flexible family of distributions applicable as an approximation for the conditional distribution of scaled exceedances. The excesses over high thresholds can be modeled by *GPDs*. The *GPDs* are used as a natural approximation for conditional excess distribution function for very high thresholds. This approximation is used in the peaks-over-threshold method and in the conditional value-at-risk forecasts. The distribution of the maximum of a Poisson distributed number of *IID* excesses over a high threshold is a *GEV* distribution function. The appropriateness of the *GPD* approximation requires that the threshold level must be taken sufficiently high enough that it is some sense, approaching the endpoint of the distribution. See Figure 2 for an illustration of a log-returns distribution where the notion of a model for a tail loss region and the tail threshold point are explicitly indicated on the left tail.

7.1 The Distribution of Exceedances

In many situations, the risk manager may not be interested in the maximum loss (or minimum gain), but may be interested in the loss events associated with exceedances of certain threshold losses. This situation may arise when the risk manager knows the level of loss that they can withstand, and is interested in the frequency of exceedance of that level. This allows the risk manager to concern themselves with the frequencies of occurrences instead of the value of the random variable itself. Note that the exceedances above a given threshold contain information about the extreme events, but their use avoids the loss of information that occurs when only uses the extreme events. The method of exceedances consists of fitting all the exceedances of the chosen threshold to a particular family of distributions. A common practice is to backcast the K largest values of a sample of size n , and then refer to the smallest value among the K largest values as the threshold (see section 1.6.1 of Castillo (1994), section 5.2.2 of Castillo (1994), Davison and Smith (1990), Chapter 4 and pages 319-321 of Embrechts, Kluppelberg, and Mikosch (1997) and Chapter 1 of Reiss and Thomas (1997)).

In the rest of this section the probability distribution of the value of the exceedances will be examined. Let $\{X_i; i = 1, 2, \dots, n\}$ be a given set of *IID* random variables with common parent distribution function $F_X(x) = \Pr ob(X \leq x)$ and denote u to be a threshold that is smaller than which is the right end point $s(F_X) = \sup\{x: F_X(x) < 1\}$ of the support of F_X . A threshold u is said to be an upper or high threshold if it is close to the right endpoint $s(F_X)$ and this implies $p = 1 - F_X(u)$ is small and the number of exceedances

over the threshold u , denoted K may be regarded as a Poisson random variable. Exceedances are realized conditioned on the event that an observation is larger than the threshold u . For a given data set $\{x_i; i=1,2,\dots,n\}$, let y_j denote an exceedance by for $j=1,2,\dots,K$ where $y_j = x_i$ for those i such that $x_i > u$ where u is a predetermined upper threshold, and K is the number of exceedances over the threshold u in the sample of size n . The values $v_j = y_j - u = x_i - u$ are the *excesses over u* can be defined such that $v_i > 0$. The *excess distribution function of u* or *exceedance distribution function of u* , denoted $F_Y(y) = F_X^{(u)}(y) = F_X^{(u)}(x)$ is the following conditional distribution function:

$$\begin{aligned} F_X^{(u)}(y) &= \text{Pr ob}(X \leq x | X > u) \\ &= \frac{F_X(y) - F_X(u)}{1 - F_X(u)} \end{aligned} \quad (7.1.1)$$

for $y = x = v + u$ such that $x \geq u$. $F_X^{(u)}(y)$ is a censored distribution function. This expression can be rewritten in terms as *excess V over u distribution function*:

$$\begin{aligned} F_X^{(u)}(v) &= \text{Pr ob}(X - u \leq v | X > u) \\ &= \frac{F_X(u+v) - F_X(u)}{1 - F_X(u)} \end{aligned} \quad (7.1.2)$$

for $v = x - u$, or in terms of the underlying distribution F_X :

$$\begin{aligned} F_X(u+v) &= F_X^{(u)}(y)(1 - F_X(u)) + F_X(u) \\ &= 1 - (1 - F_X(u))(1 - F_X(u+v)), \end{aligned} \quad (7.1.3)$$

which is an expression that gives a representation of the probability of realizing an observation below a given threshold u . The exceedance distribution function $F_X^{(u)}(x)$ at u is the conditional probability of the excesses. The threshold u is equal to the left endpoint $i(F_X^{(u)}) = \inf(x: F_X^{(u)}(x) > 0)$ of the support of the truncated distribution function $F_X^{(u)}(x)$. Generalized Pareto distribution functions can be fitted to the exceedance distribution function $F_X^{(u)}(x)$ for sufficiently high thresholds.

7.2 Limiting Distributions of Exceedances

This limiting expression for the scaled exceedances (or excesses) over a high threshold u of the random variable X can be approximated as follows:

$$\lim_{u \uparrow x_u} \text{Pr ob} \left(\frac{X-u}{a(u)} \leq x \mid X > u \right) = \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1 - \exp(-x) & \text{if } \gamma = 0 \end{cases}, \quad (7.2.1)$$

where $x \geq 0$ if $\gamma \geq 0$ and $0 \leq x \leq -a(u)/\gamma$ if $\gamma < 0$. The distribution functions of the right hand side is the generalized Pareto distribution where γ is the tail index and $a(u) = \sigma$ is the scale factor. This result is due to Balkema and de Haan (1974) and Pickands (1975).

If the exceedance distribution function $F_X^{(u)}(b_n + a_n x)$ has a continuous limiting distribution functions as u approaches the right end point $s(F_X)$ of the support of $F_X(x)$:

$$\left| F_X^{(u)}(b_n + a_n x) - F_{GP,\gamma} \left(\frac{x-u}{\sigma_u} \right) \right| \rightarrow 0 \quad (7.2.2)$$

as $u \rightarrow s(F_X)$ for a suitable choice of constants for the shape, location and scale parameters γ , b_n and $a_n > 0$, $\sigma_u > 0$ is a constant, $F_{GP,\gamma}(x)$ is a γ -parameterized *GPD* function and $s(F_X) = \sup\{x: F_X(x) < 1\}$ is the right end point of the support of $F_X(x)$. Note that the exceedance distribution function $F_X^{(u)}$ and the approximating general Pareto distribution function $F_{GP,\alpha}((x-u)/\sigma_u)$ have the same left endpoint. If eq. (7.2.2) holds, then $F_X(x)$ belongs to the *POT*-domain of attraction of $F_{GP,\gamma}(x)$. This limit theorem can also be formulated in terms of a α -parameterization of the generalized Pareto distribution function $F_{GP,\alpha}(x)$ (See Chapter 1 of Reiss and Thomas (2000)).

It can be shown that every extreme value distribution function $F_{EV,\gamma}(x)$ belongs to the *POT*-domain of attraction of a general Pareto distribution function $F_{GP,\gamma}(x)$. The method to show this is to first verify the following relationship

$$\left| \bar{F}_{EV,\gamma}(x) - \bar{F}_{GP,\gamma}(x) \right| \approx O\left(\bar{F}_{GP,\gamma}^2(x)\right) \quad (7.2.3)$$

for the survivor functions $\bar{F}_{EV,\gamma}(x)$ and $\bar{F}_{GP,\gamma}(x)$ using a Taylor series expansion of the exponential function about zero. Then the following expression can be deduced:

$$\left| \overline{F}_{EV,\gamma}^{(u)}(x) - \overline{F}_{GP,\gamma}^{(u)}(x) \right| \approx O\left(\overline{F}_{GP,\gamma}^{(u)}(u)\right). \quad (7.2.4)$$

Eq. (7.2.3) follows from the γ -parameterization of the standard generalized Pareto distribution with $\mu < u$

$$F_{GP,\alpha}^{(u)}\left(\frac{x-\mu}{\sigma_u}\right) = F_{GP,\alpha}\left(\frac{x-u}{\sigma_u + \gamma(u-\mu)}\right) \quad (7.2.5)$$

whereby $\sigma_u = 1 + \gamma u$.

The distributions $F_{GP0}^n(x + \ln(n))$ and $F_{GP,\alpha}^n(a_n x)$ where $a_n = n^{1/\gamma}$ are the extreme value distribution functions $F_{GP0}(x)$ and $F_{GP,\alpha}(x)$ in the limit as $n \rightarrow \infty$ and through the use of Eq. (7.2.3).

8 Fundamental Results of Extreme Value Theory

EVT is concerned with the properties of the maximum (or the minimum) in large samples, which then allows for the determination of the shape of the tails of a distribution. *EVT* can “conveniently” be thought of as a complement to the central limit theorem: The former addresses the properties of the sample fluctuations of normalized and centered cumulative sums, whereas the latter addresses the properties of the sample fluctuations of normalized and centered maxima. In this section a summary of the fundamental mathematical result of *EVT*, *EVT*-based models for statistical inference, and the mathematical and statistical applications of *EVT* will be presented.

8.1 Fundamental Mathematical EVT Results

There are three fundamental mathematical results that illustrate the importance of *EVT* in risk management applications:

1) Convergence of Types Theorem

The convergence-of-types theorem describes the conditions that guarantee the convergence of empirical distributions to either of the following types of distributions:

- 1.1) The general extreme value family of distributions, which are used in the modeling of the limiting behavior of the maximum (and/or minimum) of a sequence of random variables. This result is useful in setting position limits and margin levels in risk control.

- 1.2) The general Pareto family of distributions, which can be used to model the tail region of a distribution for some large threshold value. This is useful in the context of calculating the probabilities of extreme samples from the tail of a distribution. This is useful in the context of formulating *VaR* forecasts.

This result was first proved in Fisher and Tippett (1928) for the *IID* case.

2) Maximum Domains of Attraction Theorem

The maximum domains of attraction theorem describes the conditions that guarantee that a tail of a distribution will be in the domain of attraction of a given extreme value distribution, if and only if the theorem of Fisher and Tippett (1928) holds. This result was first proved in Gnedenko (1943).

3) Criterion for Choosing a High Threshold

An important characteristic of an extreme value distribution is the tail index parameter $\alpha = 1/\gamma$. There are two main approaches to modeling *EVT* distributions (see sections 1.5 and 1.6 of Beirlant, Teugels, and Vyncker (1996), and pages 143-147, page 16 of Embrechts, Kluppelberg, and Mikosch (1997):

- 1) The method of exceedances or the peaks over thresholds method.
- 2) Procedures to directly estimate the tail index parameter.

Examples of the value of the tail index include:

- 1) The Student t distribution is equal to the degrees of freedom.
- 2) The tail index is greater than two for an *ARCH* process.
- 3) The tail index is greater than zero and less than two for the sum-stable distributions

To estimate a tail index parameter, a choice must be made on what is the high threshold, which designates where the tail region begins. The criterion for choosing a high threshold describes the conditions that guarantee that a distribution function of excesses above a threshold converges (in the limit) to a one of the general extreme value (or general Pareto families of distributions). The high threshold can be chosen in any of the following ways:

- 3.1) A quantile-quantile plot or $Q-Q$ plot, which is defined such that the quantiles of the empirical or fitted distribution function are plotted against the quantiles of the hypothesized distribution (see pages 143-147 of Cleveland (1985) and Section 1.5 of Beirlant, Teugels, and Vyncker (1996)). For example; a normal probability plot is such that the quantiles of the empirical or fitted distribution function are plotted against the quantiles of the hypothesized normal distribution. Concave or convex departures from a straight line indicate a heavier- or shorter-tailed distribution relative to the normal distribution. See Beirlant, Vynckier and Teugels (1996) for an example of the use of the Pareto quantile plot in the estimation of the tail index parameter.
- 3.2) There are a number of estimators of the tail index. One of the most popular is the Hill estimator (see Hill (1975)). The Hill estimator, which is a ratio moment estimator for the tail index parameter, is the inverse average of the log ratios of the order statistics of the data set. The Hill estimator is used to approximate the tail index or shape index $\alpha = 1/\gamma$ for $\alpha > 0$, i.e. for fat-tailed distributions. There are other tail index estimators including the Pickands estimator (see De Haan and Resnick (1980) or Pickands (1975))
- 3.3) The mean excess function can be implemented as a graphical method of choosing a high threshold (see section 1.6 of Beirlant, Teugels, and Vyncker (1996)). The mean excess function is the sum of the excesses above the threshold divided by the total number of times the threshold was exceeded. That is, the mean excess function is the expected overshoot of a threshold conditioned that a given threshold was exceeded. For the general Pareto family of distributions, the expected value of the exceedances is a linear function of the threshold level. This allows for a criterion for choosing the region above a threshold in the context of tail fitting. A positive slope indicates a heavy-tailed distribution.

8.2 Additional Probabilistic EVT Tools

Other important probabilistic tools used in extreme value theory include (see page 16 of Embrechts, Kluppelberg, and Mikosch (1997)):

- 1) The weak convergence of distributions of random variables such as sums, random sums, and maxima of random variables.

- 2) The weak convergence of sum processes and maximum processes to their limits in an appropriate function space.
- 3) The use of point processes for describing the random distribution of point in space and time with application to extreme value theory.

8.3 Statistical Models and Inferences Using EVT

Some of the modeling areas where *EVT* has been applied include (see page 17 of Embrechts, Kluppelberg, and Mikosch (1997)):

- 1) Fat-tailed distributions
- 2) Time series processes with heavy-tailed innovations
- 3) General asymptotic theory
- 4) Point process theory
- 5) Long-memory theory
- 6) Self-similar processes

The following quantities may serve as variables or indicators that can be used to make statistical inferences about the extreme value or high quantiles of a sample (see page 17 of Embrechts, Kluppelberg, and Mikosch (1997)):

- 1) The distributions of periodic extremes such as annual extremes.
- 2) The distribution of the maximum (or minimum) such as the largest (or smallest) values in a portfolio.
- 3) The return or recurrence period for some rare event.
- 4) The frequency of occurrence of extreme events.
- 5) The frequency of exceedances of a threshold.
- 6) The mean excess over a given threshold.
- 7) The distribution of excesses.

- 8) The distribution of records, which are local maxima (or minima).
- 9) The record times between rare events.
- 10) The time-dependent properties of the records underlying a marked process.

8.4 Summary of the Mathematical Applications of Extreme Value Theory

Extreme value theory provides the necessary tools and techniques to answer the following (partial) list of questions:

- 1) What parametric distribution models have the potential to fit the profit/loss (P/L) distributions realized in risk management problems?
- 2) What are the potential models that can be used to model the conditional excess and shortfall distributions that arise in risk management problems?
- 3) How to choose the threshold values for the boundary values defining the beginning of the extreme profit region and end of the extreme loss region of a P/L distribution?
- 4) Can more complex risk management problems can be solved using extreme value methods such as accounting for time-dependent models, i.e. nonstationarities such as trends and seasonalities, nonlinear relationships between the financial variables, or the use of exogenous economic variables in the P/L process?
- 5) Can the univariate extreme value methods be extended to a multivariate setting? That is instead of modeling portfolio returns as one univariate aggregated entity, can the returns of the individual instruments be model as a multivariate returns process in the portfolio be used in measuring risk.

