Extreme Value Theory for Risk Managers

Alexander J. McNeil *
Departement Mathematik
ETH Zentrum
CH-8092 Zürich
Tel: +41 1 632 61 62
Fax: +41 1 632 15 23
mcneil@math.ethz.ch

17th May 1999

Abstract

We provide an overview of the role of extreme value theory (EVT) in risk management (RM), as a method for modelling and measuring extreme risks. We concentrate on the peaks-over-threshold (POT) model and emphasize the generality of this approach. Wherever the tail of a loss distribution is of interest, whether for market, credit, operational or insurance risks, the POT method provides a simple tool for estimating measures of tail risk. In particular we show how the POT method may be embedded in a stochastic volatility framework to deliver useful estimates of Value-at-Risk (VaR) and expected shortfall, a coherent alternative to the VaR, for market risks. Further topics of interest, including multivariate extremes, models for stress losses and software for EVT, are also discussed.

1 A General Introduction to Extreme Risk

Extreme event risk is present in all areas of risk management. Whether we are concerned with market, credit, operational or insurance risk, one of the greatest challenges to the risk manager is to implement risk management models which allow for rare but damaging events, and permit the measurement of their consequences.

This paper may be motivated by any number of concrete risk management problems. In market risk, we might be concerned with the day to day determination of the Value-at-Risk (VaR) for the losses we incur on a trading book due to adverse market movements. In credit or operational risk management our goal might be the determination of the risk capital we require as a cushion against irregular losses from credit downgradings and defaults or unforeseen operational problems.

Alongside these financial risks, it is also worth considering insurance risks; the insurance world has considerable experience in the management of extreme risk and many methods which we might now recognize as belonging to extreme value theory have a long

*Alexander McNeil is Swiss Re Research Fellow at ETH Zürich and gratefully acknowledges the financial support of Swiss Re.

history of use by actuaries. In insurance a typical problem might be pricing or building reserves for products which offer protection against catastrophic losses, such as excess-of-loss (XL) reinsurance treaties concluded with primary insurers.

Whatever the type of risk we are considering our approach to its management will be similar in this paper. We will attempt to *model* it in such away that the possibility of an extreme outcome is addressed. Using our model we will attempt to *measure* the risk with a measurement which provides information about the extreme outcome. In these activities extreme value theory (EVT) will provide the tools we require.

1.1 Modelling Extreme Risks

The standard mathematical approach to modelling risks uses the language of probability theory. Risks are random variables, mapping unforeseen future states of the world into values representing profits and losses. These risks may be considered individually, or seen as part of a stochastic process where present risks depend on previous risks. The potential values of a risk have a *probability distribution* which we will never observe exactly although past losses due to similar risks, where available, may provide partial information about that distribution. Extreme events occur when a risk takes values from the *tail* of its distribution.

We develop a model for a risk by selecting a particular probability distribution. We may have estimated this distribution through *statistical analysis* of empirical data. In this case EVT is a tool which attempts to provide us with the best possible *estimate* of the tail area of the distribution. However, even in the absence of useful historical data, EVT provides guidance on the kind of distribution we should select so that extreme risks are handled conservatively.

1.2 Measuring Extreme Risks

For our purposes, measuring a risk means summarising its distribution with a number known as a risk measure. At the simplest level, we might calculate the mean or variance of a risk. These measure aspects of the risk but do not provide much information about the extreme risk. In this paper we will concentrate on two measures which attempt to describe the tail of a loss distribution - VaR and expected shortfall. We shall adopt the convention that a loss is a positive number and a profit is a negative number. EVT is most naturally developed as a theory of large losses, rather than a theory of small profits.

VaR is a high *quantile* of the distribution of losses, typically the 95th or 99th percentile. It provides a kind of upper bound for a loss that is only exceeded on a small proportion of occasions. It is sometimes referred to as a confidence level, although this is a misnomer which is at odds with standard statistical usage.

In recent papers Artzner, Delbaen, Eber & Heath (1997) have criticized VaR as a measure of risk on two grounds. First they show that VaR is not necessarily subadditive so that, in their terminology, VaR is not a *coherent* risk measure. There are cases where a portfolio can be split into sub-portfolios such that the sum of the VaR corresponding to the sub-portfolios is smaller than the VaR of the total portfolio. This may cause problems if the risk-management system of a financial institution is based on VaR-limits for individual books. Moreover, VaR tells us nothing about the potential size of the loss that exceeds it.

Artzner et al. propose the use of *expected shortfall* or *tail conditional expectation* instead of VaR. The tail conditional expectation is the *expected* size of a loss that exceeds VaR and is coherent according to their definition.

1.3 Extreme Value Theory

The approach to EVT in this paper follows most closely Embrechts, Klüppelberg & Mikosch (1997); other recent texts on EVT include Reiss & Thomas (1997) and Beirlant, Teugels & Vynckier (1996). All of these texts emphasize applications of the theory in insurance and finance although much of the original impetus for the development of the methods came from hydrology.

Broadly speaking, there are two principal kinds of model for extreme values. The oldest group of models are the *block maxima* models; these are models for the largest observations collected from large samples of identically distributed observations. For example, if we record daily or hourly losses and profits from trading a particular instrument or group of instruments, the block maxima method provides a model which may be appropriate for the quarterly or annual maximum of such values. We see a possible role for this method in the definition and analysis of *stress losses* (McNeil 1998) and will return to this subject in Section 4.1.

A more modern group of models are the *peaks-over-threshold* (POT) models; these are models for all large observations which exceed a *high* threshold. The POT models are generally considered to be the most useful for practical applications, due to their more efficient use of the (often limited) data on extreme values. This paper will concentrate on such models.

Within the POT class of models one may further distinguish two styles of analysis. There are the semi-parametric models built around the Hill estimator and its relatives (Beirlant et al. 1996, Danielsson, Hartmann & de Vries 1998) and the fully parametric models based on the generalized Pareto distribution or GPD (Embrechts, Resnick & Samorodnitsky 1998). There is little to pick and choose between these approaches - both are theoretically justified and empirically useful when used correctly. We favour the latter style of analysis for reasons of simplicity - both of exposition and implementation. One obtains simple parametric formulae for measures of extreme risk for which it is relatively easy to give estimates of statistical error using the techniques of maximum likelihood inference.

The GPD will thus be the main tool we describe in this paper. It is simply another probability distribution but for purposes of risk management it should be considered as equally *important* as (if not more important than) the Normal distribution. The tails of the Normal distribution are too thin to address the extreme loss.

We will not describe the Hill estimator approach in this paper; we refer the reader to the references above and also to Danielsson & de Vries (1997).

2 General Theory

Let $X_1, X_2, \ldots$ be identically distributed random variables with unknown underlying *distribution function* $F(x) = P\{X_i \leq x\}$. (We work with distribution functions and not densities.) The interpretation of these random risks is left to the reader. They might be:

- Daily (negative) returns on financial asset or portfolio – losses and profits
- Higher or lower frequency returns

- Operational losses
- Catastrophic insurance claims
- Credit losses

Moreover, they might represent risks which we can directly observe or they might also represent risks which we are forced to simulate in some Monte Carlo procedure, because of the impracticality of obtaining data. There are situations where, despite simulating from a known stochastic model, the complexity of the system is such that we do not know exactly what the loss distribution F is.

We avoid assuming independence, which for certain of the above interpretations (particularly market returns) is well-known to be unrealistic.

2.1 Measures of Extreme Risk

Mathematically we define our measures of extreme risk in terms of the loss distribution F. Let $1 > q \geq 0.95$ say. Value-at-Risk (VaR) is the qth quantile of the distribution F

$$\mathrm{VaR}_q = F^{-1}(q),$$

where F^{-1} is the inverse of F, and expected shortfall is the expected loss size, given that VaR is exceeded

$$\mathrm{ES}_q = E\left[X \mid X > \mathrm{VaR}_q\right].$$

Like F itself these are theoretical quantities which we will never know. Our goal in risk measurement is estimates $\widehat{\mathrm{VaR}}_q$ and $\widehat{\mathrm{ES}}_q$ of these measures and in this chapter we obtain two explicit formulae (6) and (10). The reader who wishes to avoid mathematical theory may skim through the remaining sections of this chapter, pausing only to observe that the formulae in question are simple.

2.2 Generalized Pareto Distribution

The GPD is a two parameter distribution with distribution function

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

where $\beta > 0$, and where $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

This distribution is generalized in the sense that it subsumes certain other distributions under a common parametric form. ξ is the important *shape* parameter of the distribution and β is an additional scaling parameter. If $\xi > 0$ then $G_{\xi,\beta}$ is a reparametrized version of the ordinary Pareto distribution, which has a long history in actuarial mathematics as a model for large losses; $\xi = 0$ corresponds to the exponential distribution and $\xi < 0$ is known as a Pareto type II distribution.

