Extreme Value Theory and Statistics for Heavy Tail Data

Silvia Caserta[1]

EURONEXT N.V. and Tinbergen Institute

Casper G. de Vries[2]

Erasmus University Rotterdam and Tinbergen Institute

A scientific way of looking beyond the worst-case return is to employ statistical extreme value methods. Extreme Value Theory (EVT) shows that the probability on very large losses is eventually governed by a simple function, regardless the specific distribution that underlies the return process. This limit result can be exploited to construct semi-parametric portfolio Value at Risk (VaR) estimates around and beyond the largest observed loss. Such extreme VaR estimates can be useful inputs for scenario analysis and stress testing. The aim of this chapter is to introduce the reader to extreme value theory and the statistics of extremes.

1. Introduction

In EVT one studies the distribution of the maximum and minimum values of random variables as the sample size increases. EVT is widely used in engineering problems like the determination of dike height as a function of the highest flood levels and their frequency. Paralleling the growth of risk management, there is a recent interest for EVT in finance. EVT has proven to be a useful theoretical and statistical tool to calculate risk measures like VaR (i.e. from the optimist's point of view the portfolio value that will be exceeded with a high probability).

This chapter presents in a simple way the basic concepts of EVT and the statistical techniques used to analyse extreme market movements. We start with a visual analysis of market index returns, to emphasize the frequent occurrence of large market swings. This shows that the return distribution has heavier tails than the normal distribution. If the distribution has heavy tails, EVT shows that the probability on the most extreme loss returns is governed by a particular function. This function has the nice property that it is to a first order self-additive. This property can be exploited to reduce the computational burden of the risk manager. We introduce the statistical techniques, which are used to estimate this function, and provide a small case study.

2. Extremes in Financial Returns

Even though it is often expedient and fruitful to assume that asset returns are normally distributed, it is also well known that empirical return distributions contain an excessive amount of extremes relative to the normal model; see Campbell, Lo and MacKinlay (1997). To show this data feature we plotted in Figure 1 the log-returns

---

[1] The views expressed are those of the authors and do not necessarily represent those of Euronext.

[2] Corresponding author address: Dept. of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands; email: <cdevries@few.eur.nl>.

from the daily closing prices of the AEX stock market index over the period March 17, 1983 - May 15, 2002. One can easily recognize the various crashes and market rallies at the end of 1987, the Asian crisis in 1997, the Russian crisis in 1998 and the recent market turmoil.