8.5 Summary of the Statistical Applications of Extreme Value Theory

The statistical applications of extreme value theory include:

- 1) Modeling heavy-tailed distributions or the tail regions of distributions using the generalized extreme value and generalized Pareto distributions. The generalized extreme value distributions are used to model the behavior of the extremal values. The generalized Pareto distributions are used as an approximation for the distribution of the excesses above a given high threshold. Both of these results hold if the extremes and

tail regions are observed over a long period of time, which improves the validity of the asymptotic results.

- 2) Time series with heavy tailed innovations including ARCH/GARCH models. The terms “ARCH” and “GARCH” are typically used to describe the general family of models, which use autoregressive models and “similar” to autoregressive moving average models, respectively, to characterize the phenomena of conditional heteroskedasticity. Conditional heteroskedasticity usually refers to the situations where large changes in the returns tend to be followed by other large changes, and small changes in the returns tend to be followed by other small changes. ARCH and GARCH processes were introduced to explicitly recognize the difference between the unconditional and conditional variances, i.e. long-term and short-term volatility forecasts.
- 3) General asymptotic theory including the convergence-of-types theorem. In a similar manner to the central limit theorem, the distribution for the maximum of a collection of random variables from the same parent distribution will converge to one of the following three types of distributions: the Fréchet, Gumbel, and Weibull distributions.
- 4) Theory of point processes including the point process of exceedances. A *point process* is a random sequence of successive occurrences of point events where a point event is the time instant at which the phenomenon such as a trade of a number of shares in a stock occurs. A point process is a random distribution of points in space.
- 5) Long-memory processes, where long-range correlated is characterized by a lack of a typical temporal scale.
- 6) Self-similarity and fractals. The term “*self-similarity*” is used to characterize an object, which looks the same at all length scales, i.e. it is invariant under a change of magnification. The related term “*fractal*” describes dilation invariance objects that exhibit irregularities at all scales of magnification.

8.6 Distributional Models Based on Extreme Value Theory

Some of the important classes of distributions that can be modeled using extreme value theory include (see page 15 of Embrechts, Kluppelberg, and Mikosch (1997)):

- 1) The subexponential distributions as realistic models for heavy-tailed random variables. The stochastic behavior of a sum of n independent observations from a

subexponential distribution is governed by the largest of these observations. This observation indicates the strong influence that the maximum of a sample has on the total sum derived from the sample.

- 2) The α -stable random variables for $\alpha < 2$ as the limit laws for sums of infinite variance *IID* random variables.
- 3) The Fréchet, Gumbel and Weibull distributions, as limit laws for the maxima of *IID* random variables. These three distributions collectively form the generalized extreme value distribution (*GEVD*).
- 4) The exponential, Pareto, and Beta distributions, as approximations for the conditional excess distribution function for high thresholds. These three distributions collectively form the generalized Pareto distributions (*GPD*).
- 5) The normal distribution as the limit law for sums of *IID* finite variance random variables. The distribution of a sum of properly normalized and centered independent finite variance random variables will converge to a normal distribution. This result holds independent of the original distribution of the random variables forming the sum.
- 6) The Poisson distribution as the limit law of binomial distributions for large value of the probability parameter p , which represent a counting measure for rare events. Historically, the Poisson distribution $P(\lambda)$ appeared as the weak limit of a binomial distribution $BI(n, p, 1, 0)$ for which $np \rightarrow \lambda$ as $n \rightarrow \infty$ and $p \rightarrow 0$.

These distributions and their corresponding range of possible shapes make *EVT* applicable to the modeling of the following classes of financial distributions:

- 1) Profit and loss distributions
- 2) Returns distributions
- 3) Credit-loss distributions
- 4) Operational risk loss distributions
- 5) Insurance claims distributions.

9 Modeling High Quantiles and the Tail Index Parameter

9.1 Modeling of the High Quantiles using Tail Index Estimator and Regular Variation Assumption for the Tail Behavior

In this section, an out-of-sample estimator for the quantile and probability estimators based on the unconditional distribution will be discussed. The unconditional estimators are useful in the modeling of extreme risks in financial markets since it is very difficult, if not impossible, to find counterparties to hedge against very turbulent market states. The results from this section will be applied in latter sections of this paper to extension of Value-at-Risk (*VaR*). The parametric form for the of the univariate exceedance probability for a high quantile x_q :

$$\begin{aligned}\bar{F}_X(x_q) &= \text{Pr ob}(X > x_q) \\ &= 1 - F_X(x_q),\end{aligned}\tag{9.1.1}$$

where $F_X(x)$ is the distribution function of the random variable X , can be approximated using a second-order expansion of $F_X(x)$. The tail probability $\bar{F}_X(x_q) = q$ should be very small in magnitude in order for the justification of the approximations (See section 5.4.3 of Dacorogna, Gençay, Müller, Olsen and Pictet (2001), Danielsson and de Vries (1997), pages 90-91 of Danielsson and de Vries (2000), De Haan and Stadtmuller (1996) and De Haan, Jansen, Koedijk, and De Vries (1994)).

Usually the upper tail distributions are modeled using Pareto-type distributions. The defining characteristics of heavy-tailed or Pareto-type distributions is the property of regular variation at infinity:

$$\lim_{t \rightarrow \infty} \frac{1 - F_X(tx)}{1 - F_X(t)} = x^{-\alpha}\tag{9.1.2}$$

where $x > 0$ and $\alpha > 0$, and no finite upper endpoint. The regular variation property implies that the unconditional moments of X for orders larger than α are unbounded, and that the class of regularly varying distributions is heavy-tailed. This approach is appropriate if the interest is not in specifically modeling $F_X(x)$.

The first order expansion of $F_X(x)$ as $x \rightarrow \infty$ using the regular variation condition, is given by

$$F_x(x) = 1 - a x^{-\alpha} + o(x^{-\alpha}) \quad (9.1.3)$$

as $x \rightarrow \infty$, and where $x > 0$, $\alpha > 0$, and a is a positive constant. This is the usual expression used in the expression for the univariate exceedance probabilities based on fat-tailed regularly varying distributional assumption.

A more general parametric form for the upper tail of $F_x(x)$ can be obtained by using a second-order expansion of $F_x(x)$ as $x \rightarrow \infty$. De Haan and Stadtmuller (1996) show that there are only two non-trivial second-order expansions of $F_x(x)$:

$$F_x(x) = 1 - a x^{-\alpha} (1 + b x^{-\beta}) + o(x^{-\beta}) \quad (9.1.4)$$

as $x \rightarrow \infty$, and where $x > 0$, $\alpha > 0$, $\beta > 0$, and a and b are constants and:

$$F_x(x) = 1 - a x^{-\alpha} (1 + b \ln(x)) + o(\ln(x)) \quad (9.1.5)$$

Under an additional mild condition, the second-order expansion of the density function corresponding to the expansion in eq. (9.1.4) is given by:

$$f_x(x) = a\alpha x^{-(\alpha+1)} + ab(\alpha + \beta) x^{-(\alpha+\beta+1)} + o(x^{-(\alpha+\beta+1)}). \quad (9.1.6)$$

The second-order expansion of the density function in eq. (9.1.6) is applicable to the non-normal sum-stable, Student t, Fréchet, and other Type-II extreme value, i.e. fat-tailed distributions. The sum-stable distributions are symmetric and with characteristic exponents $\alpha \in (1, 2)$. The second-order term in the expansion in eq. (9.1.5) decays at a slower rate than the algebraic rate of the expansion in eq. (9.1.4). Therefore, an estimator for the univariate exceedance probabilities can be developed based on the second-order expansion of the distribution and density function in eqs. (9.1.4) and (9.1.6).

An estimator for the univariate exceedance probabilities based on the fat-tailed feature of regularly varying distributions functions will be given: Let p and q represent two excess probabilities such that $p < 1/n < q$ where n is the sample size, and denote x_p and x_q the associated quantiles of p and q , respectively, such that

$$x_p : 1 - F_x(x_p) = p \text{ and } x_q : 1 - F_x(x_q) = q \quad (9.1.7)$$

Note that in general the following inequality should hold:

$$x_p > \max(X_1, X_2, \dots, X_n) \quad (9.1.8)$$

since $p < 1/n$ and where $\{X_1, X_2, \dots, X_n\}$ is the observed sample of size n . The quantile x_p can be estimated by extrapolating the empirical distribution function $S_x^n(x)$ by means of its regular variation properties. Recall that $S_x^n(x)$ can be defined as follows:

$$S_x^n(x) = \frac{1}{n} \sum_{i \leq n} I(X_i \leq x) \quad (9.1.9)$$

where the summation runs from $i = 1, 2, \dots, n$ and $I(y \leq x)$ is an indicator function:

$$I(y \leq x) = \begin{cases} 1 & \text{if } y \leq x \\ 0 & \text{if } y > x \end{cases} \quad (9.1.10)$$

Using the expression in equation (1) with $\beta > 0$

$$1 - F_x(x) = a x^{-\alpha} (1 + b x^{-\beta}) + o(x^{-\beta}), \quad (9.1.11)$$

the ratio q/p can be expressed in terms of the excess probabilities:

$$\frac{q}{p} = \frac{\left(\frac{x_p}{x_q}\right)^\alpha \left(1 + b x_q^{-\beta}\right) + o(x_q^{-\beta})}{\left(1 + b x_p^{-\beta}\right) + o(x_p^{-\beta})} \quad (9.1.12)$$

Upon further rearrangement, the high quantile x_p can be given in terms of x_q , which is defined within the sample and use of the second-order expansion of $F_x(x)$:

$$x_p = x_q \left(\frac{q}{p}\right)^{1/\alpha} \left(\frac{1 + b x_p^{-\beta} + o(x_p^{-\beta})}{1 + b x_q^{-\beta} + o(x_q^{-\beta})}\right)^{1/\alpha} \quad (9.1.13)$$

Let $\hat{x}(j)$ for $j = 1, 2, \dots, n$ represent the sample order statistics of the observed sample $\{X_1, X_2, \dots, X_n\}$ of size n . Now if the higher order terms in eq. (9.1.13) are ignored, and the following substitutions are made into eq. (9.1.13): x_q is replaced by the sample order statistic $\hat{x}(n - k_q)$ where k_q is the sample rank of $\hat{x}(n - k_q)$ and the sample expression for q is given by $q = k_q/n$, and α is replaced by an estimated value $\hat{\alpha}$, then the estimator for the univariate exceedance probabilities based on fat-tailed regularly varying distributional assumption is given by the following expression:

$$\hat{x}_p = \hat{x}(n - k_q) \left(\frac{k_q}{n\hat{p}}\right)^{1/\hat{\alpha}} \quad (9.1.14)$$

$\hat{\alpha}$ can be determined using any of the univariate semiparametric tail index estimators such as the Hill estimator, which is appropriate for Pareto-type distribution. From the expression for $\hat{\chi}_p$ in eq. (9.1.14), the excess probability $\hat{p} = \hat{F}_x(\hat{x}(n - k_p))$ can be estimated from the following expression:

$$\hat{F}_x(\hat{x}(n - k_p)) = \frac{k_q}{n} \left(\frac{\hat{x}(n - k_q)}{\hat{\chi}_p(k_p)} \right)^{\hat{\alpha}} \quad (9.1.15)$$

9.2 The Generalized Pareto Distribution Functions, the Peaks-Over-Threshold Method and the Modeling of Exceedances

The *peaks-over-threshold (POT) method* is used to model the excessive moves above a threshold u where the threshold is chosen such that a significant fraction of the observations are above the threshold, but a much larger fraction does not exceed it. *POT* is the term that hydrologists use for this procedure, whereas it is referred to as the *point process of exceedances* by mathematicians. Value-at-Risk (*VAR*) is an example of the *POT* method, where the threshold is a high quantile on the loss distribution. Also using the *POT* method, the mean excess function for some high quantile can be used to model the expected shortfall, i.e. given that there is the loss is greater than *VAR*, what is its expected value. The *POT* method can be summarized as follows: For a given data set $\{\chi_i; i=1,2,\dots,n\}$, let y_j denote an exceedance by for $j=1,2,\dots,K$ defined where $y_j = \chi_i$ for those i such that $\chi_i > u$ and where: u is a predetermined upper threshold, K is the number of exceedances over the threshold, and $v_j = y_j - u$ for $j=1,2,\dots,K$ are referred to as the *excesses* over a given threshold u . The *excesses* are also referred to as *magnitude* or *size* of the *exceedances*. The *GPDs* can be used to model the distribution of the *exceedances*. The observations can be taken in the original order of their outcome or in any other ordering.

9.3 Applicability of the GPD to the Modeling of Exceedances

The appropriateness of the *GPD* functions as an approximate model for the exceedances of a given threshold (and the excesses) of a sample of *IID* random variables can be summarized as follows (see Davison and Smith (1990), section 6.5 of Embrechts, Kluppelberg, and Mikosch (1997), and Rootzén and Tajvidi (2000)):

- 1) The probability of exceeding a given threshold can be modeled using Bernoulli random variables. The Bernoulli random variable is used in the first step to modeling each of the exceedances of a given threshold.
- 2) The modeling of a number of exceedances of a threshold can be modeled as a sum of Bernoulli random variables, which is a binomial random variable.
- 3) The Poisson family of distributions is the appropriate family for the approximate modeling of the distribution of the number of exceedances over a given threshold.
- 4) The Poisson distribution can be used to approximate the sums of Bernoulli random variables. That is, the Poisson distribution is an accurate approximation of the excess distribution function if the threshold for the excesses is chosen large enough and the *law of small numbers* is applicable. The applicability of the Poisson distribution is appropriate if the threshold level is extremely high and the associated tail probability is extremely close to zero.
- 5) The Poisson random variable can be approximated by the generalized Pareto distribution if the threshold is high enough and the probability of exceeding the threshold is very close to zero. The excess-distribution-to-types theorem suggests that the *GPD* is an appropriate approximation for the excess distribution function for a sufficiently high threshold.