The first case is the most relevant for risk management purposes since the GPD is *heavy-tailed* when $\xi > 0$. Whereas the normal distribution has *moments* of all orders, a heavy-tailed distribution does not possess a complete set of moments. In the case of the GPD with $\xi > 0$ we find that $E[X^k]$ is infinite for $k \geq 1/\xi$. When $\xi = 1/2$, the GPD is an infinite variance (second moment) distribution; when $\xi = 1/4$, the GPD has an infinite fourth moment.

Certain types of large claims data in insurance typically suggest an infinite second moment; similarly econometricians might claim that certain market returns indicate a distribution with infinite fourth moment. The normal distribution cannot model these phenomena but the GPD is used to capture precisely this kind of behaviour, as we shall explain in the next sections. To make matters concrete we take a popular insurance example. Our data consist of 2156 large industrial fire insurance claims from Denmark covering the years 1980 to 1990. The reader may visualize these losses in any way he or she wishes.

2.3 Estimating Excess Distributions

The distribution of *excesses losses* over a high threshold u is defined to be

$$F_u(y) = P\left\{X - u \leq y \mid X > u\right\}, \tag{1}$$

for $0 \leq y < x_0 - u$ where $x_0 \leq \infty$ is the *right endpoint* of F, to be explained below. The excess distribution represents the probability that a loss exceeds the threshold u by at most an amount y, given the information that it exceeds the threshold. It is very useful to observe that it can be written in terms of the underlying F as

$$F_u(y) = \frac{F(y+u) - F(u)}{1 - F(u)}. \tag{2}$$

Mostly we would assume our underlying F is a distribution with an infinite right endpoint, i.e. it allows the possibility of arbitrarily large losses, even if it attributes negligible probability to unreasonably large outcomes, e.g. the normal or t– distributions. But it is also conceivable, in certain applications, that F could have a finite right endpoint. An example is the beta distribution on the interval $[0, 1]$ which attributes zero probability to outcomes larger than 1 and which might be used, for example, as the distribution of credit losses expressed as a proportion of exposure.

The following limit theorem is a key result in EVT and explains the importance of the GPD.

Theorem 1 *For a large class of underlying distributions we can find a function $\beta(u)$ such that*

$$\lim_{u \to x_0} \sup_{0 \leq y < x_0 - u} |F_u(y) - G_{\xi,\beta(u)}(y)| = 0.$$

That is, for a large class of underlying distributions F, as the threshold u is progressively raised, the excess distribution F_u converges to a generalized Pareto. The theorem is of course not mathematically complete, because we fail to say exactly what we mean by a large class of underlying distributions. For this paper it is sufficient to know that the class contains *all* the common continuous distributions of statistics and actuarial science (normal, lognormal, χ^2, t, F, gamma, exponential, uniform, beta, etc.).

In the sense of the above theorem, the GPD is the *natural model* for the unknown excess distribution above sufficiently high thresholds, and this fact is the essential insight on which our entire method is built. Our model for a risk X_i having distribution F assumes that, for a certain u, the excess distribution above this threshold may be taken to be exactly GPD for some ξ and β

$$F_u(y) = G_{\xi,\beta}(y). \tag{3}$$

Assuming we have realisations of $X_1, X_2, \ldots$ we use statistics to make the model more precise by choosing a sensible u and estimating ξ and β. Supposing that N_u out of a total of n data points exceed the threshold, the GPD is fitted to the N_u excesses by some statistical fitting method to obtain estimates $\hat{\xi}$ and $\hat{\beta}$. We favour maximum likelihood estimation (MLE) of these parameters, where the parameter values are chosen to maximize the joint probability density of the observations. This is the most general fitting method in statistics and it also allows us to give estimates of statistical error (standard errors) for the parameter estimates.

Choice of the threshold is basically a compromise between choosing a sufficiently high threshold so that the asymptotic theorem can be considered to be essentially exact and choosing a sufficiently low threshold so that we have sufficient material for estimation of the parameters. For further information on this data-analytic issue see McNeil (1997).

For our demonstration data, we take a threshold at 10 (million Krone). This reduces our n=2156 losses to N_u=109 threshold exceedances. On the basis of these data ξ and β are estimated to be 0.50 and 7.0; the value of ξ shows the heavy-tailedness of the data and suggests a good explanatory model may have an infinite variance. In Figure 1 the estimated GPD model for the excess distribution is shown as a smooth curve. The empirical distribution of the 109 extreme values is shown by points; it is evident the GPD model fits these excess losses well.

2.4 Estimating Tails of Distributions

By setting $x = u + y$ and combining expressions (2) and (3) we see that our model can also be written as

$$F(x) = (1 - F(u))G_{\xi,\beta}(x - u) + F(u), \tag{4}$$

for $x > u$. This formula shows that we may move easily to an interpretation of the model in terms of the tail of the underlying distribution $F(x)$ for $x > u$.

Our aim is to use (4) to construct a *tail estimator* and the only additional element we require to do this is an estimate of $F(u)$. For this purpose we take the obvious *empirical* estimator $(n - N_u)/n$. That is, we use the method of *historical simulation* (HS).

An immediate question is, why do we not use the HS method to estimate the whole tail of $F(x)$ (i.e. for all $x \geq u$)? This is because historical simulation is a poor method in the tail of the distribution where data become sparse. In setting a threshold at u we are judging that we have sufficient observations exceeding u to enable a reasonable HS estimate of $F(u)$, but for higher levels the historical method would be too unreliable.

Putting our HS estimate of $F(u)$ and our maximum likelihood estimates of the parameters of the GPD together we arrive at the tail estimator

$$\hat{F}(x) = 1 - \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x - u}{\hat{\beta}}\right)^{-1/\hat{\xi}}; \tag{5}$$

it is important to observe that this estimator is only valid for $x > u$. This estimate can be viewed as a kind of HS estimate augmented by EVT and it can be constructed whenever we believe data come from a common distribution, although its statistical properties are best understood in the situation when the data may also be assumed independent or only weakly dependent.

For our demonstration data the HS estimate of $F(u)$ is 0.95 $((2156-109)/2156)$ so that our threshold is positioned (approximately) at the 95th sample percentile. Combining this with our parametric model for the excess distribution we obtain the tail estimate shown in Figure 2. In this figure the y-axis actually indicates the tail probabilities $1-F(x)$. The top left corner of the graph shows that the threshold of 10 corresponds to a tail probability of 0.05 (as estimated by HS). The points again represent the 109 large losses and the solid curve shows how the tail estimation formula allows extrapolation into the area where the data become a sparse and unreliable guide to their unknown parent distribution.

2.5 Estimating VaR

For a given probability $q > F(u)$ the VaR estimate is calculated by inverting the tail estimation formula (5) to get

$$\widehat{\mathrm{VaR}}_q = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-q)\right)^{-\hat{\xi}} - 1\right). \tag{6}$$

In standard statistical language this is a quantile estimate, where the quantile is an unknown parameter of an unknown underlying distribution. It is possible to give a *confidence interval* for $\widehat{\mathrm{VaR}}_q$ using a method known as profile likelihood; this yields an asymptotic interval in which we have confidence that VaR lies. The asymmetric interval reflects a fundamental asymmetry in the problem of estimating a high quantile for heavy-tailed data: it is easier to bound the interval below than to bound it above.

In Figure 3 we estimate $\mathrm{VaR}_{0.99}$ to be 27.3. The vertical dotted line intersects with the tail estimate at the point $(27.3, 0.01)$ and allows the VaR estimate to be read off the x-axis. The dotted curve is a tool to enable the calculation of a confidence interval for the VaR. The second y-axis on the right of the graph is a *confidence* scale (*not* a quantile scale). The horizontal dotted line corresponds to 95% confidence; the x-coordinates of the two points where the dotted curve intersects the horizontal line are the boundaries of the 95% confidence interval $(23.3, 33.1)$. Two things should be observed: we obtain a wider 99% confidence interval by dropping the horizontal line down to the value 99 on the confidence axis; the interval is asymmetric as desired.