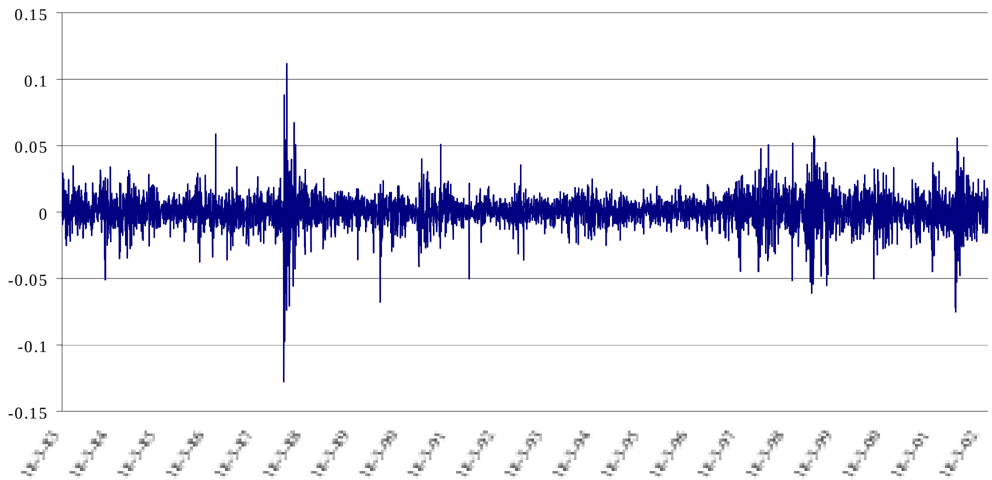

Figure 1 AEX daily logarithmic returns 13-03-1983/15-05-2002

In Figure 2 we have generated an equal amount of pseudo random numbers by drawing from a normal distribution with the same mean and standard deviation as in the AEX return data (y-axes have the same scale). Relative to the normal data, the true returns do exhibit many larger and smaller spikes, which sometimes appear in clusters. For risk management with its focus on outliers this deviation from normality is crucial and cannot be ignored. This data feature is the so-called heavy tail feature, referring to the power shape of the tails of the density (the normal has a light tail as the tails of the density fall towards zero at an exponential rate). For example a Student-t distribution would fit the picture much better.

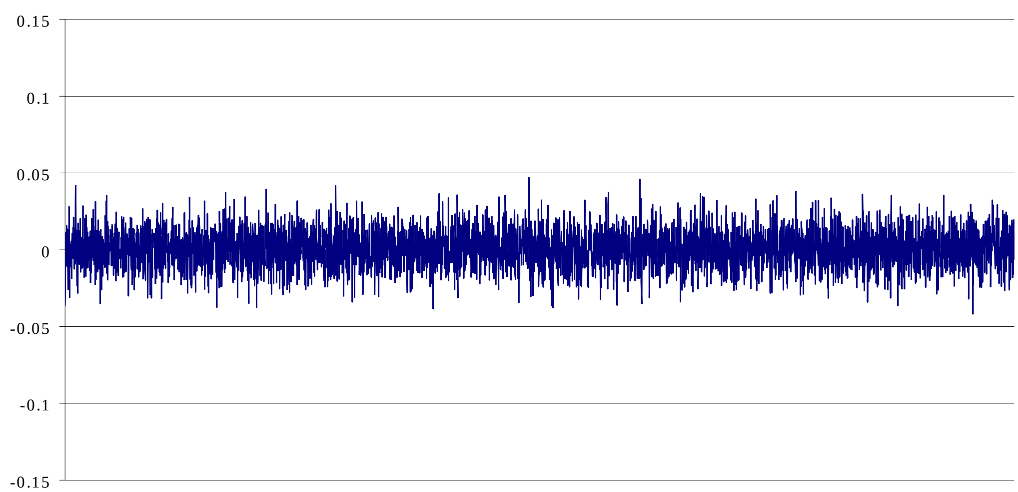

Figure 2 Normally generated returns

3. Extreme Value Theory

The EVT gives an approximation to the distribution of the maximum and minimum values of random variables as the sample size increases. The advantage is that the limit laws provided by EVT do not require a detailed knowledge of the distribution, which generates the returns (this is in analogy with the advantage of the central limit law for averages). Moreover, EVT implies that the probability on very large losses is governed by a simple function, again regardless the specific distribution that underlies the return process.[i] This limit result is exploited in risk management to construct semi-parametric portfolio Value at Risk (VaR) estimates around and beyond the largest loss.

Let $X_1, X_2, \ldots, X_n$ be a sample of random variables of size n. We can think of $X_1, X_2, \ldots, X_n$ as the returns on the AEX index. The maximum M_n and the minimum m_n of the returns are defined as

$$M_n = \max \{ X_1, X_2, \ldots, X_n \} \quad \text{and} \quad m_n = \min \{ X_1, X_2, \ldots, X_n \}.$$

The worst case VaR might thus be calculated as $-m_n$ times the portfolio value. Examples of maximum and minimum values of the AEX are reported in Table 1.