9.4 Overview of the Peaks-Over-Threshold Method

The peaks-over-threshold (*POT*) method can be used to calculate risk measures such as the value-at-risk and the expected shortfall. Statistically, the main challenge in implementing an estimation procedure for one of these risk measures is obtaining a reasonable estimate of the tails of the profit/loss distribution. These risk measures are straightforward to calculate once a good estimate is obtained. The basic assumptions required for the modeling the time between occurrences and the magnitudes of excesses over a high threshold u of a time series X include (See section 6.5 of Embrechts, Kluppelberg, and Mikosch (1997)):

- 1) The excesses over a high threshold u occur at times governed by a Poisson process with intensity parameter λ . These times are called exceedance times. Note that the Poisson process for the exceedances may not be a homogenous Poisson process, i.e. λ may a function of time $\lambda = \lambda(t)$

- 2) The corresponding magnitudes of the excesses over u , i.e. $X - u$ are independent and have a generalized Pareto distribution function with shape parameter $\alpha = 1/\gamma$ and scale parameter σ .
- 3) The exceedance magnitudes and the exceedance times are independent of each other.

The selection of the optimal threshold is an important technical issue in the implementation of the *POT* method. A good rule to follow in the selection of the optimal threshold is given by Paul Embrechts (see page xvii of Embrechts (2000B)):

“Let the tails speak for themselves!”

but he follows this statement with:

“Easily said, not so easily implemented.”

The main issues in the estimation of optimal threshold are as follows:

- 1) A value of the threshold that is too high results in too few exceedances, and this choice results in too large an estimate of the sample variance.
- 2) A value of the threshold that is too low results in an estimator that is very biased. That is, if the data is too scarce in the tail region, the resulting estimator will have low statistical power.

Theoretically it is possible to choose an asymptotically optimal threshold level by quantifying the bias-versus variance trade-off. There are problems in quantifying this trade-off including the checking of the required underlying assumptions and the dependence inherent to the data.

9.5 The Peaks-Over-Threshold Method

The *peaks-over-threshold (POT) method* is used to make statements about the excessive moves above a threshold u where it is the threshold is chosen such that a significant fraction of time the series is above the threshold but the fraction of the series. Value-at-Risk is an example of the *POT* where the threshold is a given upper quantile and mean excess function for some quantile is used to model the expected shortfall. The *POT* method basically is the fitting of a model to the excesses over a threshold. The *POT* method is used to extract the upper extremes from a data set. *POT* can be viewed as a statistical method for the estimation of tail events. Hydrologists first used the *POT* method. The basis of the *POT* method is that extreme events, i.e. losses, larger than a

given threshold u occur as a Poisson process with intensity parameter λ and that the magnitudes of the excesses, i.e. the amounts by which the losses exceed the threshold, can be fitted to a generalized Pareto distribution function. The normalized time of excesses above a high level u asymptotically as a Poisson process, using the Poisson process limit for a binomial process. The *POT method* tends to be more stable than tail index estimator such as the Hill Estimator with respect to the choice of the high threshold (see McNeil and Frey (2000)).

For a given data set $\{x_i; i = 1, 2, \dots, n\}$, let y_j denote an exceedance by for $j = 1, 2, \dots, K$ where $y_j = x_i$ for those i such that $x_i > u$ where u is a predetermined upper threshold, K is the *number of exceedances over the threshold*, and noting that the observations can be taken in the original order of their outcome or in any other order. The values $v_j = y_j - u$ for $j = 1, 2, \dots, K$ are called the *excesses over u* , i.e. the amounts by which the observations overshoot the threshold. They are also referred to as *magnitude* or *size of the exceedances*.

The *POT* allows for the study of the behavior of subset of the K largest losses out of a total of n losses. Note that in general K is a random variable and can be defined as follows:

$$K = \sum_{i \leq n} I(X_i > u) \quad (9.5.1)$$

for a given set of *IID* random variables $\{X_i; i = 1, 2, \dots, n\}$ with common distribution function $F_X(x) = \text{Prob}(X \leq x)$ and $I(X_i > u)$ is the indicator function defined as

$$I(X_i > u) = \begin{cases} 1 & \text{if } X_i > u \\ 0 & \text{if } X_i \leq u \end{cases} \quad (9.5.2)$$

The probability distribution function of K exceedances is given by

$$\begin{aligned} f_K(k) &= \text{Prob}(K = k) \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned} \quad (9.5.3)$$

for $k = 1, 2, \dots, n$, which is a $BI(n, p, 1, 0)$ binomial distribution where $p = 1 - F_X(u)$. The mean number of exceedances is given by

$$E[K] = np$$

$$= n(1 - F_x(u)) \quad (9.5.4)$$

The following limit holds if it is assumed that the threshold u depends on n denoted as u_n , which is a sequence of real numbers:

$$\lim_{n \rightarrow \infty} n(1 - F_x(u_n)) = \tau \quad (9.5.5)$$

where $\tau \in [0, \infty]$.

A Poisson method consists of increasing the threshold level until consistency with the assumptions of the Poisson approximation to the binomial distribution conditioned the events are independent. Basically the idea is that the use of the point process of exceedances has assumed that such trials can be modeled as repeated Bernoulli trials with two possible outcomes, which are an “*exceedance*” with probability $p(u)$ or “*not an exceedance*” with probability $1 - p(u)$, converges weakly to a homogenous Poisson process with rate parameter τ .

The excesses can be modeled using the *GPD*:

$$F_{GP\gamma}(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}_+\right)^{-1/\gamma} \quad (9.5.6)$$

where $1 + \gamma/\sigma > 0$, $\gamma \neq 0$ is the shape parameter, $\sigma > 0$ is the scale parameter, and “ $(\cdot)_+$ ” signifies the positive part, implying that for $\gamma < 0$, $F_{GP\gamma}(x) = 1$ for $x \geq -\gamma/\sigma$, i.e. the distribution has finite positive endpoint $x_u = -\gamma/\sigma$.

The Pareto distribution functions are defined for $x < 0$ and $\gamma \geq 0$, and are the appropriate fat-tailed distributions for modeling loss distributions. For the expression is interpreted as the limit as $\gamma \rightarrow 0$:

$$F_{GP\gamma}(x) = 1 - \exp\left(-\frac{x}{\sigma}\right), \quad (9.5.7)$$

which is the exponential distribution. This model is called the “*peaks-over-threshold model with general Pareto distributed excesses*.” or, as it is referred to in hydrology, “*the partial duration series model with general Pareto distributed excesses*.” For large losses γ is always positive and is often between 0.5 and 1. For positive γ , the generalized Pareto distribution is closely related to the ordinary Pareto distribution (See section 17.6 of Brockhaus, Farkas, Ferraris, Long, and Overhaus (2000), Davison and Smith (1990), section

6.5 of Embrechts, Kluppelberg, and Mikosch (1997), Rootzén and Tajvidi (2000), and Chapter 1 of Reiss and Thomas (2000)).

10 Explanatory Data Analysis and Extreme Value Theory

Graphical methods can be used to estimate the Pareto index and to qualitatively examine the dependencies between random variables. They also address the following question:

“How does one choose the threshold level?”

The basic graphical methods used in extreme value theory include (see Beirlant, Teugels, and Vyncker (1996), section 6.2 of Embrechts, Kluppelberg, and Mikosch (1997), and Gavin (2000)):

1) The Box and Dot Plots

Boxplots are graphical representations of *quantile summaries* where the latter is a summary of various aspects of a density function. They use the following variables: the median, the maximum, the minimum and two hinges which are the midpoints between the maximum and the median and the minimum and the median. *Dot diagrams* are used to describe a frequency distribution when there are only a small number of observations in a sample of quantitative variables. Dot diagrams are plots of the individual data along a line segment where each observation is represented by a “dot” at its value on the line segment.

2) The $Q-Q$ and Probability Plots

Probability plots are also known as *quantile-quantile plots* or $Q-Q$ plots, and are an approach to qualitatively examining the “goodness-of-fit” of the data. They are sometimes called draftsman’s displays. The $Q-Q$ plot displays the quantiles of one attribute or phenomena of one data set against the quantiles of one attribute or phenomena of another data set. The probability plot displays the observed quantiles of the time series on one axis against the quantiles of the hypothesized distribution on the other axis. A concave or convex departure from the straight line indicates a heavier- or shorter-tailed empirical distribution (see section 6.2.1 of Embrechts, Kluppelberg, and Mikosch (1997)).

3) The Mean Excess Plot

The mean excess plot is a plot of the mean of all excesses (divided by the total number of data points) over a given threshold versus the threshold level. The mean excess function is useful in graphically discriminating between different types of tails. When the plot of the mean excess function against the threshold level exhibits a straight line, this is indicative of the generalized Pareto distribution where the expected number of exceedances is a linear function of the threshold. This implies that a positive slope of the mean excess plot is an indication of heavy-tailed behavior. If the plot of the mean excess function versus the threshold level is a constant, this is indicative of an exponential distribution. That is, the mean excess function for $v = x - u$ given $X > u$ is given by:

$$e_{xs}^u(v) = e_{xs}^u(X - u) = \lambda^{-1}$$

where λ is the parameter of the exponential distribution. Note that median excess functions are also used due to their robustness in small samples (see section 6.2.2 of Embrechts, Kluppelberg, and Mikosch (1997)).

4) The Scatter Plot

A *bivariate scatterplot* (or *scatter diagram*) is probably the simplest, most widely used and understood graphical method for displaying the relationship between two time series. One of the time series is a response variable and the other is the predictor variable.

5) Gumbel's Method of Exceedances

Gumbel's method of exceedances is also referred to as the *annual maxima method* or *blocks maxima method*. Gumbel's method of exceedances is used to determine how many future records exceed a given high threshold, which is an observation from a past record. Basically, the procedure can also be used to take block maxima, where the blocks are periods of time such as years or months (see section 6.2.3 of Embrechts, Kluppelberg, and Mikosch (1997) and Gumbel (1958)).

11 Estimation of the Tail Index Parameter

11.1 Overview of the Estimation of the Tail Index Parameter

The behavior of the extreme quantiles and tail probabilities are governed by the value of the *tail index parameter* $\alpha = 1/\gamma$ of the underlying extreme value distribution. The tail

index is also called the *Pareto index*, *extreme value index*, or the *shape parameter*. The shape parameter is an intrinsic parameter of an extreme value distribution and does not depend on the number of returns used in defining the tail section. The “*fatness*” of the tails is directly related to tail index parameter.

The values of the tail index denoted γ for the three extreme value distributions include:

- 1) $\gamma > 0$ for the *Fréchet distribution*
- 2) $\gamma = 0$ for the *Gumbel distribution*
- 3) $\gamma < 0$ for the *Weibull distribution*

The extreme value distribution for financial time series, which have empirically exhibited both the non-Gaussian and heavy tails properties, is the Fréchet distribution.

The shape parameter denoted $\alpha = 1/\gamma$ represents the maximal order of finite moments.

- 1) If $\alpha > 1$, at least the mean of the extreme value distribution exists;
- 2) If $\alpha > 2$, at least the mean and variance exist;
- 3) If $\alpha > 3$, at least the first three moments exist; and so forth.

The shape parameter corresponds to the degrees of freedom for a Student t distribution and characteristic exponent of a stable Paretian distribution.

11.2 EVT-based Techniques and the Estimation of High Quantiles

In risk assessment and management applications, frequently estimates of loss quantiles and probabilities are required not only from near the boundary of the range of the observed data, but also at points outside the range. *EVT* techniques are capable of estimating extreme quantiles and the corresponding probabilities through the use of models that fit the empirical survival function of a set of observations using only samples from the extreme regions of the range of the sample, as opposed to the entire sample.

Benefits of using *EVT* to fit the tail regions are as follows:

- 1) The *EVT*-based estimation methods were developed specifically for modeling the tails regions of a distribution, and not the center of the distribution.
- 2) A reasonable functional form can be formulated for the tail region based on *a priori* considerations.

There are two basic principal model approaches to the estimation of the extremal properties of the extreme values and tail events of a probability distribution or a stochastic process:

1) **Block Maxima Models**

The *block maxima method* or *per-period maxima method* is the oldest of the models of classical extreme value analysis, and was proposed in Gumbel (1958). This method is based directly on the sample extremes from non-overlapping subsamples from a given time series. The block maxima models are based on the largest observations collected from large periodic samples of identically distributed observations. From the sequence of maxima, descriptive statistic and maximum likelihood estimates of the distribution of the maxima can be calculated. A drawback to the use of the block maxima method is that the use of subsamples reduces the precision of the estimators. Gumbel used the maxima from annual time series. In this approach one of the three limiting extreme value distributions are fitted to some time series with periodic observation that have been decomposed into blocks. The block maxima approach can be used to determine the seasonality, i.e. quarterly or annual, maximum of a collection of hourly or daily profits and losses derived from transacting in an individual instrument or a portfolio of instruments. McNeil (1998) has proposed the use of block maxima in the definition and analysis of stress losses in a risk management context. Byström (March 1999) uses the block maxima method to create conditional *VaR* forecasts.

2) **Peaks-over-Threshold Models**

The *peaks-over-threshold (POT)* is the more recent approach to modeling the distribution of the extreme values, which focuses on modeling the characteristics of the largest values of a sequence that exceed a given high threshold. The peaks-over-threshold models are also referred to as the exceedances over thresholds models. This approach models the exceedances over a threshold using the distribution of excess values over a threshold value, i.e. value from only the “*tail*” region. These models are more efficient in practical applications than the blocks maxima method when there is a limited amount of data for the extreme values. The key result for the peaks-over-threshold models is that the general asymptotic form of the distribution of the excess value is a *GPD*. This result was first presented in Pickands (1975). This approach is usually applied by fitting the *GPD* to the observed excesses over a threshold. One advantage of this approach over the block maxima method is that since each exceedance is associated with a specific event, making it is possible for the parameter

of the *GPD* to depend on other variables, i.e. covariates. There are various approaches to the implementation of the peaks-over-threshold models (see section 4.1 of K ellezi and Gilli (2000), McNeil (2000), McNeil and Saladin (1997) and Pickands (1975)):

2.1) Semi-parametric models that are derived from the Hill estimator for the Pareto tail index. The tail index of a density function is an indicator of the probability of obtaining a sample that represents a large deviation from the average of central value of this density function.

2.2) Parametric models that are based on heavy-tailed *GPDs*. These methods include methods based on absolute extremes of epochs or events and methods based on exceedances over a given threshold.