2.6 Estimating ES

Expected shortfall is related to VaR by

$$\mathrm{ES}_q = \mathrm{VaR}_q + E[X - \mathrm{VaR}_q \mid X > \mathrm{VaR}_q], \tag{7}$$

where the second term is simply the mean of the excess distribution $F_{\mathrm{VaR}_q}(y)$ over the threshold VaR_q. Our model for the excess distribution above the threshold u (3) has a nice stability property. If we take any higher threshold, such as VaR_q for $q > F(u)$, then the excess distribution above the higher threshold is also GPD with the same shape parameter, but a different scaling. It is easily shown that a consequence of the model (3) is that

$$F_{\mathrm{VaR}_q}(y) = G_{\xi,\beta+\xi(\mathrm{VaR}_q-u)}(y). \tag{8}$$

The beauty of (8) is that we have a simple explicit model for the excess losses above the VaR. With this model we can calculate many characteristics of the losses *beyond* VaR. By noting that (provided $\xi < 1$) the mean of the distribution in (8) is $(\beta+\xi(\mathrm{VaR}_q-u))/(1-\xi)$, we can calculate the expected shortfall. We find that

$$\frac{\mathrm{ES}_q}{\mathrm{VaR}_q} = \frac{1}{1-\xi} + \frac{\beta - \xi u}{(1-\xi)\mathrm{VaR}_q}. \tag{9}$$

It is worth examining this ratio a little more closely in the case where the underlying distribution has an infinite right endpoint. In this case the ratio is largely determined by the factor $1/(1-\xi)$. The second term on the right hand side of (9) becomes negligibly small as the probability q gets nearer and nearer to 1. This asymptotic observation underlines the importance of the shape parameter ξ in tail estimation. It determines how our two risk measures differ in the extreme regions of the loss distribution.

Expected shortfall is estimated by substituting data-based estimates for everything which is unknown in (9) to obtain

$$\widehat{\mathrm{ES}}_q = \frac{\widehat{\mathrm{VaR}}_q}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}}. \tag{10}$$

For our demonstration data $1/(1-\hat{\xi}) \approx 2.0$ and $(\hat{\beta} - \hat{\xi}u)/(1-\hat{\xi}) \approx 4.0$. Essentially $\widehat{\mathrm{ES}}_q$ is obtained from $\widehat{\mathrm{VaR}}_q$ by doubling it. Our estimate is 58.2 and we have marked this with a second vertical line in Figure 4. Again using the profile likelihood method we show how an estimate of the 95% confidence interval for $\widehat{\mathrm{ES}}_q$ can be added (41.6,154). Clearly the uncertainty about the value of our coherent risk measure is large, but this is to be expected with such heavy-tailed data. The prudent risk manager should be aware of the magnitude of his uncertainty about extreme phenomena.

3 Extreme Market Risk

In the market risk interpretation of our random variables

$$X_t = -(\log S_t - \log S_{t-1}) \approx (S_{t-1} - S_t)/S_{t-1}, \tag{11}$$

represents the loss on a portfolio of traded assets on day t, where S_t is the closing value of the portfolio on that day. We change to subscript t to emphasize the temporal indexing of our risks. As shown above, the loss may be defined as a relative or logarithmic difference, both definitions giving very similar values.

In calculating daily VaR estimates for such risks, there is now a general recognition that the calculation should take into account *volatility* of market instruments. An extreme value in a period of high volatility appears less extreme than the same value in a period of low volatility. Various authors have acknowledged the need to *scale* VaR estimates by current volatility in some way (see, for example, Hull & White (1998)). Any approach which achieves this we will call a *dynamic* risk measurement procedure. In this chapter we focus on how the dynamic measurement of market risks can be further enhanced with EVT to take into account the extreme risk over and above the volatility risk.

Most market return series show a great deal of common structure. This suggests that more sophisticated modelling is both possible and necessary; it is not sufficient to assume they are independent and identically distributed. Various stylized facts of empirical finance argue against this. While the correlation of market returns is low, the serial correlation of absolute or squared returns is high; returns show *volatility clustering* – the tendency of large values to be followed by other large values, although not necessarily of the same sign.

3.1 Stochastic Volatility Models

The most popular models for this phenomenon are the stochastic volatility (SV) models, which take the form

$$X_t = \mu_t + \sigma_t Z_t, \tag{12}$$

where σ_t is the volatility of the return on day t and μ_t is the expected return. These values are considered to depend in a deterministic way on the past history of returns. The randomness in the model comes through the random variables Z_t, which are the *noise* variables or the *innovations* of the process.

We assume that the noise variables Z_t are *independent* with an identical unknown distribution $F_Z(z)$. (By convention we assume this distribution has mean zero and variance 1, so that σ_t is directly interpretable as the volatility of X_t.) Although the structure of the model causes the X_t to be *dependent* we assume that the model is such that the X_t are identically distributed with unknown distribution function $F_X(x)$. In the language of time series we assume that X_t is a *stationary* process.

Models which fit into this framework include the ARCH/GARCH family. A simple example is

$$\begin{aligned} \mu_t &= \lambda X_{t-1}, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 (X_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{13}$$

with α_0, α_1, $\beta > 0$, $\beta + \alpha_1 < 1$ and $|\lambda| < 1$. This is an autoregressive process with GARCH(1,1) errors and with a suitably chosen noise distribution this is a model which mimics many features of real financial return series.

3.2 Dynamic Risk Management

Suppose we have followed daily market movements over a period of time and we find ourselves at the close of day t. In dynamic risk management we are interested in the *conditional* return distribution

$$F_{X_{t+1}+\ldots+X_{t+k}|\mathcal{F}_t}(x), \tag{14}$$

where the symbol $\mathcal{F}_t$ represents the *history* of the process X_t up to and including day t. In looking at this distribution we ask, what is the distribution of returns over the next $k \geq 1$ days, given the present market background? This is the issue in daily VaR (or ES) calculation.

This view can be contrasted with static risk management where we are interested in the *unconditional* or stationary distribution $F_X(x)$ (or $F_{X_1+\ldots+X_k}(x)$ for a k-day return). Here we take a complementary view and ask questions like, how large is a 100 day loss in general? What is the magnitude of a 5-year loss?

We redefine our risk measures slightly to be quantiles and expected shortfalls for the distribution (14) and we introduce the notation $\mathrm{VaR}_q^t(k)$ and $\mathrm{ES}_q^t(k)$. The subscript t shows that these are dynamic measures designed for calculation at the close of day t; k denotes the time horizon. If we drop k we consider a one day horizon.

3.3 One day horizons

The structure of the model (12) means that the dynamic measures take simple forms for a 1 day horizon.

$$\begin{aligned} \mathrm{VaR}_q^t &= \mu_{t+1} + \sigma_{t+1}\mathrm{VaR(Z)}_q \\ \mathrm{ES}_q^t &= \mu_{t+1} + \sigma_{t+1}\mathrm{ES(Z)}_q, \end{aligned} \tag{15}$$

where $\mathrm{VaR(Z)}_q$ denotes the qth quantile of a noise variable Z_i and $\mathrm{ES(Z)}_q$ is the corresponding expected shortfall.

The simplest approaches to estimating a dynamic VaR make the assumption that $F_Z(z)$ is a known standard distribution, typically the normal distribution. In this case $\mathrm{VaR(Z)}_q$ is easily calculated. To estimate the dynamic measure a procedure is required to estimate tomorrow's expected return μ_{t+1} and tomorrow's volatility σ_{t+1}. Several approaches are available for forecasting the mean and volatility of SV models. Two possibilities are the exponentially weighted moving average model (EWMA) as used in Riskmetrics or GARCH modelling. The problem with the assumption of conditional normality is that this tends to lead to an underestimation of the dynamic measures. Empirical analyses suggest the conditional distribution of appropriate SV models for real data is often heavier-tailed than the normal distribution.

The trick as far as augmenting the dynamic procedure with EVT is concerned, is to apply it to the random variables Z_t rather than X_t. In the EVT approach (McNeil & Frey 1998) we avoid assuming any particular form for $F_Z(z)$; instead we apply the GPD tail estimation procedure to this distribution. We assume that above some high threshold u the excess distribution is exactly GPD. The problem with the statistical estimation of this model is that the Z_t variables cannot be directly observed, but this is solved by the following two stage approach.

Suppose at the close of day t we consider a time window containing the last n returns $X_{t-n+1}, \ldots, X_t$.

1. A GARCH-type stochastic volatility model, typically an AR model with GARCH errors, is fitted to the historical data by *pseudo* maximum likelihood (PML). From this model the so-called *residuals* are extracted. If the model is tenable these can be regarded as realisations of the unobserved, independent noise variables $Z_{t-n+1}, \ldots, Z_t$. The GARCH-type model is used to calculate 1-step predictions of μ_{t+1} and σ_{t+1}.

2. EVT is applied to the residuals. For some choice of threshold the GPD method is used to estimate $\mathrm{VaR(Z)}_q$ and $\mathrm{ES(Z)}_q$ as outlined in the previous chapter. The risk measures are calculated using equations (15).

3.4 Backtesting

The procedure above, which we term dynamic or conditional EVT, is a successful way of adapting EVT to the special task of daily market risk measurement. This can be verified by backtesting the method on historical return series.