Other approaches to modeling the tail regions that are based on variants of the basic theory of extreme values include:

- 3) The modeling of the joint distribution of the k largest order statistics. This is an extension of Gumbel's annual-maxima approach to the modeling of the joint distribution of the k largest- or smallest-order statistics in each year. Smith (1986) and Weissman (1991) fit a generalized extreme value distribution to the annual maxima of a time series using a method based on the asymptotic joint distribution of a fixed number of extreme order statistics. (See Smith (1986), Tawn (1988), Leadbetter, Lindgren, and Rootz en (1983) and Weissman (1991))
- 4) The point process approach to modeling high-level exceedances. In this approach the exceedance times and excess values are modeled as a two-dimensional point process. If the point process is stationary and satisfies a condition that asymptotically there are no clusters among the high-level exceedances, the limiting form will be a non-homogeneous Poisson process with an intensity measure on a set A of the form $(t_1, t_2) \times (y, \infty)$ where $y > u$ and (t_1, t_2) is bounded interval on the real line. (See Leadbetter, Lindgren, and Rootz en (1983), Resnick (1987), Smith (1989) and Smith (2000))
- 5) A $Q-Q$ plot where the sample quantiles are plotted against the quantiles of a generalized Pareto distribution (*GPD*). The choice of *GPD* is the choice where the parameters results in a $Q-Q$ plot that is approximately a straight line (see Section 1.5 of Beirlant, Teugels, and Vyncker (1996)).

11.3 Overview of Tail Index Estimation Procedures

The likelihood of an extreme event can be determined using the estimate of the tail index. The tail index estimate allows for the extrapolation of the estimated empirical distribution outside its domain of observed samples. The “*fatness*” of the tails and the shape of the tail region are directly related to tail index parameter. The shape or tail index parameter denoted $\alpha = 1/\gamma$ represents the maximal order of finite moments. For example, if $\alpha > 1$, at least the mean of the extreme value distribution exists; if $\alpha > 2$, at least the mean and variance exist; if $\alpha > 3$, at least the first three moments exist; and so forth. The shape parameter is an intrinsic parameter of an extreme value distribution and does not depend on the number of returns used in defining the tail section. The shape parameter corresponds to the degrees of freedom for a Student t distribution and characteristic exponent of a stable Paretian distribution (See Section 2.3.3 of Beirlant, Teugels, and Vyncker (1996), section 17.10 of Brockhaus, Farkas, Ferraris, Long, and Overhaus (2000), Dekkers and de-Haan (1989), DuMouchel (1983), Section 6.4.2 of Embrechts, Kluppelberg, and Mikosch (1997), Kearns and Pagan (1997), Pictet, Dacorogna and Müller (1998), and Smith (1987)).

If $\alpha = \infty$, the tails decline exponentially, and in many cases the distribution is normal. If $\alpha < \infty$, the tails are fat and they decline as a power law. Some characteristics of the tails index in this case include:

- 1) α is the number of finite moments.
- 2) For the Student t distributions, α is the number of degrees of freedom.
- 3) For the stable distributions, α is the characteristic exponent.

The general procedure to estimate the tail index $\alpha = 1/\gamma$ of the distributions that are in any of the maximum domain of attraction of the three cases of the generalized extreme value distribution is as follows:

- 1) Choose an appropriate estimator for the shape parameter γ of the general extreme value distribution function.
- 2) Choose appropriate estimators for the normalization constants.
- 3) Test that the proposed estimators from steps one and two result in reasonable approximations to the tail in the end of the tail where the high quantiles are located.
- 4) Determine the statistical properties of these estimators.

The important question that must be addressed in the estimation of the tail index is:

How many extreme values to include in the optimal estimator of the tail index?

That is, the choice for the number of upper order statistics to include in the tail index estimator is a function of the sample size n , i.e. $k = k(n)$ and the following mathematical conditions are usually imposed of the choice of $k(n)$:

- 1) Use a sufficiently large number of order statistics, i.e. let $k(n) \rightarrow \infty$.
- 2) Make sure the choice of the number of upper order statistics concentrates on the upper tail, i.e. let $n/k(n) \rightarrow \infty$.

DuMouchel (1983) recommends that k should not exceed ten percent of the sample n .

Major issues in the identification and estimation of the tails include:

- 1) Where does the tail section of the distribution start?
- 2) The infrequency of the tail events in samples.
- 3) The effects that the drying up of liquidity that is typically associated with an event market events.

Various parametric methods for estimating the generalized Pareto distribution parameters σ_U and γ are reviewed in Hosking and Wallis (1987).

11.4 Tail Index Estimation using Maximum Likelihood Methods

The tail index can be parametrically using the maximum likelihood method. That, is, the parameters of the generalized Pareto distribution can be estimated using the maximum likelihood method. Whereas maximum likelihood estimators exist in large samples provided that $\gamma < 1$, they are asymptotically normal and efficient when $\gamma < 1/2$ (see Smith (1985)). Using the approach in described in Neftci (2000) or Smith (1985), the parameters σ_U and γ , which are the normalizing coefficients, are obtained by maximizing the log likelihood function of a *GPD*.

The tail index can be estimated parametrically using the maximum likelihood method. There are drawbacks associated with maximum likelihood method applied to the estimation of the tail index including:

- 1) Maximum likelihood assumes that the extreme values exactly follow the limiting distribution. The convergence-to-types theorem is an “*approximation*” in finite samples. Therefore, the Cramer-Rao lower bound may not be realizable, and more efficient estimators exist.
- 2) The normalizing coefficients must be estimated along with the tail index when maximum likelihood is used.
- 3) A sampling interval must be specified

Assuming that x_U is high enough so that the generalized Pareto distribution $F_{GPD}(v_t; \sigma_U, \gamma)$ with $\gamma \neq 0$ is a good approximation for the probability $F_U(v_t)$, then:

$$\text{Prob}(x_t < v_t) \cong 1 - \left(1 - \frac{\mathcal{W}_t}{\sigma_U}\right)^{1/\gamma} \quad (11.4.1)$$

The above equation holds for $\gamma \leq 0$. In the case that $\gamma > 0$, the condition where $1 + \mathcal{W}_t/\sigma_U > 0$ must be satisfied for the density to be well defined. Following the eq. (11.4.1), the density function of x_t can be approximated at an arbitrary observation point v_t , by the density $F_{GPD}(v_t; \sigma_U, \gamma)$:

$$F_{GPD}(v_t; \sigma_U, \gamma) = \left(\frac{\sigma_U - \mathcal{W}_t}{\sigma_U}\right)^{1/\gamma} (\sigma_U - \mathcal{W}_t)^{-1} \quad (11.4.2)$$

Finally, by applying the density of $F_{GPD}(v_t; \sigma_U, \gamma)$ at each observation point v_t , the following log likelihood function is obtained:

$$\ell(\gamma, \sigma_U, k) = -k \ln(\sigma_U) + (\gamma^{-1} - 1) \sum_{i=1}^k \ln\left(1 - \frac{\mathcal{W}_i}{\sigma_U}\right) \quad (11.4.3)$$

where, k is the number of exceedances in a sample of n observations. In this case, the number of sample extremes (k) is obtained by first estimating the standard deviation of the entire sample of the returns.

11.5 Linear Regression and Tail Index Estimation

The Hill estimator can be estimated using linear regression, although it was not originally motivated using a regression framework. This can be noted by taking the natural logarithm of both sides of the power law function (see eq. (4.3.2)):

$$\ln(\text{Prob}(X > x)) = \ln(k) - \alpha \ln(x), \quad (11.5.1)$$

implying that α is the slope coefficient in a simple linear relationship between the natural logarithm of the tail of the empirical survival function and the natural logarithm of the k most extreme tails (see Diebold, Schuermann and Stroughair (2000) and Kearns and Pagan (1997)).

11.6 Non-parametric Tail Index Estimators

There are a large number of nonparametric tail index estimators that have been developed. Chapters 2 and 3 of Beirlant, Teugels, and Vyncker (1996), section 6.4 of Embrechts, Kluppelberg, and Mikosch (1997) and Pictet, Dacorogna and Müller (1998) contain discussions of a large percentage of these nonparametric estimators. The nonparametric estimators that are used in the comparison are the De Haan and Resnick (1980), Hill (1975) and Pickands (1975). The Hill estimator is an estimator for the Pareto index, γ of the heavy-tailed distributions, which are in the maximum domain of attraction of the heavy tailed Fréchet distribution, i.e. for positive values of the Pareto index. The Pickands and the De Haan and Resnick tail index estimators can be used to estimate the tail index over the entire range of the tail index for all three limiting extreme value distributions

Let k be the number of upper order statistics to include in the estimator out of a sample of size n where $n \gg k$. The De Haan and Resnick (1988), the Hill (1975), and Pickands (1975) estimator are based on the ordered sample

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(k)} \leq x_{(k+1)} \dots \leq x_{(n-1)} \leq x_{(n)} \quad (11.6.1)$$

of the observed sample of log returns $\{x_1, x_2, \dots, x_n\}$ where $\{x_{(j)} : j = 1, 2, \dots, n\}$ denote the order values of $\{x_1, x_2, \dots, x_n\}$. The Hill (1975) estimator is given by

$$\mathcal{Y}_{k,n}^H = -\frac{1}{k-1} \sum_{i=1}^{k-1} \ln(x_{(i)}) - \ln(x_{(k)}) \quad (11.6.2)$$

The Hill estimator can be interpreted as the average excess vertical excess of the log-transformed data above a given threshold. The Pickands (1975) estimator is given by

$$\mathcal{Y}_{k,n}^P = \frac{1}{\ln(2)} \ln \left(\frac{x_{(k)} - x_{(2k)}}{x_{(2k)} - x_{(4k)}} \right), \quad (11.6.3)$$

and the De Haan and Resnick (1988) estimator is given by

$$\gamma_{k,n}^{HR} = \frac{\ln(x_{(1)}) - \ln(x_{(k)})}{\ln(k)} \quad (11.6.4)$$

11.7 Comparison of Maximum Likelihood Estimator with Non-Parametric Estimators

As discussed in the previous sections, the tail index can be parametrically or non-parametrically. There are drawbacks associated with maximum likelihood method when applied to the estimation of the tail index. The maximum likelihood method assumes that the extreme values exactly follow the limiting distribution. Maximum likelihood estimation is fully, asymptotically, efficient when the Cramer regularity conditions hold. In finite samples, especially for the generalized Pareto Family of distributions, the endpoints are typically not the exact. Therefore, the Cramer-Rao lower bound may not be realizable, and more efficient estimators exist. Also, the normalizing coefficients must be estimated along with the tail index when maximum likelihood is used, and a sampling interval must be specified.

As a comparison, the results of three common non-parametric tail index estimators will be compared the maximum likelihood estimator. Non-parametric tail index estimation techniques are based on simple functions of the order statistics of the entire sample. They also do not assume that the order statistics exactly follow the limiting distribution. In the nonparametric estimation of the tail index of a *GPD* one chooses a high threshold and then fits the tail index to the realizations in the corresponding tail region. Additionally non-parametric tail index estimators are not seriously affected by autocorrelations. A drawback to these nonparametric estimators is that they are not asymptotically efficient estimators.

The tail index parameter can also be estimated using a Bayesian framework (see section 11.5 of Reiss and Thomas (2000) with Hosking, Smith (2000) and Smith and Goodman (1994))

12 Applications of Extreme Value Theory

EVT has been applied to financial modeling, risk management, hydrology and insurance (see Embrechts (editor) (2000), which is an recent edited volume of applications in risk management and insurance). *EVT* is a useful alternative in the modeling of market, credit, and operational risk exposures. One of the main applications of extreme value theory is the measurement of risk relative to a loss distribution function. *EVT* to evaluate risk measures such as the Value-at-Risk (*VaR*), which is defined as the probable loss at a high quantile and the *expected shortfall* or “*beyond VaR*” measure of risk, which is defined as the

expected amount by which VaR may be exceeded in the rare event that VaR is exceeded (see section 6.1 of Embrechts, Kluppelberg, and Mikosch (1997)). The determination of the extreme tails distributions of the loss distribution is an important aspect of the functions of the risk regulators especially in the context of the potential-loss determination measures such the capital at risk. EVT also allows for the procedures to check for the probability of exceedance of some critical regulatory capital level during a fixed time interval.

EVT has been applied to the following areas:

- 1) Loss claims from major natural disasters such as droughts, earthquakes, flood, and hurricanes.
- 2) Flood levels of rivers
- 3) Large upward or downward movements in financial markets over a given time period.
- 4) Extreme levels of environmental indicators such as ozone or carbon monoxide levels.
- 5) Wind speeds at certain measurement sites.
- 6) Waves heights during storms
- 7) The maximal or minimal performance levels of a portfolio of financial assets.
- 8) The modeling of financial crises.
- 9) The modeling of loss distributions in the context of credit and operational risk management

The estimators for VaR at extremely high quantile and the “*beyond VaR*” measure of risk will be discussed in detail in next major section (Section 12.1). Then sections of other modeling financial time series (Section 12.2), risk management applications (Section 12.3) and insurance applications (Section 12.4) follow, respectively.

12.1 Extreme Value Theory and Value-at-Risk

12.1.1 Extreme Events and Value-at-Risk

Value-at-risk (VaR) is a useful measure of market risk for “*business as usual*” (or “*normal*” or “*stable*” or “*ordinary*” or “*quiet*”) market conditions, but it has serious limitations in the measurement of extreme event or catastrophic risk under extreme (or “*abnormal*” or “*catastrophic*” or “*hectic*” or “*unusual*”) market conditions. Basically, VaR is a statistical estimate of a portfolio loss under normal market conditions, such that there is a given (small) probability that a loss of that size or greater will occur over a specific time

horizon. In other words, *VaR* is the upper bound on the losses that should not be exceeded more than a small preset fraction of the time during a future time horizon and under normal market conditions. This exceedance probability of the upper loss bound is defined relative to a coverage level or confidence level and the institution's risk preference. For example, if a bank's daily *VaR* forecast is US\$ 50 million relative to 99 percent coverage level, then under normal market conditions, there is a chance of one in one hundred that a daily loss of greater than US\$ 50 million will occur. *VaR* developed into one of the most widely accepted methodology for measuring market risk during the late 1990s used by regulators, financial and nonfinancial institutions, and asset managers. Many of the early *VaR* techniques such assumed that log-returns of the individual components of the portfolio were jointly normally distributed, were the log-returns are the natural logarithm of the price relatives. The normal distribution for log-returns is a more than suitable theoretical choice since has very tractable properties and it can be completely described by its mean and variance. Unfortunately, the normality assumption has been continually criticized as inadequate in numerous empirical financial studies beginning with Mandelbrot (1963).

At a general level *VaR* measures market risk for a stable market environment where future price changes reflect historical movements. In the context of extreme or even catastrophic markets events such as excessive currency devaluations or a stock market crash, *VaR* is rendered useless since the relationship between historical and future price breaks down. The problem is that standard *VaR* measures will not reflect the extremal events in the risk measurement using *VaR* unless such events fall within the "look-back" period. The issue is that using the most recent historical information in a relatively short look-back period gives a relevant measure of *VaR* under normal conditions. Unfortunately short lock-back periods are not likely to extreme or catastrophic events, which will affect the results of the measurement of risk. Even if an extreme or catastrophic event is in the lock-back period, it may not influence the *VaR* estimate depending on how the estimate is calculated since *VaR* is a measure of the potential loss at a given one-side confidence level and the extreme or events may be beyond this boundary and they might no influence the risk measurement. To overcome this problem one should use a risk measure such as the expected shortfall, which gives the expected loss given that the loss was greater than a threshold such as *VaR*.