Figure 5 shows a dynamic VaR estimate using the conditional EVT method for daily losses on the DAX index. At the close of every day the method is applied to the last 1000 data points using a threshold u set at the 90th sample percentile of the residuals. Volatility and expected return forecasts are based on an AR(1) model with GARCH(1,1) errors as in (13). The dashed line shows how the dynamic VaR estimate reacts rapidly to volatility changes. Superimposed on the graph is a static VaR estimate calculated with static EVT as in Chapter 2. This changes only gradually (as extreme observations drop occasionally from the back of the moving data window).

A VaR estimation method is backtested by comparing the estimates with the actual losses observed on the next day. A VaR violation occurs when the actual loss exceeds the estimate. Various dynamic (and static) methods of VaR estimation can be compared by counting violations; tests of the violation counts based on the binomial distribution can show when a systematic underestimation or overestimation of VaR seems to be taking place. It is also possible to devise backtests which compare dynamic ES estimates with actual incurred losses exceeding the VaR on days when VaR violation takes place (see McNeil & Frey (1998) for details).

	S&P	DAX
Length of Test	7414	5146
$\mathrm{VaR}_{0.95}$		
Expected violations	371	257
Dynamic EVT violations	366 (0.41)	258 (0.49)
Dynamic normal violations	384 (0.25)	238 (0.11)
Static EVT violations	402 (0.05)	266 (0.30)
$\mathrm{VaR}_{0.99}$		
Expected violations	74	51
Dynamic EVT violations	73 (0.48)	55 (0.33)
Dynamic normal violations	104 (0.00)	74 (0.00)
Static EVT violations	86 (0.10)	59 (0.16)
$\mathrm{VaR}_{0.995}$		
Expected violations	37	26
Dynamic EVT violations	43 (0.18)	24 (0.42)
Dynamic normal violations	63 (0.00)	44 (0.00)
Static EVT violations	50 (0.02)	36 (0.03)

Table 1: Some VaR backtesting results for two major indices. Values in brackets are p-values for a statistical test of the success of the method; values smaller than 0.05 indicate failure.

McNeil and Frey compare conditional EVT with other dynamic approaches which do not explicitly model the tail risk associated with the innovation distribution. In particular, they compare the approach with methods which assume normally distributed or t–distributed innovations (which they label the dynamic normal and dynamic t methods). They also compare dynamic with static EVT.

The VaR violations relating to the DAX data in Figure 5 are shown in Figure 6 for the dynamic EVT, dynamic normal and static EVT methods, these being denoted respectively by the circular, triangular and square plotting symbols. It is apparent that although the dynamic normal estimate reacts to volatility, it is violated more often than the dynamic EVT estimate; it is also clear that the static EVT estimate tends to be violated several times in a row in periods of high volatility because it is unable to react swiftly enough to the changing volatility. These observations are borne out by the results in Table 1, which is a sample of the backtesting results in McNeil & Frey (1998).

The main results of their paper are

- Dynamic EVT is in general the best method for estimating VaR_q^t for $q \geq 0.95$. (Dynamic t is an effective simple alternative, if returns are not too asymmetric.) For $q \geq 0.99$, the dynamic normal method is not good enough.

- The dynamic normal method is useless for estimating ES_q^t, even when $q = 0.95$. To estimate expected shortfall a dynamic procedure has to be enhanced with EVT.

It is worth understanding in more detail why EVT is particularly necessary for calculating expected shortfall estimates. In a stochastic volatility model the ratio $\mathrm{ES}_q^t/\mathrm{VaR}_q^t$ is essentially given by $\mathrm{ES(Z)}_q/\mathrm{VaR(Z)}_q$, the equivalent ratio for the noise distribution. We have already observed in (9) that this ratio is largely determined by the weight of the tail of the distribution $F_Z(z)$ as summarized by the ξ parameter of a suitable GPD approximation.

We have tabulated some values for this ratio in Table 2 in the case when $F_Z(z)$ admits a GPD tail approximation with $\xi = 0.22$ (the threshold being set at $u = 1.2$ with $\beta = 0.57$). The ratio is compared with the equivalent ratio for a normal innovation distribution.

q	0.95	0.99	0.995	$q \to 1$
GPD tail	1.52	1.42	1.39	1.29
Normal	1.25	1.15	1.12	1.00

Table 2: ES to VaR ratios under two models for the noise distribution.

Clearly the ratios are smaller for the normal distribution. If we erroneously assume conditional normality in our models, not only do we tend to underestimate VaR, but we also underestimate the ES/VaR ratio. Our error for ES is magnified due to this double underestimation.

3.5 Multiple day horizons

For multiple day horizons ($k > 1$) we do not have the simple expressions for dynamic risk measures which we had in (15). Explicit estimation of the risk measures is difficult and it is attractive to want to use a simple scaling rule, like the famous square root of time rule, to turn one day VaR into k-day VaR. Unfortunately, square root of time is designed for the case when returns are normally distributed and is not appropriate for the kind of SV model driven by heavy-tailed noise that we consider realistic. Nor is it necessarily appropriate for scaling dynamic risk measures, where one might imagine current volatility should be taken into account.

It is possible to adopt a Monte Carlo approach to estimating dynamic risk measures for longer time horizons. Possible future paths for the SV model of the returns may be simulated and possible k-day losses calculated.

To calculate a single future path on day t we could proceed as follows. The noise distribution is modelled with a composite model consisting of GPD tail estimates for *both* tails and a simple empirical (i.e. historical simulation) estimate based on the model residuals in the centre. k independent values $Z_{t+1}, \ldots, Z_{t+k}$ are simulated from this model (for details of the necessary random number generator see McNeil & Frey (1998)). Using the noise values and the current estimated volatility from the fitted GARCH-type model, future values of the return process $X_{t+1}, \ldots, X_{t+k}$ are recursively calculated and summed to obtain the k-day loss. This loss is taken as a realisation from the conditional distribution of the k-day loss (14).

By repeating this process many times to obtain a sample of values (perhaps 1000) from the target distribution and then applying the GPD tail estimation procedure to these simulated data, reasonable estimates of the risk measures may be obtained.

Such simulation results can then be used to examine the nature of the implied scaling law. McNeil and Frey conduct such an experiment and suggest that for horizons up to 50 days $\mathrm{VaR}_q^t(k)$ typically obeys a power scaling law of the form

$$\mathrm{VaR}_q^t(k)/\mathrm{VaR}_q^t \approx k^{\lambda_t},$$

where λ_t depends on the current volatility. Their results are summarized in Table 3. In their experiment square root of time scaling ($\lambda_t = 0.5$) is appropriate on days when estimated volatility is high; otherwise a a larger scaling exponent is suggested.

q	0.95	0.99
low volatility	0.65	0.65
average volatility	0.60	0.59
high volatility	0.48	0.47

Table 3: Typical scaling exponents for multiple day horizons. Low, average and high volatilities are taken to be the 5th, 50th and 95th percentiles of estimated historical volatilities respectively.

4 Other Issues

In this section we provide briefer notes on some other relevant topics in EVT.

4.1 Block Maxima Models for Stress Losses

For a more complete understanding of EVT we should be aware of the block maxima models. Although less useful than the threshold models, these models are not without practical relevance and could be used to provide estimates of stress losses.

Theorem 1 is not really a mathematical result as it presently stands. We could make it mathematically complete by saying that distributions which admit the asymptotic GPD model for their excess distribution are precisely those distributions in the *maximum domain of attraction* of an *extreme value distribution*. To understand this statement we must first define the generalized extreme value distribution (GEV).

The distribution function of the GEV is given by

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

where $1 + \xi x > 0$ and ξ is the shape parameter. As in the case of the GPD, this parametric form subsumes distributions which are known by other names. When $\xi > 0$ the distribution is known as Fréchet; when $\xi = 0$ it is a Gumbel distribution; when $\xi < 0$ it is a Weibull distribution.

The GEV is the *natural limit* distribution for *normalized maxima*. Suppose that X_1, X_2, ... are independent identically distributed losses with distribution function F as earlier and define the maximum of a block of n observations to be $M_n = \max(X_1, \ldots, X_n)$. Suppose it is possible to find sequences of numbers $a_n > 0$ and b_n such that the distribution of $(M_n - b_n)/a_n$, converges to some limiting distribution H as the block size increases. If this occurs F is said to be in the maximum domain of attraction of H. To be absolutely technically correct we should assume this limit is a non-degenerate (reasonably behaved) distribution. We also note that the assumption of independent losses is by no means important for the result that now follows and can be dropped if some additional minor technical conditions are fulfilled.

Theorem 2 *If F is in the maximum domain of attraction of a non-degenerate H then this limit must be an extreme value distribution of the form $H(x) = H_\xi((x-\mu)/\sigma)$, for some ξ, μ and $\sigma > 0$.*

This result is known as the Fisher-Tippett Theorem and occupies an analogous position with respect to the study of maxima as the famous central limit theorem holds for the study of sums or averages. Fisher-Tippett essentially says that the GEV is the only possible limiting distribution for (normalized) block maxima.

If the ξ of the limiting GEV is strictly positive, F is said to be in the maximum domain of attraction of the Fréchet. Distributions in this class include the Pareto, t, Burr, loggamma and Cauchy distributions. If $\xi = 0$ then F is in the maximum domain of attraction of the Gumbel; examples are the normal, lognormal and gamma distributions. If $\xi < 0$ then F is in the maximum domain of attraction of the Weibull; examples are the uniform and beta distributions. Distributions in these three classes are precisely the distributions for which excess distributions converge to a GPD limit.

To implement an analysis of stress losses based on this limiting model for block maxima we require a lot of data, since we must define blocks and reduce these data to block maxima only. Suppose, for the sake of illustration, that we have daily (negative) return data which we divide into k large blocks of essentially equal size; for example, we might take yearly or semesterly blocks. Let $M_n^{(j)} = \max(X_1^{(j)}, X_2^{(j)}, \ldots, X_n^{(j)})$ be the maximum of the n observations in block j. Using the method of maximum likelihood we fit the GEV to the block maxima data $M_n^{(1)}, \ldots, M_n^{(k)}$. That is we assume that our block size is sufficiently large so that the limiting result of Theorem 2 may be taken as approximately exact.

Suppose that we fit a GEV model $H_{\hat{\xi},\hat{\mu},\hat{\sigma}}$ to semesterly maxima of daily negative returns. Then a quantile of this distribution is a stress loss. $H^{-1}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(0.95)$ gives the magnitude of daily loss level we might expect to reach every 20 semesters or 10 years. This stress loss is known as the 20 semester *return level* and can be considered as a kind of *unconditional* quantile estimate for the unknown underlying distribution F. In Figure 7 we show the 20 semester return level for daily negative returns on the DAX index; the return level itself is marked by a solid line and an asymmetric 95% confidence interval is marked by dotted lines. In 23 years of data 4 observations exceed the point estimate; these 4 observations occur in 3 different semesters. In a full analysis we would of course try a series of different block sizes and compare results. See Embrechts et al. (1997) for a more detailed description of both the theory and practice of block maxima modelling using the Fisher-Tippett Theorem.

4.2 Multivariate Extremes

So far we have been concerned with *univariate* EVT. We have modelled the tails of univariate distributions and estimated associated risk measures. In fact, there is also a multivariate extreme value theory (MEVT) and this can be used to model the tails of multivariate distributions in a theoretically supported way. In a sense MEVT is about studying the *dependence structure* of extreme events, as we shall now explain.

Consider the *random vector* $\mathbf{X} = (X_1, \ldots, X_d)'$ which represents losses of d different kinds measured at the same point in time. We assume these losses have *joint* distribution $F(x_1, \ldots, x_d) = P\{X_1 \leq x_1, \ldots, X_d \leq x_d\}$ and that individual losses have continuous *marginal* distributions $F_i(x) = P\{X_i \leq x\}$. It has been shown by Sklar (see Nelsen (1999)) that every joint distribution can be written as

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)),$$

for a unique function C that is known as the *copula* of F. A copula may be thought of in two equivalent ways: as a function (with some technical restrictions) that maps values in the unit hypercube to values in the unit interval; as a multivariate distribution function with standard uniform marginal distributions. The copula C does not change under (strictly) increasing transformations of the losses $X_1, \ldots, X_d$ and it makes sense to interpret C as the dependence structure of $\mathbf{X}$ or F, as the following simple illustration in $d = 2$ dimensions shows.

We take the marginal distributions to be standard univariate normal distributions $F_1 = F_2 = \Phi$. We can then choose any copula C (i.e. any bivariate distribution with uniform marginals) and apply it to these marginals to obtain bivariate distributions with normal marginals. For one particular choice of C, which we call the Gaussian copula and denote C_ρ^{Ga}, we obtain the standard bivariate normal distribution with correlation ρ. The Gaussian copula does not have a simple closed form and must be written as a double integral - consult Embrechts, McNeil & Straumann (1999) for more details. Another interesting copula is the Gumbel copula which does have a simple closed form,

$$C_\beta^{\mathrm{Gu}}(v_1, v_2) = \exp\left[-\left\{(-\log v_1)^{1/\beta} + (-\log v_2)^{1/\beta}\right\}^\beta\right],\ 0 < \beta \leq 1. \qquad (16)$$

Figure 8 shows the bivariate distributions which arise when we apply the two copulas $C_{0.7}^{\mathrm{Ga}}$ and $C_{0.5}^{\mathrm{Gu}}$ to standard normal marginals. The left-hand picture is the standard bivariate normal with correlation 70%; the right-hand picture is a bivariate distribution with approximately equal correlation but the tendency to generate extreme values of X_1 and X_2 simultaneously. It is, in this sense, a more dangerous distribution for risk managers. On the basis of correlation, these distributions cannot be differentiated but they obviously have entirely different dependence structures. The bivariate normal has rather weak *tail dependence*; the normal-Gumbel distribution has pronounced tail dependence. For more examples of parametric copulas consult Nelsen (1999) or Joe (1997).

One way of understanding MEVT is as the study of copulas which arise in the limiting multivariate distribution of componentwise block maxima. What do we mean by this? Suppose we have a family of random vectors $\mathbf{X_1}, \mathbf{X_2}, \ldots$ representing d-dimensional losses at *different* points in time, where $\mathbf{X_i} = (X_{i1}, \ldots, X_{id})'$. A simple interpretation might be that they represent daily (negative) returns for d instruments. As for the univariate discussion of block maxima, we assume that losses at different points in time are independent. This assumption simplifies the statement of the result, but can again be relaxed to allow serial dependence of losses at the cost of some additional technical conditions.

We define the vector of componentwise block maxima to be $= (M_{1n}, \ldots, M_{dn})'$ where $M_{jn} = \max(X_{1j}, \ldots, X_{nj})$ is the block maximum of the jth component for a block of size n observations. Now consider the vector of normalized block maxima given by $((M_{1n} - b_{1n})/a_{1n}, \ldots, (M_{dn} - b_{dn})/a_{dn})'$, where $a_{jn} > 0$ and b_{jn} are normalizing sequences as in Section 4.1. If this vector converges in distribution to a non-degenerate limiting distribution then this limit must have the form

$$C\left(H_{\xi_1}\left(\frac{x_1 - \mu_1}{\sigma_1}\right), \ldots, H_{\xi_d}\left(\frac{x_d - \mu_d}{\sigma_d}\right)\right),$$

for some values of the parameters ξ_j, μ_j and σ_j and some copula C. It must have this form because of univariate EVT. Each marginal distribution of the limiting multivariate distribution must be a GEV, as we learned in Theorem 2.

MEVT characterizes the copulas C which may arise in this limit - the so-called MEV copulas. It turns out that the limiting copulas must satisfy $C(u_1^t, \ldots, u_d^t) = C^t(u_1, \ldots, u_d)$ for $t > 0$. There is no single parametric family which contains all the MEV copulas, but certain parametric copulas are consistent with the above condition and might therefore be regarded as *natural* models for the dependence structure of extreme observations.

In two dimensions the Gumbel copula (16) is an example of an MEV copula; it is moreover a versatile copula. If the parameter β is 1 then $C_1^{\mathrm{Gu}}(v_1, v_2) = v_1 v_2$ and this copula models *independence* of the components of a random vector $(X_1, X_2)'$. If $\beta \in (0, 1)$ then the Gumbel copula models dependence between X_1 and X_2. As β decreases the dependence becomes stronger until a value $\beta = 0$ corresponds to *perfect dependence* of X_1 and X_2; this means $X_2 = T(X_1)$ for some strictly increasing function T. For $\beta < 1$ the Gumbel copula shows tail dependence - the tendency of extreme values to occur together as observed in Figure 8. For more details see Embrechts et al. (1999).

The Gumbel copula can be used to build tail models in two dimensions as follows. Suppose two risk factors $(X_1, X_2)'$ have an unknown joint distribution F and marginals F_1 and F_2 so that, for some copula C, $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$. Assume that we have n pairs of data points from this distribution. Using the univariate POT method we model the tails of the two marginal distributions by picking high thresholds u_1 and u_2 and using tail estimators of the form (5) to obtain

$$\hat{F}_i(x) = 1 - \frac{N_{u_i}}{n}\left(1 + \hat{\xi}_i \frac{x - u_i}{\hat{\beta}_i}\right)^{-1/\hat{\xi}_i}, x > u_i, \ i = 1, 2.$$

We model the dependence structure of observations exceeding these thresholds using the Gumbel copula $C_{\hat{\beta}}^{\mathrm{Gu}}$ for some estimated value $\hat{\beta}$ of the dependence parameter β. We put tail models and dependence structure together to obtain a model for the joint tail of F.

$$\hat{F}(x_1, x_2) = C_{\hat{\beta}}^{\mathrm{Gu}}\left(\hat{F}_1(x_1), \hat{F}_2(x_2)\right), x_1 > u_1, \ x_2 > u_2.$$

The estimate of the dependence parameter β can be determined by maximum likelihood, either in a second stage after the parameters of the tail estimators have been estimated or in a single stage estimation procedure where all parameters are estimated together. For further details of these statistical matters see Smith (1994). For further details of the theory consult Joe (1997).