In the analysis of fat-tailed financial time series, the returns distributions can be decomposed into two different regions:

- 1) A bell-shaped or approximately normally distributed center region, which includes the business-as-usual events. These makes up the majority of the observations, and occur

in a market volatility regime that has a stable, normal relationship between the magnitude of changes and the frequency. In the short-term, this stable relationship allows for future forecast of the distribution and to calculate risk measurements and determine the risk capital. Business-as-usual regions change slowly and this is the basis of their use in *VaR* calculations.

- 2) A section of outliers or extreme tail observations, which usually represents a major structural change or disruption in the markets. Due to the scarcity of observations in these regions it is difficult to fit general empirical distribution to the data in these regions. Parametric distributions associated with extreme value theory can be used to overcome to some extent the scarcity of data.

The difficulties in properly assessing the influence of the extreme events in a standard risk measurement such as *VaR* can render the measurement useless. Standard *VaR* forecast models can be extended using *EVT*-based techniques to measure specific types of event risk, but it cannot, by construction, describe the magnitude of the extremely rare unexpected returns. In this context, the appropriate measurement of event risk should incorporate complementary measures of risk such as the *EVT*-based expected shortfall, which measures the expected loss, given that returns have fallen below some specified threshold. Both of these extensions of the standard *VaR* model will be briefly discussed in the subsection (Sections 12.1.2).

12.1.2 The *EVT*-based Value-at-Risk Extension and the Expected Shortfall

EVT methods can be used to approximate risk measures that are defined at high loss quantile levels such as the extreme *VaR* and the *expected shortfall* or a “*beyond VaR*” risk measure. The *GPD* is the appropriate model for approximating the conditional distribution of exceedances above a high loss threshold using only samples from the extreme region as opposed to the entire data set, such that a significant fraction of the values are above this threshold, but a much larger fraction do not exceed it. The appropriateness of the *GPD* approximation requires the threshold level to be taken sufficiently high enough such that the probability of exceeding it is very close to zero, and such that the law of small numbers is applicable. The focus of the presentation for the *EVT*-based risk measures will be on the loss distribution of the log-returns, implying the extreme losses are located in the right most tail of the loss distribution. The loss distribution is located on the left side of the log-returns distribution. This section uses the mathematical results from section (9.1) (Also see

sections 17.6.1 and 17.6.2 of Brockhaus, Farkas, Ferraris, Long, and Overhaus (2000), Embrechts, Resnick and Samorodnitsky (January 1998), section 3.2 of K ellezi and Gilli (2000), and pages 7-8 of McNeil (2000)).

The approximate *EVT*-based extreme *VaR* forecast at a high loss quantile with a log-returns $r = r(t)$ parent distribution is:

$$VaR_q = r_u + \frac{\sigma}{\gamma} \left(\frac{n}{k_u} (1-q)^{-\gamma} - 1 \right) \quad (12.1.2.1)$$

where $VaR_q > r_u$, r_u is the upper tail threshold, k_u is the number of the n samples of $r(t)$ that exceed r_u , $VaR_q = VaR_q(r_u)$ is the q -th loss quantile estimated relative to r_u , and u is the quantile of the threshold and is such that $q > u$. The q -th loss quantile x_q is such that $100 \cdot q$ percent of the observations are less than x_q . The choice of r_u allows for the VaR_q estimate to “fall beyond” the observed range. For example, r_u may be defined at the 95-th loss quantile, i.e. $u = 0.95$, whereas VaR_q may be defined at $q = 0.99995$.

The “beyond *VaR*” forecast will be defined using the mean excess function at a high tail threshold r_u , which is the expected shortfall for a loss distribution and given by:

$$\begin{aligned} \mu_{r_u} &= E[r - r_u | r > r_u] \\ &= E[r | r > r_u] - r_u \end{aligned} \quad (12.1.2.2)$$

where $X - r_u$ is an excess over r_u . The expected shortfall over VaR_q defined relative to r_u is given by:

$$\begin{aligned} \mu_{VaR_q} &= E[r - VaR_q | r > VaR_q] \\ &= \frac{\sigma}{1-\gamma} + \gamma \left(\frac{VaR_q - r_u}{1-\gamma} \right) \end{aligned} \quad (12.1.2.3)$$

The following is the expected loss conditioned that the loss is greater than a VaR_q forecast defined relative to r_u :

$$E[r | r > VaR_q] = VaR_q + \mu_{VaR_q} \quad (12.1.2.4)$$

If the log-returns $r = r(t)$ are normally distributed:

$$\begin{aligned}
E[r|r > VaR_q] &\approx VaR_q + VaR_q^{-1} \\
&\approx VaR_q
\end{aligned}
\tag{12.1.2.5}$$

since as one moves towards the end of the tail of a normal distribution, the inverse of VaR_q approaches zero, i.e. $VaR_q^{-1} \rightarrow 0$ as $q \rightarrow 1$. The implication is that for normally distributed log-returns, standard VaR techniques are adequate. This is not the case for fat-tailed loss distributions such as a Pareto distribution with tail index $\gamma > 1$:

$$\begin{aligned}
E[r|r > VaR_q] &= VaR_q + \left(\frac{1 + VaR_q}{\gamma - 1} \right) \\
&\approx \frac{\gamma}{\gamma - 1} VaR_q
\end{aligned}
\tag{12.1.2.6}$$

The estimated values of the tail index from empirical studies of financial time series exhibit a range of $1.5 < \gamma < 5$. This range implies that for fat-tailed Pareto log-returns:

$$1.2 VaR_q < E[r|r > VaR_q] < 3 VaR_q . \tag{12.1.2.7}$$

These lower and upper bounds on the “*beyond VaR*” forecast are a justification for using the more sophisticated *EVT* approaches. The upper bound is also a justification of the suggested Bank of International Settlement (*BIS*) risk capital multiplier of three.

12.2 Extreme Events, Extreme Value Distributions and Tails of Financial Time Series

The following features of extreme value theory (*EVT*) make it attractive method in the estimation of the tail region (see De Vries, Caserta, Reiss, and Thomas (2000)):

- 1) *EVT* is based on a well-developed statistical theory.
- 2) *EVT* presents a parametric form for the tail of a distribution.
- 3) *EVT* is useful in discriminating between alternative probability models for financial market returns.

These characteristics allows for the extrapolation beyond the range of the data. Note, in the context of *EVT*-based tails, values taken from beyond the range of the data need to be used with care.

Extreme value distributions and the corresponding tail region analysis have been applied to the following areas of financial time series modeling and applied portfolio theory:

1) **Equity Markets**

EVT can be used to discriminate between alternative heavy-tailed probability models for stock market returns. *EVT* has been applied to equity prices to investigate the following:

- 1.1) Whether tail region of the returns process should be modeled using fat-tailed distributions.
- 1.2) The likelihood of an extreme movement, such as 20 percent in a single trading day.
- 1.3) Whether the likelihood of a extreme movement changes after market crashes such as the October 1987 crash.

Jansen and De Vries (1991) estimated the tail index for US stock market returns. Example 6.2.9 on pages 298, 302 and 303 of Embrechts, Kluppelberg, and Mikosch (1997) analyze the daily log-returns for BMW over the period from January 2, 1973 to July 12, 1996, which results in 6,146 observations. The BMW data set is an example where the upper and lower tails are quite different. Also see Broussard and Booth (1998), Gavin (2000), Jansen and De Vries (1991), Longin (1996) Loretan and Phillips (1994), and Vilasuso and Katz (2000) for other examples of applying *EVT* to the modeling of equity returns.

2) **FOREX Markets**

Another example of a financial application of *EVT* concerns the upper limits on open position to foreign currency dealers by the treasurers of a *FOREX* dealing room of an international bank. The trading limits should be a function of the probability of a single negative return that is large enough to endanger the solvency of the bank. The tail index can also be used to detect for regime switches that affect the type of distribution. Koedijk, Schafgans and de Vries (1990) examined the properties of exchange rates under the European monetary system. Hols and De Vries (1991) studied the US Dollar-Canadian exchange rate. Danielsson and de Vries (1997) model high frequency foreign exchange records using *EVT* methods. for other examples of applying *EVT* to the modeling of equity returns. Also see Huisman, Koedijk, Kool and Palm (1998),

Koedijk and Kool (1992), Koedijk, Schafgans and de Vries (1990), Koedijk, Stork and De Vries (1992), Starica (1998) or Starica (2000) for other examples of applying *EVT* to the modeling of equity returns.

3) **Commodities Markets**

Kofman and de Vries (1989) apply extreme value theory to the modeling of the potato futures. Their results point towards the use of a Student *t* distribution with 3 degrees of freedom or a low-order *ARCH* process.

4) **Interest Rates**

Marakov (1997) present an example that explicitly depicts the utility in modeling the center and tail regions separately. Figure 6 on page 49 of Marakov (1997) exhibits the distribution of monthly observation of the three month LIBOR rate from 1973 to 1991. In this figure one can observe some very large movements in interest rate in the early 1980's. This was during a time when the US Federal Reserve Bank decided to attempt to reduce inflation by abandoning its long-standing policy of maintaining stable levels of interest rates and significantly increasing short-term interest rates. The result was a time of increased market volatility as reflect in Figure 6 where the extremal observations are from the period of May 1980 to May 1981. There are two different regions to this distribution:

- 1) A bell-shaped or approximately normal region, which includes the “*business-as-usual*” events.
- 2) A section of outliers or extreme tail regions, which usually represents a major structural change or disruption in the markets.

Borkovec and Klüppelberg (2001) and Kuan and Webber (1998) also analyze interest rate series using *EVT* methods.

5) **Portfolio Theory and Safety First Behavior**

Safety-first investors are those investors who are worried about guaranteeing a minimum return. This implies they measure portfolio risk using the tail probabilities of the extreme excesses of the returns distribution for the portfolio. Recognizing this is important since the performance of a portfolio over a given long-investment horizon is a function of a few exceptional trading days such as price corrections or positive reaction to good news, where most of the other days contribute marginally to the

bottom line. Extreme value statistics are the appropriate tools to address this problem. This is opposed to the standard mean-variance portfolio theory approach that uses the variance-covariance matrix to measure portfolio risk. Risk diversification for the safety-first investor is induced by the dependence of the extreme returns and not through the linear correlations of the asset's returns. (see De Haan, Jansen, Koedijk, and De Vries (1994))

10) Volatility of the Rate of Return for Common Stocks

Parkinson (1980) and Wiggins (1991) used intra-period maximum and minimum to generate volatility forecasts.

6) Pension Funds and Asset Allocation

Also consider a board of a pension funds that must decide the maximum percentage of the managed funds that can be allocated to the most risky investment categories such as equities, foreign currency and derivatives. The quotas are a function of the solvency and liquidity requirements of the pension fund, conditioned that the fund must be able to pay out every period. In this case the board is worried about an extremely negative return wiping out too much of the capital that the fund becomes illiquid or insolvent. (see Longin (1996))

7) The Modeling of Volatility Skews and Smiles

Options pricing models have been developed where heavy-tailed distributions are used for the underlying asset. Eberlein, Keller and Prause (1998) used a hyperbolic Lévy process to model various German equity data. Ahn, Bouabci and Penaud (February 2000) use a kernel density estimator to fit a hyperbolic distribution for underlying assets that consisted of Brazilian equity instruments. (Also see Caserta, Danielsson and de Vries (2001))

8) Stock Returns in Emerging and Mature Markets

Jondeau and Rockinger (April 1999) examine the tail behavior of five mature markets and nine Asian markets, six Eastern Europe, and seven Latin American emerging markets. (Also see Pownall and Koedijk (1999))

9) Stability of Stock Markets Returns and Financial Crises

Quintos, Fan, and Phillips (2001) use the tail index estimator of Hill (1975) to develop test of the constancy of the tail behavior over time when there is a known (or exogenous) break date in the time series leading to a financial crisis. Quintos, Fan, and Phillips (2001) apply their tests for time-varying tail index parameters to stock market returns over the time period January 2, 1995 to October 16, 1998.

12.3 EVT, Risk Management and Risk Measurement

EVT and the corresponding tail region analysis has been applied to the following areas of risk management and risk measurement:

1) Value-at-Risk and Extensions

Classical *VaR* techniques use information from the entire range of the profit/loss (P/L) distribution, whereas *EVT* procedures only take into consideration the relevant information from the extremes contained in the left tail of the P/L distribution. Basically, *EVT* addresses the extreme-down-side risk of the portfolio and not its overall global risk. *EVT* allows senior management and regulators to set minimal requirements for the level of risk to be held against unexpected losses or even as a maximal allowable limit on potential losses. *EVT*-based measures of risk can be viewed as the natural fat-tailed extension of the standard variance-covariance *VaR* technique, which is suitable for short-tailed profit/loss distributions. In this context, the following questions are relevant:

- 1.1) How does one estimate *EVT*-based *VaR* forecast for a given portfolio?
Answering this question requires estimation of a high quantile, which in many cases is outside the range of the observed data.
- 1.2) Given that a *VaR* forecast has been exceeded: how does one estimate the probability that it was exceeded by a specific amount? This question can be addressed by calculating the excess or short-fall distribution. The generalized Pareto distribution serves as the parametric model that is the natural approximation for this situation.

Longin (2000) used *EVT* to formulate a *VaR* measure that is compatible with the presence of extremely rare market events. Longin (2000) used an *ad-hoc* aggregation formula to approximate the value-at-risk measure. Also see Danielsson and de Vries (1996), Danielsson and de Vries (December 1997), Danielsson and de Vries (2000), or

Example 6.1.6 on pages 287-288 of Embrechts, Kluppelberg, and Mikosch (1997), Longin (2000)) for others applications of *EVT* to extend standard *VaR* risk measure.

2) The Expected Shortfall as a Risk Measure

Artzner, Delbaen, Elber, and Heath (July 1999) have proposed the *expected shortfall* as an alternative risk measure to *VaR*. The expected shortfall is the expected magnitude of the loss, conditioned that the loss has exceeded a *VaR* forecast defined relative to the same high quantile and the corresponding coverage level. The expected shortfall is usually a coherent risk measure as defined using the axiomatic approach of Artzner, Delbaen, Elber, and Heath (July 1999).