This is perhaps the simplest bivariate POT model one can devise and it could be extended to higher dimensions by choosing extensions of the Gumbel copula to higher dimensions. Realistically, however, parametric models of this kind are only viable in a small number of dimensions. If we are interested in only a few risk factors and are particularly concerned that joint extreme values may occur, we can use such models to get useful descriptions of the joint tail. In very high dimensions there are simply too many parameters to estimate and too many different tails of the multivariate distribution to worry about - the so-called *curse of dimensionality.* In such situations collapsing the problem to a univariate problem by considering a whole portfolio of assets as a single risk and collecting data on a portfolio level seems more realistic.

4.3 Software for EVT

We are aware of two software systems for EVT. EVIS (Extreme Values In S-Plus) is a suite of free S-Plus functions for EVT developed at ETH Zurich. To use these functions it is necessary to have S-Plus, either for UNIX or Windows.

The functions provide assistance with four activities: exploring data to get a feel for the heaviness of tails; implementing the POT method as described in Section 2; implementing analyses of block maxima as described in Section 4.1; implementing a more advanced form of the POT method known as the point process approach. The EVIS functions provide simple templates which an S-Plus user could develop and incorporate into a customized risk management system. In particular EVIS combines easily with the extensive S-Plus time series functions or with the S+GARCH module. This permits dynamic risk measurement as described in Section 3.

XTREMES is commercial software developed by Rolf Reiss and Michael Thomas at the University of Siegen in Germany. It is designed to run as a self-contained program under Windows (NT, 95, 3.1). For didactic purposes this program is very successful; it is particularly helpful for understanding the different sorts of extreme value modelling that are possible and seeing how the models relate to each other. However, a user wanting to adapt XTREMES for risk management purposes will need to learn and use the Pascal-like integrated programming language XPL that comes with XTREMES. The stand-alone nature of XTREMES means that the user does not have access to the extensive libraries of pre-programmed functions that packages like S-Plus offer.

EVIS may be downloaded over the internet at `http://www.math.ethz.ch/~mcneil`. Information on XTREMES can be found at `http://www.xtremes.math.uni-siegen.de`.

5 Conclusion

EVT is here to stay as a technique in the risk manager's toolkit. We have argued in this paper that whenever tails of probability distributions are of interest, it is *natural* to consider applying the theoretically supported methods of EVT. Methods based around assumptions of normal distributions are likely to underestimate tail risk. Methods based on historical simulation can only provide very imprecise estimates of tail risk. EVT is the most scientific approach to an inherently difficult problem - predicting the size of a rare event.

We have given one very general and easily implementable method, the parametric POT method of Section 2, and indicated how this method may be adapted to more specialised risk management problems such as the management of market risks. The reader who wishes to learn more is encouraged to turn to textbooks like Embrechts et al. (1997) or Beirlant et al. (1996). The reader who has mistakenly gained the impression that

EVT alone will solve all risk management problems should read Diebold, Schuermann & Stroughair (1999) for some commonsense caveats to uncritical use of EVT. We hope, however, that all risk managers are persuaded that EVT will have an important role to play in the development of sound risk management systems for the future.

References

Artzner, P., Delbaen, F., Eber, J. & Heath, D. (1997), 'Thinking coherently', *RISK* **10**(11), 68–71.

Beirlant, J., Teugels, J. & Vynckier, P. (1996), *Practical analysis of extreme values*, Leuven University Press, Leuven.

Danielsson, J. & de Vries, C. (1997), 'Tail index and quantile estimation with very high frequency data', *Journal of Empirical Finance* **4**, 241–257.

Danielsson, J., Hartmann, P. & de Vries, C. (1998), 'The cost of conservatism', *RISK* **11**(1), 101–103.

Diebold, F., Schuermann, T. & Stroughair, J. (1999), Pitfalls and opportunities in the use of extreme value theory in risk management, *in* 'Advances in Computational Finance', Kluwer Academic Publishers, Amsterdam. To appear.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling extremal events for insurance and finance*, Springer, Berlin.

Embrechts, P., McNeil, A. & Straumann, D. (1999), 'Correlation and dependency in risk management: properties and pitfalls', preprint, ETH Zürich.

Embrechts, P., Resnick, S. & Samorodnitsky, G. (1998), 'Living on the edge', *RISK Magazine* **11**(1), 96–100.

Hull, J. & White, A. (1998), 'Incorporating volatility updating into the historical simulation method for value at risk', *Journal of Risk* **1**(1).

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

McNeil, A. (1997), 'Estimating the tails of loss severity distributions using extreme value theory', *ASTIN Bulletin* **27**, 117–137.

McNeil, A. (1998), 'History repeating', *Risk* **11**(1), 99.

McNeil, A. & Frey, R. (1998), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', preprint, ETH Zürich.

Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer, New York.

Reiss, R. & Thomas, M. (1997), *Statistical Analysis of Extreme Values*, Birkhäuser, Basel.

Smith, R. (1994), Multivariate threshold methods, *in* J. Galambos, ed., 'Extreme Value Theory and Applications', Kluwer Academic Publishers, pp. 225–248.

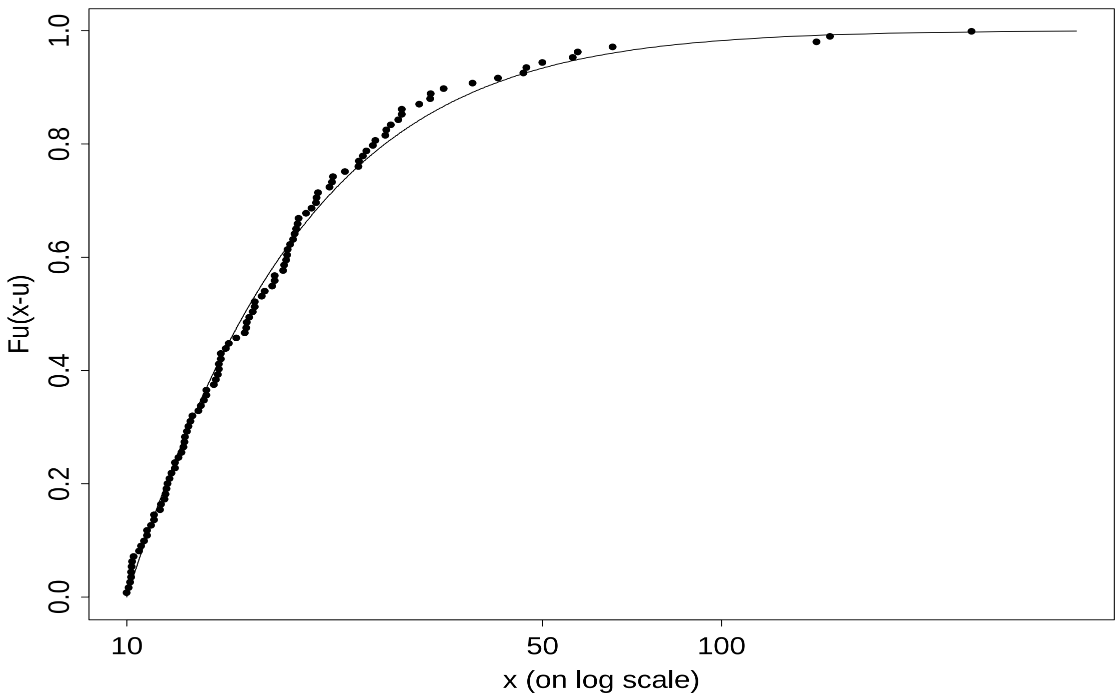

Figure 1: Modelling the excess distribution.

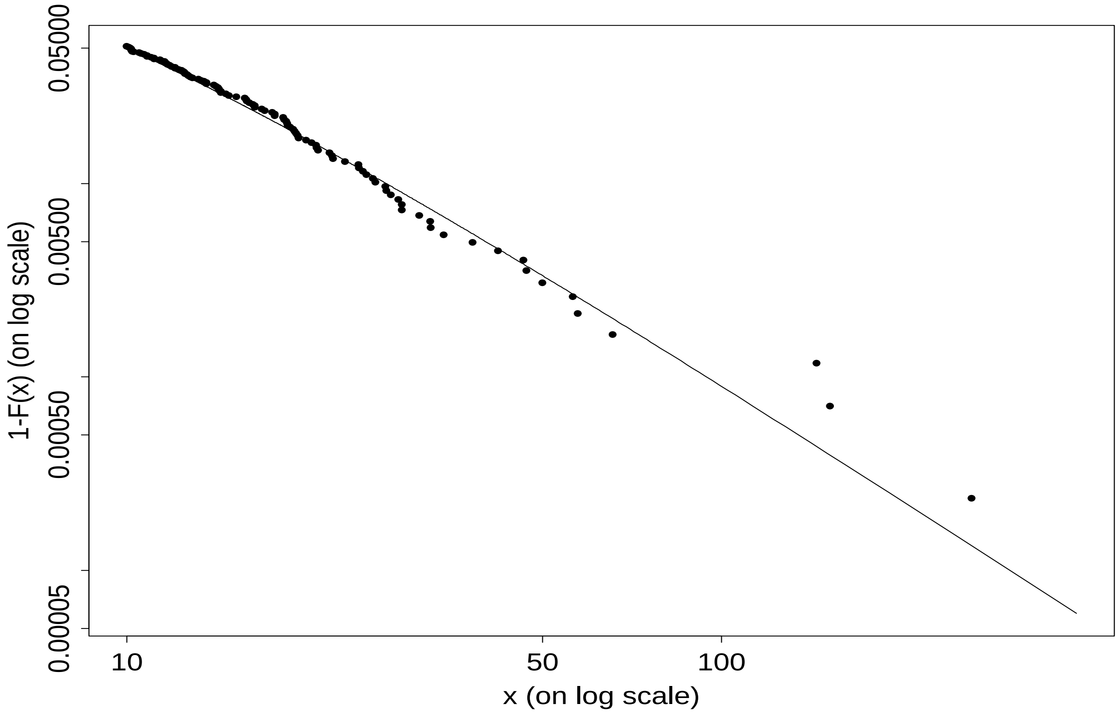

Figure 2: Modelling the tail of a distribution.

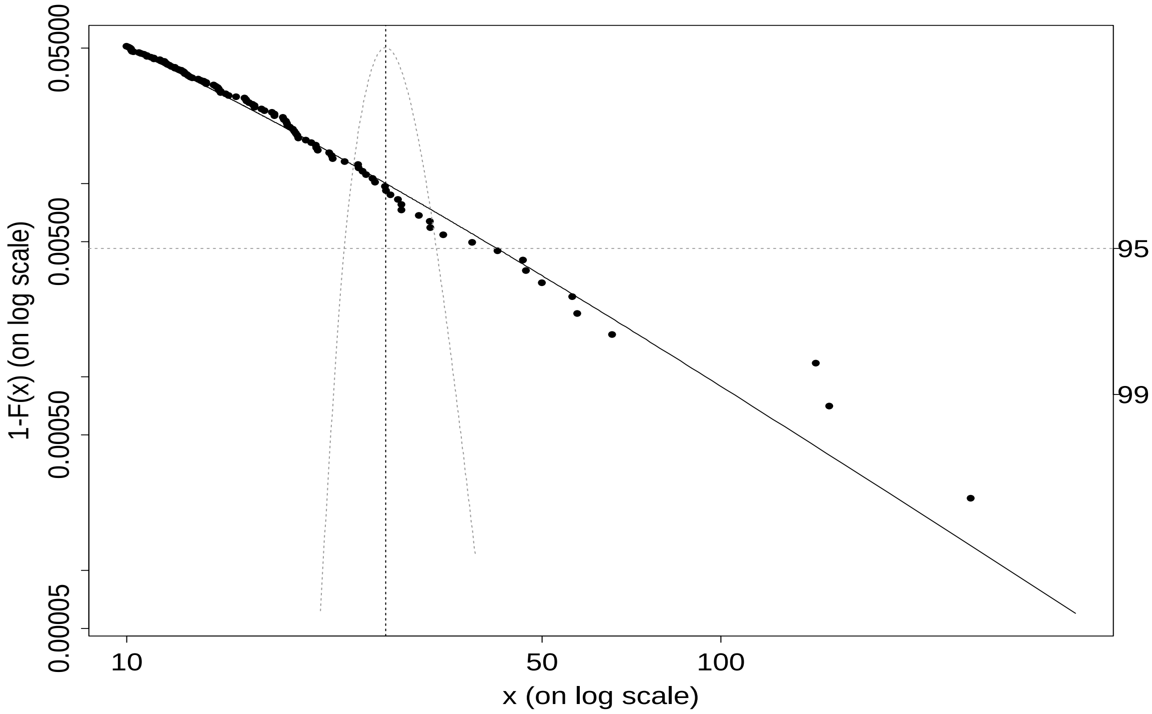

Figure 3: Estimating VaR.

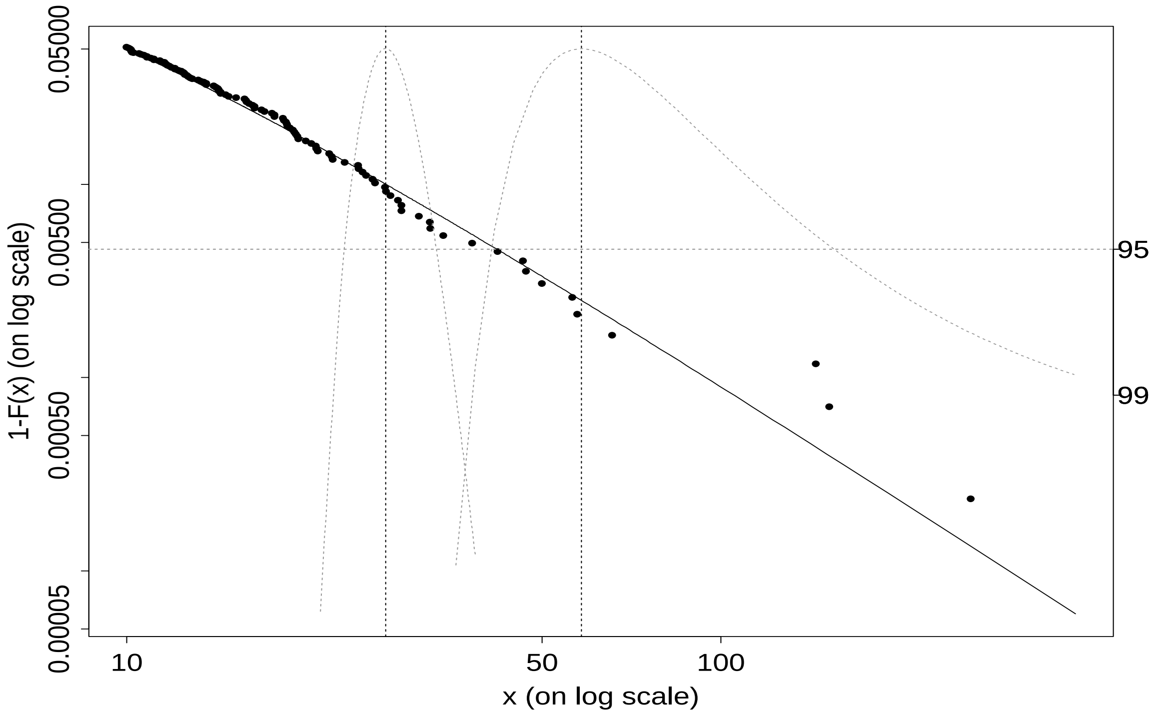

Figure 4: Estimating Expected Shortfall.

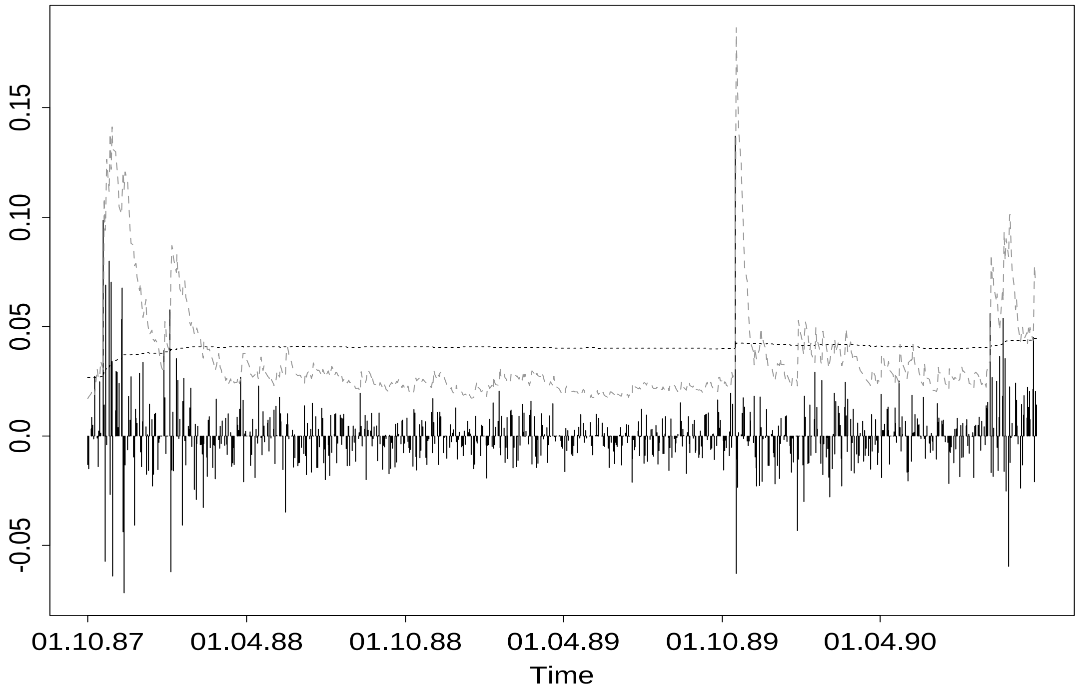

Figure 5: Dynamic VaR using EVT for the DAX index.