3) Regulatory Capital Adequacy and Extreme Value Theory

The use of extreme value theory is suitable for situations where the risk is associated with the tail region of a profit/loss distribution rather than the entire or center region of the distribution. Regulators concerned with capital requirements for securities firms should be interested in the possibility of bankruptcy, which could result form an extremely large change in the value of a firm's portfolio. From the regulatory viewpoint, extreme value theory can be used to determine the capital that must be put aside by a financial institution to cover the largest possible loss in such a way that the firm can stay solvent even after the largest of market shocks. This regulation issue is not currently imposed in the United States nor is it suggested in the recommendations of the Basle Committee reports on banking supervision. In these situations, risk measured by extreme value statistics may be more efficient than the common risk measure of variance. At the catastrophic extreme, regulators are interested in the market conditions during a crisis since these events can be a source of systemic risk (See Danielsson, Hartmann and de Vries (January 1998)).

4) Futures Margins

Another financial application of *EVT* concerns the setting of margins in the derivatives markets: Margin setting in the futures markets is known to be sensitive to the occurrence of large price changes. Margin committees and brokers in the future markets must deal with the following trade-off in the process of setting the margin level: A high margin level protects brokers against insolvent customers, which in turn reinforces market integrity, but it also increases the costs that must be supported by the investors, which in the end makes the market less attractive. Extreme value theory can

be used to derive the margin level for a given probability of margin violation desired by margin committees or brokers. Longin (1994) proposed a new method to set margins that takes into account the appropriate amount of extremes in the distribution of price changes and provides a simple analytical formula to compute the “*optimal*” margin level. Broussard (2001) and Broussard and Booth (1998) also analyze future margins series using *EVT* methods.

5) **Stress Testing**

McNeil (1998) has proposed the use of block maxima in the definition and analysis of stress losses. Byström (March 1999) use the block maxima method to create conditional VaR forecast and applies the method to two different stock indexes; the Swedish Affärvärlden Generalindex and the Dow Jones Industrial Index.

6) **Scenario Analysis**

In the area of risk management, *EVT* methods have been used in the context of scenario analysis and the modeling of loss distributions. McNeil and Saladin (2000) used a portfolio of windstorm losses to develop scenarios for forecasting future extreme losses using the peaks-over-thresholds method.

7) **Modeling of Credit Risk Exposures**

In the assessment of credit risk exposures, in conjunction with economic capital calculations, the normality assumption is not adequate since credit profit/loss portfolios are typically highly skewed and therefore, have extreme long and heavy tails. *EVT* can be used for quantifying the boundaries between these different loss classes, which are the expected loss, the unexpected loss, and the extreme loss. *EVT* offers a scientific language for translating management guidelines on the loss boundaries into actual numbers. The expected and unexpected loss can be quantified standard techniques, whereas the extreme loss is the most difficult part of the loss distribution and requires the use of *EVT*. This implies that *EVT* can be used to model default probabilities of bonds and in the estimation of the so-called diversification factors in the management of bond portfolios (see Section 4.3 of Arvanitis and Gregory (2001) and Chapter 10 of Ong (1999)).

8) **Portfolio Credit Risk**

Lucas, Klaassen, Spreij and Straetmans (2001) derive an explicit characterization of the extreme tail behavior of credit losses in terms of the underlying portfolio characteristics. They state that under quite general conditions credit losses have a polynomial or fat tail rather than an exponential or thin tail. They also investigate the relationship between the credit quality as measured by the probability of default and the credit loss tail index.

9) Credit Spread Risk

Phoa (Spring 1999) uses extreme value theory to model the maximum one-day shifts in swaps spreads, specifically the Australian dollar swap spread. The driving force behind this application was the dramatic widening of credit spreads that occurred in the third quarter of 1998. The widening of swaps spreads surprised most market participants and caused losses that greatly exceeded the reported values of standard *VaR* forecasts.

10) Operational Risk

In the assessment of operational risk exposures, the normality assumption is not adequate since operational risk loss distributions are typically highly skewed to the right and therefore, have extreme long and heavy right tails. Cruz (November 1999) and Cruz, Coleman, and Salkin (2000) use the *GPD* in the modeling of catastrophic operational losses using operational risk-linked bonds. Medova (November 2000) uses Bayesian-version of the peaks-over-threshold (*POT*) method to explicitly model the excess loss distribution in the extreme tail of the operational loss distribution. Balestra (April 2000) and Ceske and Hernández (November 1999) also discuss the application of *EVT* to operational risk measurement.

11) Integrated Risk Management

EVT, which was originally developed for the analysis of rare events that were acts of nature, is currently receiving attention in the area of integrated risk management. Integrated risk management is concerned with the measuring the potential effects that rare events cause since they translate into changes in the regulatory capital requirements (see Embrechts (2000A), Embrechts (2000B), and Embrechts, Resnick, and Samorodnitsky (January 1998)).

12.4 EVT Applications in Actuarial Science and Insurance

When the claims distribution is fat-tailed, the influence of the extremely large, but unlikely claims in the determination of the final premium calculations can be considerable. *EVT*

techniques can be used to model these extremely large, but unlikely claims and therefore help determine more plausible final premium calculations. An early actuarial science application of *EVT* techniques was the development of the Benktander distributions, which were introduced in the context of modeling fat-tailed non-life insurance claims by Gunnar Benktander (see Benktander (1970)). A large number of the analytical tools that form the actuarial techniques have been applied to the integrated risk management framework. This is especially true in the modeling of credit and operational loss distributions where the losses occur at discrete stochastic time intervals; are random in size; and such that the size can possibly grow over time, which is similar to the structure of non-life insurance claims. Also the frequency and size of the losses may depend on exogenous factors. (see Chapter 1 of Embrechts, Kluppelberg, and Mikosch (1997), Radtke, Reiss, and Thomas (2000) and Teugels (1970))

In the context of the ruin problem, some of the important variables inherent of the Lundberg-Cramér model of insurance portfolios, say up to a time t include (section 1.1 of Embrechts, Kluppelberg, and Mikosch (1997)):

- 1) The initial reserve or capital of the insurance company at some initial time t_0 such that $t_0 \leq t < T$.
- 2) The realized premium income up to time t and the expected premium income after time t and before some future time T , such that $t < T$.
- 3) The epochs of the claims up to time t and the epochs of the expected future claims after time t and before some future time T , such that $t < T$.
- 4) The claim size process over the period. A number of small (large) claim size distributions are listed in Table 1.2.5 on page 34 (Table 1.2.6 on page 35) of Embrechts, Kluppelberg, and Mikosch (1997)
- 5) The claim times process, which is the time when claims occur, and the interarrival time between claims. The claim size process and the interarrival time between claims are independent of each other.
- 6) The number of claims process over the period
- 7) The risk reserve held at time t . The risk reserves are funds that are usually set aside at the beginning of a book year.

The volume edited by Hélyette Geman contains some recent applications of *EVT* in the areas of insurance and weather derivatives (see Geman (editor) (1999)). The following are examples of where *EVT* has been applied in the actuary modeling of insurance policies:

1) **Fat-tailed Claims Distributions and the Expected Total Claim Amount**

Important stochastic variables in the modeling of the expected total claim amount and the loss distribution from claims include:

- 1.1) The number of claims
- 1.2) The claim size
- 1.3) The claim times process or the interarrival time between claims.

The process of modeling claim sizes of a portfolio contaminated by a small number of extremely large claims requires knowledge of the claim number process and the claim size distribution. Large interarrival times between the realizations of the extremely large claims are to be expected. The distribution of total claims and the ruin probabilities, which are the basic tools of so-called solvency analysis within the collective risk theory framework, can be extended to a context where the claims distributions are heavy-tailed. In this context, the claim number process and the process of claim sizes are assumed to be independent. Note if the claims loss distribution is subexponential, i.e. fat tailed, then the behavior of the entire portfolio of claims is basically determined by the extreme claims. Subexponential distributions are a subclass of heavy-tailed distributions. The extremely low probability of these extremely large claims makes it difficult to properly model the loss distribution from the claims, especially due to the detrimental effects that these one-in-ten-thousand-year loss events can have on the solvency of an insurance company. Note: a distribution is said to be *subexponential* (*superexponential*) if it shows more (less) excentricity than the exponential distributional model, i.e. fatter (thinner) tails than the exponential distribution (see Sections 5.1 and 5.2 of Beirlant, Teugels, and Vyncker (1996) and sections 1.3.2 and 1.3.3 of Embrechts, Kluppelberg, and Mikosch (1997))

2) **Ruin Probabilities for Fat-tailed Claims Distributions**

The classical Lundberg-Cramér ruin probability result, which holds for superexponential distributions, does not hold for subexponential distributions. The analogous results for large claim distributions is given in Embrechts and Veraverbeke (1982), which was a generalization of the earlier results of Teugels, and Veraverbeke

(1973). The heavy-tailed version of the Lundberg-Cramér model of claims is given in section 1.4 of Embrechts, Kluppelberg, and Mikosch (1997). (Also see Section 5.3 of Beirlant, Teugels, and Vyncker (1996) and section 1.4 of Embrechts, Kluppelberg, and Mikosch (1997))

3) Insurance Claims

Smith and Goodman (1994) combine extreme value analysis with statistical methods based on Bayesian inference to analysis 425 insurance claims. They model the distribution of future losses for insurance claims of several different types using Bayesian methods. They consider trend and seasonal variations, loss aggregation affects the influence of outliers and their effect on the modeling of the future loss distribution. Issues that must be addressed in the modeling of extreme insurance claims include:

- 3.1) The determination of the yearly premium volume that is needed to cover with the high probability, the future losses generated by the portfolio of policies held.
- 3.2) The determination of what is the probable maximum loss for the portfolio of policies, where the probable maximum loss is defined at a high quantile on the distribution of loss claims.
- 3.3) The determination of the priority or attachment point a_k for a one-in- k -year event when writing an excess-of-loss cover reinsurance. a_k can be reformulated as a high quantile of the claims distributions. That is, the determination of a_k such that the probability of exceeding a_k equals $1/k$.

Heavy-tailed claim distribution can be modeled using the method of exceedances or the peaks-over-threshold method, which uses the *GPD* if certain conditions are met (see Example 6.1.3 on pages 285-287 of Embrechts, Kluppelberg, and Mikosch (1997)).

4) Fire Insurance Claims

Embrechts, Resnick, and Samorodnitsky (January 1998) apply an extreme-value theory based tail-fit to 8043 industrial fire insurance claims. Embrechts (2001B) present Danish fire loss data of 2493 claims over the period of 1980 to 1993. Example 6.2.9 on pages 298, 302 and 303 of Embrechts, Kluppelberg, and Mikosch (1997) analyzes an industrial fire data set with 8,043 samples

5) Windstorm Damage Insurance Claims

EVT techniques can be used in the prediction of meteorological events to model covariables such as time and weather. Rootzén and Tajvidi (1997) apply extreme value theory to insurance claims from windstorm losses encountered by the Swedish insurance group Länsförsäkringar over the 12-year period from 1982 to 1993. The individual claims were grouped into 46 storm events. Rootzén and Tajvidi (1997) focused on the modeling of the extreme losses, the total claim amount, and trend detection of such windstorm loss events, together with covariate analysis of the potential influences of various environmental factors. McNeil and Saladin (2000) developed scenarios for forecasting future extreme losses by applying the peaks-over-thresholds method to a portfolio of windstorm losses. They concentrated on modeling the trends in both frequency and severity of loss.

6) Non-Life Insurance Claims

EVT methods are used to model claim size distributions in non-life insurance (see Beirlant and Teugels (1992), Beirlant, Vyncker and Teugels (1994)).

7) Reinsurance

The area of reinsurance is probably the most important application of *EVT* in the field of insurance, especially in the area of contracts written on single events or a few events involving upper or catastrophic layers. *EVT* can be used to investigate the role of the extremal claims in an insurance company's portfolio through the use of reinsurance treaties. Recall that reinsurance is the safeguard that insurance companies have against excessive claims that can lead to insolvency of the company. Reinsurance treaties usually cover the excess large catastrophic claim sizes over a certain high threshold and this is specifically one of the problems where *EVT* methods using the generalized Pareto distribution are appropriate. *EVT* can be used to estimate premiums and choices for the layers of reinsurance in an excess-of-loss treaty, which insurances against a single event or few events. The most popular reinsurance treaties are the excess-of-loss reinsurance and the stop-loss reinsurance. *EVT* modeling of large or catastrophic claims that are from heavy tailed claims distributions can arise in the following reinsurance treaties: *ECOMOR* reinsurance, excess-of-loss reinsurance, facultative reinsurance, quota-share reinsurance, stop-loss reinsurance, and surplus reinsurance. Largest-claim reinsurance treaty or claims at high quantiles reinsurance treaties are referred to as reinsurance treaties of the extreme value type. For example,

the *ECOMOR* reinsurance contract commits the reinsurer to an obligation to cover the excesses above the k -th largest claim for a specific premium. Basically, *ECOMOR* forms an excess-of-loss treaty with a specific random high quantile retention. The *ECOMOR* reinsurance contract anticipated the Hill-type statistic by about 25 years (see Thépaut (1950)). Also the excess-of-loss reinsurance protection is arranged in “layers” in order for losses in the different loss ranges to be covered by different reinsurance treaties. The priority or attachment point for a layer can be reformulated as a high quantile of the claims distributions. In this context, tail estimation methods can be used to estimate the risk involved within a certain layer (see Chapter 5 of Beirlant, Teugels, and Vyncker (1996), Example 6.1.5 on page 287 and pages 517-519 of Embrechts, Kluppelberg, and Mikosch (1997), and sections 1.5 and 3.3 of Rolski, Schmidli, Schmidt, and Teugels (1999)).

8) Longevity of humans

Health care providers and pension funds are interested in the modeling of payout frequency and magnitude as a function of the distribution of the life spans of humans. The Gompertz distribution is the classical life-span distribution (see section 14.1 of Reiss and Thomas (2000) with E. Kaufmann)).

9) Stress Testing of Insurance Portfolios against Catastrophic Events

EVT methods are used to stress test insurance portfolios against a set of catastrophic events (see McNeil and Saladin (2000)).

13 Issues Inherent to Extreme Value Theory

The *EVT* approach to market risk measurement may seem like a straight forward application of theory, but the user must be warned that the devil is in the details, which is the balance between art and science of tail index estimation of infrequently occurring tail events. The tail index is parametrically using the maximum likelihood method or nonparametrically using estimators such as the Hill estimator. The main issue inherent to tail index estimation occurs when selecting the high threshold that defines where the tail begins. If the threshold is too close to the upper bound of the loss data, then there will not be enough samples for the tail index estimator to exhibit sufficient statistical power. If the threshold is too close to the centre, the *GPD* approximation for the conditional excess distribution may not be justifiable.

13.1 Difficulties Inherent to the Application of Extreme Value Theory

Some of the important aspects and faults arising when applying extreme value theory to real time series are summarized below (see pages 38-39 of Castillo (1994)):

- 1) There are two different groups of families of distributions for extreme values: The Fréchet, Gumbel and Weibull distributions for maxima, and the Fréchet, Gumbel and Weibull distributions for minima. Each family must be used in the proper corresponding situation. The use of the wrong family can lead to spurious results and bad designs.
- 2) Frequently methods such as the estimation of the tail index from data are fitted on inadequate probability paper using $Q - Q$ plots. This could be using probability paper for maxima when estimating properties of the minimum or visa versa.
- 3) The use of data only from the tail region when dealing with extreme value problems. Data from outside the tail of interest can only bias the estimate of the tail region. This highlights the difficulties in determining the optimal tail threshold, i.e. where the tail starts in an order collection of data.
- 4) The use of the maxima or minima instead of the exceedances of a given threshold value can lead to a waste of important information.
- 5) The choice of the cumulative distribution function, in the context of *EVT* procedures, is often made incorrectly by fitting families of distributions to the entire sample, and then using the relationships for the distribution of the maxima (minima) in terms of power of the distribution of the original *IID* random variables. This is erroneous, and the correct procedure is to base the estimation on only tail data using a model especially designed for the tail region.
- 6) Random variables with bounded right endpoints are wrong approximated for extremes using the Fréchet distribution, and random variables with unbounded right endpoints are wrong approximated for extremes using Weibull distribution
- 7) If the distribution function of the underlying *IID* random variables is known, then the use of the appropriate limiting distribution is not justified because, in these situations, the exact distribution of the maxima and minima are known.

- 8) The speed of convergence to the limiting distribution can be very slow, implying that very large sample sizes are needed for a good approximation.
- 9) Discussion about the choice of the Gumbel versus the Weibull or the Gumbel versus the Fréchet distributions for approximating the extremes does not make practical sense. This is because the Gumbel distributions can be approximated, as closely as desired, by either the Fréchet or Weibull distributions. This implies when in doubt use either Fréchet or Weibull distributions.
- 10) Least-squares methods must be weighted procedures with an adequate choice of weights. That is large weights for the order statistics in the tail of interest, and small or null weights for the rest of the data.
- 11) The assumption that the random variables are independent and identically distributed is usually not valid in practice. Even in case where the *IID* assumption is not valid, the asymptotic independence assumption has some theoretical justifications. It is important to determine the critical threshold value, above which extrapolation of the theoretical justifications to where real cases is valid, is crucial for the purpose of the design of an *EVT* analysis.
- 12) Extreme value distributions are very useful in problems where extrapolations from the extreme ends of the tails are needed.

13.2 Advantages to the Use of EVT in Applications

The *EVT* approach to risk measurement has the following advantages over the classical risk measurement procedures (see Diebold, Schuermann and Stroughair (2000)):

- 1) *EVT directly focuses on the extreme values, i.e. the tail region of the distribution.*

That is, the *EVT* estimation models are specifically tailored for the tail regions of the distribution rather than the center regions. Note models for the tail regions follow a convergence-to-types theorem for the maximum using the extreme value distributions in an analogous manner that models for center regions follow a convergence-to-types theorem for the stable distributions. The special case of the latter is the central limit theorem where finite variance random variables converge to the normal distribution. The convergence of the distribution of the maximum is to one of three different extreme value distributions: the Gumbel, Fréchet and Weibull distributions.

- 2) ***EVT provides a solid methodological basis for the estimation of VaR for the tail regions of a general profit/loss distribution.***

EVT allows for distributions with much heavier tails than the normal distribution to be candidate models for the tail region of the loss distribution. *EVT* produces a simple, but general approximation for *VaR*. *EVT* allows for the expected loss beyond *VaR* to be estimated. The use of the *EVT* approach to *VaR* is robust, flexible and straightforward.

- 3) ***EVT uses a justifiable and reasonable functional form for the tail region based on a priori considerations.***

That is, the *EVT* method is parametric, implying that out-of-sample *VaR* estimates are possible for extremely large losses where the probabilities of occurrence are extremely small.

- 4) ***EVT does not impose a particular form on the underlying or parent returns distribution.***

Instead, *EVT* allows the tail region of the loss distribution to be from a class of distributions called the *general Pareto distribution*. Basically, *EVT* allows the data to speak for themselves in the fitting of the loss distributions. *EVT* can considerably reduce model risk when used properly. For example, the normal distribution may fit the tail regions badly if there is the presence of a number of extreme events.

- 5) ***EVT focuses on the extreme events and thus allowing for event risk to be explicitly taken into account.***

With conditional distribution models such as *GARCH* or exponentially-weighted moving averages (*EWMA*) processes the large unexpected market shocks cannot be considered.

- 6) ***EVT produces method that is a simple, but general approximation for VaR estimates.***

- 7) ***EVT allows estimates of the expected loss beyond VaR to be formulated.***

- 8) ***The EVT VaR approach is robust, flexible and straightforward to use.***

13.3 Drawbacks to the Use of EVT in Applications

Some of the drawbacks to the use of *EVT* include:

1) *The scarcity of observations from the tail region of a distribution.*

Usually the number of observations from the tail region of a financial data record is small compared to the sample size required to estimate the form of the tail region with significant power.

2) *The problems inherent to choosing the optimal threshold or tail cut-off point.*

There is a bias-variance tradeoff issues when deciding the number of upper-order statistics to include in the tail region.

3) *The difficulty is validating the required underlying assumptions.*

For example, how does one validate the common assumption that the distribution of the maximum of the original data falls in the maximum domain of attraction of the Fréchet distribution? The validation of this assumption is important since as a consequence, the survival function can be modeled as a power law function.

4) *Serial dependencies in the data violate the usual assumption of independently distributed random variables.*

The use of *EVT* with time series that have dependent observations requires an anti-clustering condition on the extremes. The anti-clustering condition rules out volatility clustering that is frequently exhibited by financial returns series.

5) *The problem of estimating low-frequency events using short-historical records.*

This is important for macroeconomic time series that exhibit long-term mean reversion such as real output growth rates and long-run mean reversion in asset returns, implying long time lags are present in the time series.

6) *The difficulty in the extensions to the multivariate case of extreme value theory.*

The core of multivariate *EVT* has been established, but there are methodological issues remaining to be solved. Multivariate *EVT* is useful if there are only a few risk factors to be estimated and the concern is in the modeling the joint extreme values to describe

joint tail regions. In higher dimensional problems there are usually too many parameters and tails associated with multivariate *EVT* distributions to estimate. This is where the “*curse of dimensionality*” comes into play.

13.4 Summarizing the Advantages and Drawbacks of EVT

Although there are numerous drawbacks to the use of *EVT* procedures, the following comments of *EVT* researchers Richard Smith and Jonathan Tawn summarize the advantages and drawbacks and points towards the inherent value of the extreme value theory in applications (see page xvi of Embrechts (2000A)):

“There is always going to be an element of doubt as one is extrapolating into areas one doesn’t know about. But what EVT is doing is making the best use of whatever data you have about extreme phenomena.” (Richard L. Smith)

and

“The key message is the EVT cannot do magic – but it can do a whole lot better than empirical curvefitting and guesswork. My answer to the sceptics is that if people aren’t given well-founded methods like EVT, they’ll use dubious ones instead.” (Jonathan A. Tawn)

Finally, Paul Embrechts, who is the leading pioneer of *EVT* applications in integrated risk management, ends his introductory *EVT* modeling paper with the following comments (see page 274 of Embrechts (2000B)):

“The modeling of extremes is a difficult subject: by definition, data is sparse on an extreme event, and often predictions or estimates have to be made at the edge of or even beyond the available data. ... EVT does not offer a magical way out of these problems. It does, however, give better results than other approaches, many of which lack a theoretical justification.”

14 References

- [A] Ahn, H., M. Bouabci and A. Penaud (February 2000), “Tailor-made for Tails,” RISK, Vol. 13, pp. 95-98.
- [A] Artzner, P., F. Delbaen, J-M. Elber, and D. Heath (July 2000), “Coherent Measures of Risk,” Mathematical Finance, Vol. 9, No. 3, pp. 203-228.
- [A] Arvanitis, A. and J. Gregory (2001), Credit: The Complete Guide to Pricing, Hedging and Risk Management, London, England, RISK Books.

- [B] Balestra, A. (April 2000), "Quantification of Operational Risk," The Arab Bank Review: Bank Management, Vol. 2, No. 1.
- [B] Balkema, A.A. and L. de Haan (1974), "Residual Lifetime at Great Age," Annals of Probability, Vol. 2, pp. 792-804.
- [B] Bassi, F. P. Embrechts, and M. Kafetzaki (1998), "Risk Management and Quantile Estimation," A Practical Guide to Heavy Tail Distributions: Statistical Techniques for Analysing Heavy Tailed Distributions, edited by R. Adler, R. Feldman, and M. Taqqu, Boston, Massachusetts: Birkhauser Boston Inc., pp. 111-130
- [B] Beirlant, J. and J.L. Teugels (1992), "Modeling Large Claims in Non-Life Insurance," Insurance: Mathematics and Economics, Vol. 11, pp. 17-29.
- [B] Beirlant, J., P. Vyncker and J.L. Teugels (1994), "Extremes in Non-Life Insurance," Proceedings from a Conference on Extreme Value Theory and Applications, edited by J. Galambos, J. Lechner, and E. Simiu, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 489-510.
- [B] Beirlant, J., J.L. Teugels, and P. Vyncker (1996), Practical Analysis of Extreme Values, Leuven, Belgium: Leuven University Press.
- [B] Beirlant, J., P. Vyncker and J.L. Teugels (1996), "Tail Index Estimation, Pareto Quantiles Plots, and Regression Diagnostics," Journal of the American Statistical Association, Vol. 91, pp. 1659-1667
- [B] Borkovec, M. and C. Klüppelberg (2001), Extremal Behaviour of Diffusion Models in Finance, Working Paper, Munich, Germany: Center for Mathematical Sciences, Munich University of Technology.
- [B] Brockhaus, O. M. Farkas, A. Ferraris, D. Long, and M. Overhaus (2000), Equity Derivatives and Market Risk Models, London, England, RISK Publications.
- [B] Broussard, J.P. (2001), "Extreme-value and Margin Setting With and Without Price Limits," Quarterly Review of Economics and Finance, Vol. 41, pp. 365-385.
- [B] Broussard, J-P. and G. G. Booth (1998), "The Behavior of Extreme Values in Germany's Stock Index Futures: An Application to Intra-daily Margin Setting," European Journal of Operational Research, Vol. 104, pp. 393-402.
- [B] Byström, H. N. E. (March 1999), Managing Extreme Risks in Tranquil and Volatile Markets Using Conditional Extreme Value Theory, Working Paper, Department of Economics, Lund University: Lund, Sweden.
- [C] Caserta, S., J. Danielsson and C.G. de Vries (2001), Abnormal Returns, Risk, and Options in Large Data Sets, Tinbergen Institute Discussion Paper, Amsterdam, The Netherlands.
- [C] Castillo, E. (1994), "Extremes in Engineering Applications," Proceedings of the Conference on Extreme Values and Applications, Gaithersburg, Maryland, 1993, Edited by J. Galambos, J. Lechner, and E. Simiu, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 15-42.
- [C] Castillo, E. (1988), Extreme Value Theory in Engineering, New York, New York: Academic Press Inc..
- [C] Ceske, R. and J. Hernández (November 1999), "Where Theory Meets Practice," RISK, Operational Risk Supplement, Vol. 12, pp. 17-20.
- [C] Cleveland, W.S. (1985), The Elements of Graphing Data, Belmont, California: Duxbury Press of Wadsworth Publishing Company.

- [C] Cruz, M. (November 1999), "Taking Risk to the Market," RISK, Operational Risk Supplement, Vol. 12, pp. 21-24.
- [C] Cruz, M., R. Coleman, and G. Salkin (2000), "Modelling and Measuring Operational Risk," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, pp. 204-212.
- [D] Dacorogna, M.M., R. Gençay, U.A. Müller, R.B. Olsen and O.V. Pictet (2001), An Introduction to High-Frequency Finance, San Diego, California: Academic Press Inc..
- [D] Danielsson, J. and C. G. de Vries (1997), "Tail Index and Quantile Estimation with Very High Frequency Data," Journal of Empirical Finance, Vol. 4, pp. 241-257.
- [D] Danielsson, J. and C.G. de Vries (2000), "Value-at-Risk and Extreme Returns," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, pp. 85-106.
- [D] Danielsson, J., P. Hartmann and C.D. de Vries (January 1998), "The Cost of Conservatism," RISK, Vol. 11, pp. 101-103.
- [D] Davison, A. C. and R. L. Smith (1990), "Models for Exceedances of High Thresholds," Journal of the Royal Statistical Society, Series B, Vol. 52, No. 3, pp. 393-442.
- [D] DeRosa, D. F. (2001), In Defense of Free Capital Markets: The Case Against a New International Financial Architecture, Princeton, New Jersey: Bloomberg Press
- [D] Dekkers, A.L.M. and L. de-Haan (1989), "On the Estimation of the Extreme-Value Index and Large Quantile Functions ," The Annals of Statistics, Vol. 17, No. 4, pp. 1795-1832.
- [D] De Haan, L. (1976), "Sample Extremes: An Elementary Introduction," Statistica Neerlandica, Vol. 30, pp. 161-172.
- [D] De Haan, L. (1990), "Fighting the ARCH-enemy with Mathematics," Statistica Neerlandica, Vol. 44, pp. 45-68.
- [D] De Haan, L. (1994), "Extreme Value Statistics," Proceedings from a Conference on Extreme Value Theory and Applications, edited by J. Galambos, J. Lechner, and E. Simiu, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 93-122.
- [D] De Haan, L. and S.I. Resnick (1980), "A Simple Estimate for the Index of a Stable Distribution," Journal of the Royal Statistical Society: Series B, Vol. 42, pp. 83-87.
- [D] De Haan, L. and U Stadtmuller (1996), "Generalized Regular Variation of Second Order," Journal of the Australian Mathematical Society - Series A, Vol. 61.
- [D] De Haan, L., D. W. Jansen, K. Koedijk, and C. G. De Vries (1994), "Safety First Portfolio Selection, Extreme Value Theory and Long Run Asset Risks," Proceedings from a Conference on Extreme Value Theory and Applications, edited by J. Galambos, J. Lechner, E. Simiu and N. Macri, Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 471-487.
- [D] De Vries, C. G., S. Caserta, R-D Reiss, and M. Thomas (2000), "Extreme Returns in Asset Prices," Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields, 2nd Edition, Main Authors: R-D. Reiss and M. Thomas Basel, Switzerland, Birkhäuser Verlag, Chapter 13, pp. 309-339.
- [D] Diebold, F. X., T. Schuermann and J. D. Strouhair (2000), "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, pp. 51-57.
- [D] DuMouchel, W. H. (1983), "Estimating the Stable Index α in Order to Measure Tail Thickness: A Critique," The Annals of Statistics, Vol. 11, No. 4, pp. 1019-1031.