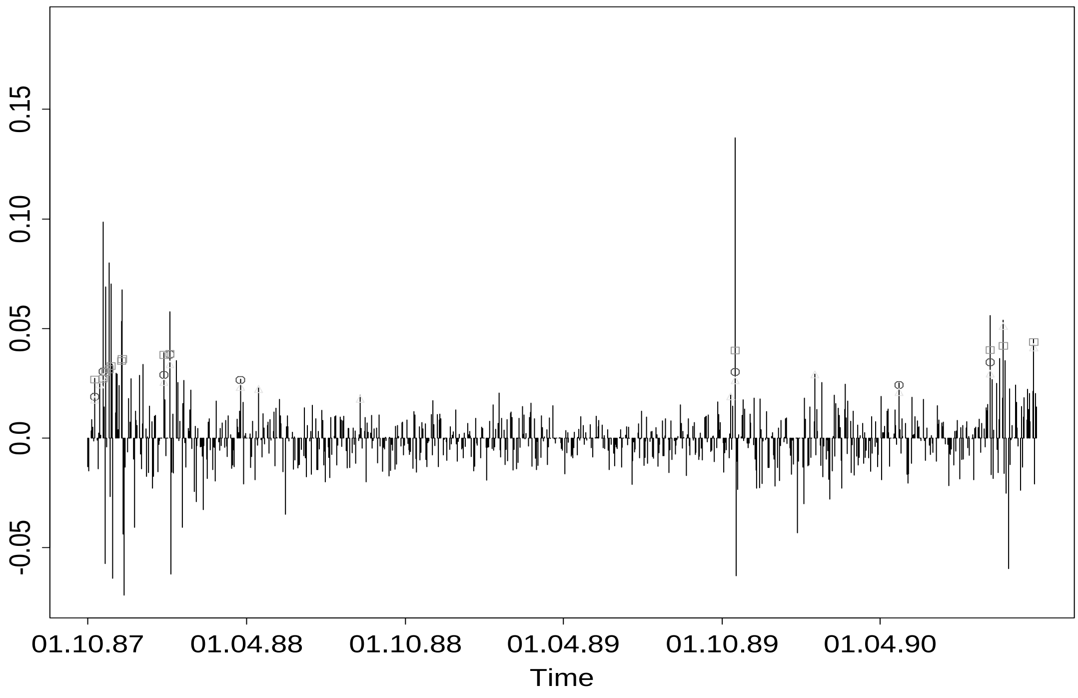

Figure 6: VaR violations in DAX backtest.

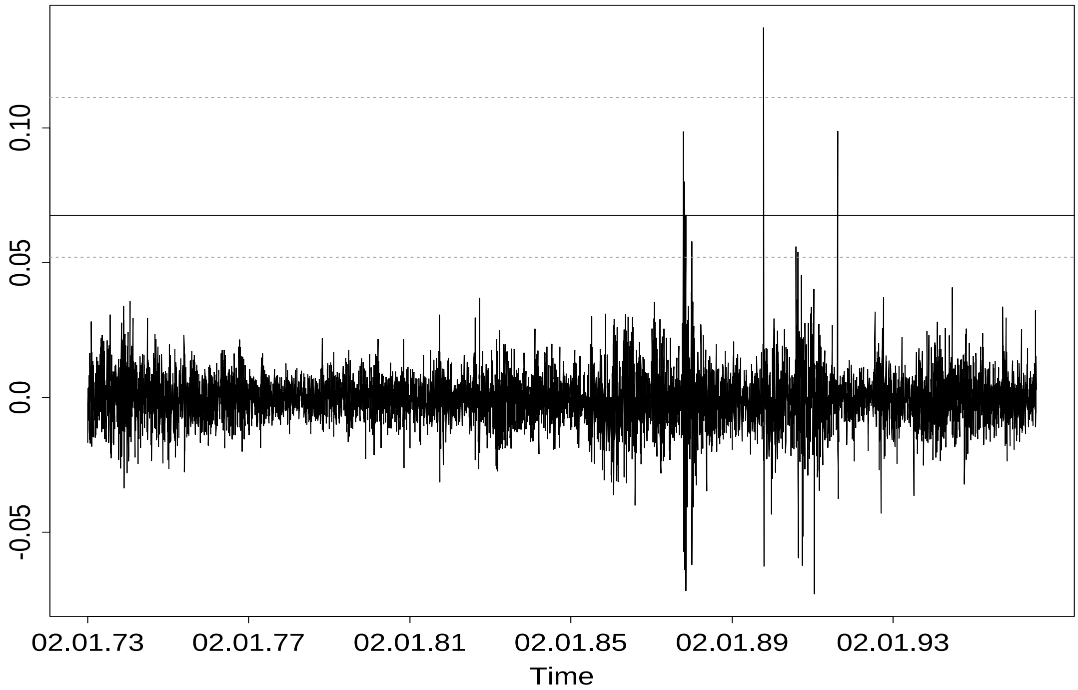

Figure 7: 20 semester return level for the DAX index (negative returns) is marked by solid line; dotted line gives 95% confidence interval.

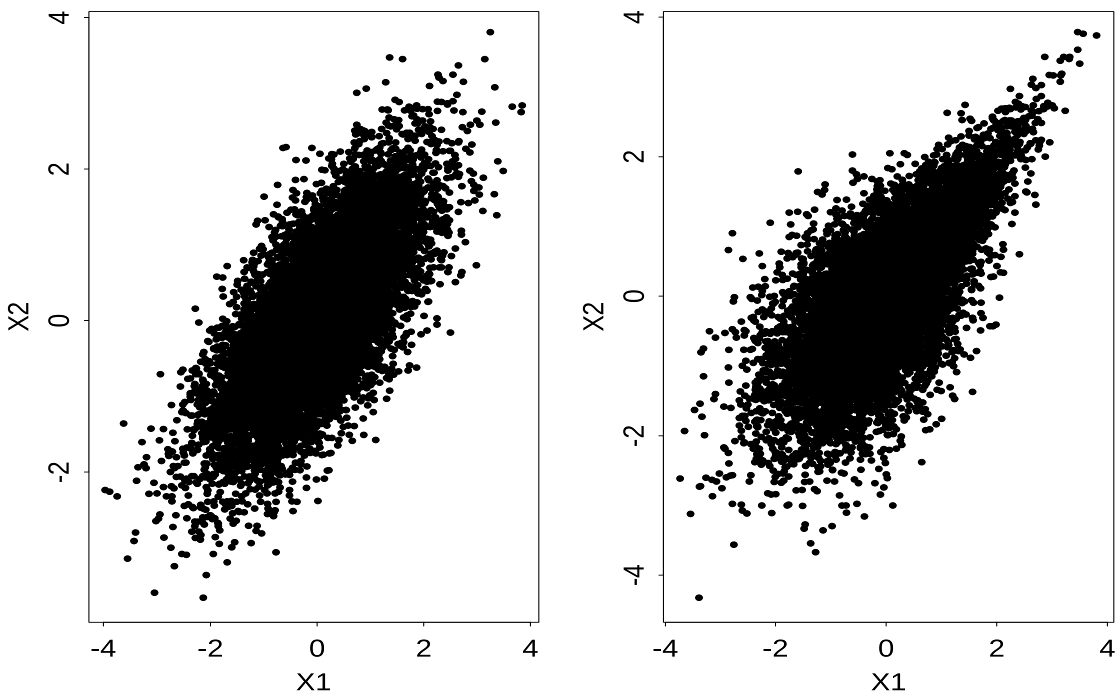

Figure 8: 10000 simulated data from two bivariate distributions with standard normal marginal distributions and correlation 0.7, but different dependence structures.