- [E] Eberlein, E., U. Keller, and K. Prause (1998), "New Insights into Smile, Mispricing and Value-at-Risk; the Hyperbolic Model," Journal of Business, Vol. 71, pp. 371-405.
- [E] Embrechts, P. (2000A), "Introduction," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, pp. xv-xxiv.
- [E] Embrechts, P. (2000B), "What is Wrong with the Bell Curve," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, pp. xxv-xxvii.
- [E] Embrechts, P. (editor) (2000C), Extremes and Integrated Risk Management, London, England, RISK Books.
- [E] Embrechts, P. (2001), "Difficult Calls in Judging Risk Extremes," Mastering Risk - Volume 1: Concepts, edited by J. Pickford, London, Great Britain, Financial Times Prentice Hall, pp. 269-274.
- [E] Embrechts, P. and N. Veraverbeke (1982), "Estimates for the Probability of Ruin with Special Emphasis on the Possibility of Large Claims," Insurance: Mathematics and Economics, Vol. 1, pp. 55-72.
- [E] Embrechts, P., C. Kluppelberg, and T. Mikosch (1997), Modeling Extreme Events for Insurance and Finance, Berlin: Germany, Springer-Verlag.
- [E] Embrechts, P., S. Resnick, and G. Samorodnitsky (January 1998), "Living on the Edge," RISK, Vol. 11, No. 1, pp. 96-100.
- [F] Fama, E. F. (1963), "Mandelbrot and the Stable Paretian Hypothesis," Journal of Business, Vol. 36, pp. 420-429.
- [F] Fisher, R.A. and H.C. Tippett (1928), "Limiting Forms of the Frequency Distributions of the Largest or Smallest Member of a Sample," Proceedings of the Cambridge Philosophy Society, Vol. 24, pp. 180-190.
- [F] Fréchet, M. (1927), "Sur la Loi de Probabilité de l'écart Maximum," Annales de la Société Polonaise de Mathématique (Cracow), Vol. 6, pp. 93-117.
- [G] Gavin, J. (2000), "Extreme Value Theory: An Empirical Analysis of Equity Risk," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, pp. 114-124.
- [G] Geman, H. (editor) (1999), Insurance and Weather Derivatives: From Exotic Options to Exotic Underlyings, London, England, RISK Books.
- [G] Gnedenko, B.V. (1943), "Sur la Distribution Limite de Terme Maximum d'un Séries Aléatoire," The Annals of Mathematics, Vol. 44, pp. 423-453.
- [G] Gumbel, E.J. (1958), Statistics of Extremes, New York, New York: Columbia University Press..
- [H] Hill, B.M. (1975), "A Simple General Approach to Inference About the Tail of a Distribution," The Annals of Statistics, Vol. 3, pp. 1163-1174.
- [H] Hols, M. C. A. B. and C. G. De Vries (1991), "The Limiting Distribution of Extremal Exchange Rate Returns," Journal of Applied Econometrics, Vol. 6, pp. 287-302.
- [H] Hosking, J. R. M. and J. R. Wallis (1987), "Parameter and Quantile Estimation for the Generalized Pareto Distribution," Technometrics, Vol. 29, No. 3, pp. 339-349.
- [H] Huisman, R., K.G. Koedijk, C.J.M. Kool, and F. Palm (1998), "The Fat-Tailedness of FX Returns," Working Paper, Limburg Institution, Maastricht University, pp. 1-31.
- [KK] Koedijk, K.G. and C.J.M. Kool (1992), "Tail Estimates of East European Exchange Rates ," Journal of Business and Economic Statistics, Vol. 10, No., pp. 83-96.

- [K] Koedijk, K., M. Schafgans and C.G. de Vries (1990), "The Tail Index Estimation for Exchange Rate Returns," Journal of International Economics, Vol. 29, pp. 93-108.
- [K] Koedijk, K.G., P.A. Stork and C. G. De Vries (1992), "Differences between Foreign Exchange Rate Regimes: The Views from the Tails," Journal of International Money and Finance, Vol. 11, pp. 462-473.
- [K] Kofman, P. and C.G. de Vries (1989), "Potato Futures Returns: A Tail Investigation," The Review of Futures Markets, Vol. 8, pp. 244-258.
- [J] 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, Vol. 73, No. 1, pp. 18-24.
- [J] Jenkinson, A.F. (1955), "The Frequency Distribution of Annual Maximum (or Minimum) Values of Meteorological Elements," Quarter Journal of The Royal Meteorological Society, Vol. 81, pp. 158-171.
- [J] Jenkinson, A.F. (1969), "Estimation of Maximum Floods," World Meteorological Organization, Technical Note, No. 98, Chapter 5, pp. 183-257.
- [J] Jondeau, E. and M. Rockinger (April 1999), "The Tail Behavior of Stock Returns: Emerging versus Mature Markets," Working Paper, Department of Finance, HEC-School of Management, Jouy-en-Josas, France.
- [K] Kearns, P. and A. R. Pagan (1997), "Estimating the Tail Density Index for Financial Time Series," Review of Economics and Statistics, Vol. 79, pp. 171-175.
- [K] Këllezi, E. and M. Gilli (2000), Proceedings of the Computational Finance 2000 Conference: 31 May – 2 June 2000, London Business School, Dordrecht, The Netherlands: Kluwer Academic Publishers.
- [K] Kuan, C.H. and N.J. Webber (1998), "Valuing Interest Rate Derivatives Consistent with a Volatility Smile," Working Paper, University of Warwick, pp. 1-30.
- [L] Leadbetter, M.R., G. Lindgren, and H. Rootzén (1983), Extreme and Related Properties of Random Sequences and Series, New York, New York, Springer-Verlag New York Inc..
- [L] Longin, F. (1994), "Optimal Margin Level in Futures Markets: A Parametric Extreme-Based Method," IFA Working Paper 192-94, London Business School, London England.
- [L] Longin, F.M. (1996), "The Asymptotic Distribution of Extreme Stock Market Returns," Journal of Business, Vol. 69, No. 3, pp. 383-408.
- [L] Longin, F. M. (2000), "From Value at Risk to Stress Testing: The Extreme Value Approach," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, Chapter 11, pp. 125-148.
- [L] Loretan, M. and P.C.B. Phillips (1994), "Testing the Covariance Stationarity of Heavy-tailed Time Series," Journal of Empirical Finance, Vol. 1, pp. 211-248.
- [L] Lucas, A., P. Klaassen, P. Spreij and S. Straetmans (2001), Tail Behavior of Credit Loss Distributions for General Latent Factor Models, Tinbergen Institute Discussion Paper TI 2001-023/2, Amsterdam, The Netherlands.
- [M] Mandelbrot, B.B. (1963), "The Variation of Certain Speculative Prices," Journal of Business, Vol. 36, pp. 394-419.
- [M] Marakov, V. (1997), "Value-at-Risk and Beyond: The Comprehensive Measurement of Market Risk," Risk Management for Financial Institutions: Advances in

- Measurement and Control, Edited by A. Aggarwal and S. Watson, London, England, RISK Publications, pages pp. 38-54.
- [M] McNeil, A. J. (1998), "Calculating Quantile Risk Measures for Financial Return Series using Extreme Value Theory," Working Paper, Department of Mathematics, ETHZ, CH-8092, Zurich, Switzerland.
- [M] McNeil, A.J. (2000), "Extreme Value Theory for Risk Managers," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, pp. 1-18.
- [M] McNeil, A. J. and R. Frey (2000), "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach," Working Paper, Department of Mathematics, ETHZ, CH-8092, Zurich, Switzerland.
- [M] McNeil, A. J. and T. Saladin (1997), "The Peaks-over-Thresholds Methods for Estimating High Quantiles of Loss Distributions," Working Paper, Department of Mathematics, ETHZ, CH-8092, Zurich, Switzerland.
- [M] McNeil, A.J. and T. Saladin (2000), "Developing Scenarios for Future Extreme Losses Using the Peaks-over-Thresholds Method," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, Chapter 18, pp. 253-267.
- [M] Medova, E. (November 2000), "Measuring Risk by Extreme Values," RISK, Operational Risk Special Report, Vol. 13, pp. S20-S25.
- [N] Neftci, S.N. (Spring 2000), "Value at Risk Calculations, Extreme Events, and Tail Estimation," The Journal of Derivatives, Vol. 7, No. 3, pp. 23-37.
- [O] Ong, M.K. (1999), Internal Credit Risk Models: Capital Allocation and Performance Measurement, London, England, RISK Books.
- [P] Parkinson, M. (1980), "The Extreme Value Method for Estimating the Variance of the Rate of Return," Journal of Business, Vol. 53, No. 1, pp. 1-65.
- [P] Phoa, W. (Spring 1999), "Estimating Credit Spread Risk using Extreme Value Theory," Journal of Portfolio Management, Vol. 25, pp. 69-73.
- [P] Pictet, O. V., M. M. Dacorogna and U. A. Müller (1998), "Hill, Bootstrap and Jackknife Estimators for Heavy Tails," A Practical Guide to Heavy Tails: Statistical Techniques and Applications, edited by R. J. Adler, R. E. Feldman and M. S. Taqqu, Boston, Massachusetts: Birkhäuser Boston Inc., pp. 283-310.
- [P] Pickands III, J. (1975), "Statistical Inference Using Extreme Order Statistics," The Annals of Statistics, Vol. 3, No. 1, pp. 119-131.
- [P] 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, Vol. 18, pp. 853-870.
- [Q] Quintos, C.E., Fan, Z-H. and P.C.B. Phillips (2001), "Structural Change in Tail Behavior and the Asian Financial Crisis", Working Paper, New York University, Finance Department; New York, New York.
- [R] Radtke, M, R-D Reiss, and M. Thomas (2000), "The Impact of Large Claims on Actuarial Decisions," Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields, 2nd Edition, main authors are R-D. Reiss and M. Thomas Basel, Switzerland, Birkhäuser Verlag, Chapter 12, pp. 279-308.
- [R] Reiss, R-D and M. Thomas (2000), Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields, 2nd Edition, Basel, Switzerland, Birkhäuser Verlag.

- [R] Resnick, S.I. (1987), Extreme Values, Regular Variation, and Point Processes, New York, New York: Springer-Verlag.
- [R] Rootzén, H. and C. Klüppelberg (May 27, 1999), A Single Number Can't Hedge Against Economic Catastrophes, Working Paper, Munich, Germany: Center for Mathematical Sciences, Munich University of Technology.
- [R] Rootzén, H. and N. Tajvidi (2000), "Extreme Value Statistics and Wind Storm Losses: A Case Study," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, Chapter 16, pp. 215-234
- [S] Smith, R.L. (1985), Maximum Likelihood Estimation in a Class of Nonregular Cases. Biometrika, Vol. 72, pp 67-92.
- [S] Smith, R.L. (1987), "Estimating Tails of Probability Distributions," The Annals of Statistics, Vol. 15, No. 3, pp. 117.
- [S] Smith, R.L. (1986), "Extreme Value Theory and the r Largest Annual Events," Journal of Hydrology, Vol. 86, 27-43.
- [S] Smith, R.L. (1989), "Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone," Statistical Science, Vol. 4, pp. 367-393.
- [S] Smith, R.L. (2000), "Measuring Risk with Extreme Value Theory," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, Chapter 2, pp. 19-35.
- [S] Starica, C. (1998), "Estimation of the Extreme Value Distribution for Constant Conditional Correlation Models," HFDF - II, Second International Conference on High Frequency Data in Finance, April 1-3, 1998: Volume 2, edited by Olsen and Associates, Zurich, Switzerland: Olsen and Associates.
- [S] Starica, C. (2000), "Multivariate Extremes for Foreign Exchange Data," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, pp. 169-185.
- [R] Rolski, T., H. Schmidli, V. Schmidt, and J. Teugels (1999), Stochastic Processes in Insurance and Finance, New York, New York: John Wiley and Sons Inc..
- [S] Smith, R.L. and D.J. Goodman (2000), "Bayesian Risk Analysis," Extremes and Integrated Risk Management, edited by P. Embrechts, London, England, RISK Books, Chapter 17, pp. 236-251.
- [T] Tawn, J.A. (1988), "An Extreme Value Theory Model for Dependent Observations," Journal of Hydrology, Vol. 101, pp. 227-250.
- [T] Teugels, J.L. (1970), "Extremes Values in Insurance Mathematics," Statistical Extremes and Applications, edited by J. Tiago de Oliveira, Dordrecht, The Netherlands: D. Reidel Publishing Company.
- [T] Teugels, J.L. and N. Veraverbeke (1973), Cramér-type Estimates for the Probability of Ruin, C.O.R.E. Discussion Paper No. 7316.
- [T] Thépaut A. (1950), "Une Nouvelle Forme de reassurance. Le Traité d'Excédent du Coût Moyen relative (ECOMOR)," Bulletin Trimestriel de l'Institut des Actuaries Français, Vol. 49.
- [V] Vilasuso, J. and D. Katz (2000), "Estimates of the Likelihood of Extreme Returns in International Stock Markets," Journal of Applied Statistics, Vol. 27, No. 1, pp. 119-130.
- [V] von Mises, R. (1936), "La Distribution de la plus Grande de n Valeurs," Rev. of Math. Union Interbalcanique, Vol. 1, pp. 141-160.

- [W] Weibull, W. (1939), "A Statistical Theory of the Strength of Materials," Proc. Royal Soc. Swedish Inst. Eng Res., No. 151.
- [WW] Weissman, I. (1991), "Estimation of Parameters and Quantiles based on the k largest Observations," Journal of the American Statistical Association, Vol. 73, pp. 812-815.
- [W] Wiggins, J. B. (1991), "Empirical Tests of the Bias and Efficiency of the Extreme-Value Variance Estimator for Common Stocks," Journal of Business, Vol. 64, No. 3, pp. 417-432.
- [Z] Zangari, P. (July, 1997), "Catering for an Event," RISK, Vol. 10, No. 7, pp. 34-36.

15 Figures