Table 1: AEX max-min returns

Period	Maximum	Minimum
15-12-2001 / 15-05-2002	2.7%	–2.2%
15-11-2001 / 15-05-2002	2.7%	–2.2%
15-10-2001 / 15-05-2002	4.1%	–4.8%
15-09-2001 / 15-05-2002	5.6%	–7.5%

When one increases the sample size by moving down in the table, the maximum increases or is unchanged and the minimum decreases or is unchanged. In the following we study the (limit) distribution of the maximum and the minimum as the sample size grows (unboundedly).

From now on we concentrate on positive random variables only, since by changing the sign on X_i one can reduce the study of minima to the study of maxima. Assume that the random variables are independent and identically distributed with (cumulative) distribution function $F(x)$. The probability that the maximum is less or equal to a pre-specified value is given by

$$\Pr\{M_n \leq x\} = \Pr\{X_1 \leq x, \ldots, X_n \leq x\} = \Pr\{X_1 \leq x\} \cdots \Pr\{X_n \leq x\} = [F(x)]^n . \qquad (1)$$

Unfortunately, $[F(x)]^n$ is mostly impractical to calculate even for moderate values of n.[ii] Worse, in most cases we do not even know $F(x)$. Fortunately, the approximation to $[F(x)]^n$ offered by EVT is very helpful.

Albeit $F(x)$ is not known, the efficiency of EVT is increased if we capitalize on the fact that the return series exhibits the spikes observed in Figure 1. To this end we need a precise definition of heavy tail distributions; see Feller (1971).

Definition 1. *We say that the distribution of the returns $F(x)$ has a heavy upper tail for the positive returns X_i, if (for large x)*

$$1-F(x)=x^{-\alpha}L(x) \quad as \quad x\to\infty, \quad \alpha>0, \tag{2}$$

and the function $L(x)$ *is such that for any x>0*

$$\lim_{t\to\infty} L(tx)/L(t)=1. \tag{3}$$

The tail of the distribution factors into two parts, the $L(x)$ function and the power part. The *L(x)* function is asymptotically unimportant since $L(tx)\approx L(x)$ for large t (one says $L(\bullet)$ varies slowly at infinity). The tail of the distribution is dominated by the power part $x^{-\alpha}$. If $L(x)$ is constant then $F(x)$ in (2) is the Pareto distribution, while in case of e.g. the Student-t distribution $L(x)$ is not constant (for the Student-t α equals the degrees of freedom). The coefficient α is called the tail index and indicates the number of bounded moments. Note that the larger is the tail index α, the less extreme is the behaviour of the returns. Due to the power part, the tail of $F(x)$ in the end always falls off more slowly than the tails of distributions such as the normal and lognormal, which have exponential like tails. Most financial data appear to be heavy tailed in this sense.

We can now state the main result from EVT. If the distribution $F(x)$ satisfies (2), then EVT shows that if the sample size n becomes large

$$\lim_{n\to\infty} \Pr\left\{M_n/a_n \le x\right\}=\lim_{n\to\infty}[\mathrm{F}(\mathrm{a_n x})]^n = e^{-x^{-\alpha}}, \ \alpha>0 \tag{4}$$

where a_n is a sequence of positive scaling numbers (needed to obtain a non-trivial limit law, cf. the central limit law).

Result (4) tells us that when we use a large but finite sample of returns, the distribution of the maximum can be approximated by

$$\Pr\left\{M_n \le x\right\}\approx G(x)=e^{-a_n{}^{\alpha}x^{-\alpha}} . \tag{5}$$

This implies an approximate density $g(x)=\alpha a_n{}^{\alpha}x^{-\alpha-1}e^{-a_n{}^{\alpha}x^{-\alpha}}$, which for large values of x reads (since $exp\left(-a_n{}^{\alpha}x^{-\alpha}\right)\to 1$ as $x\to\infty$)

$$g(x)\approx h(x)=\alpha a_n{}^{\alpha}x^{-\alpha-1} . \tag{6}$$

Note that $h(x)$ in (6) is the density of a Pareto distribution $H(x)=1-a_n{}^{\alpha}x^{-\alpha}$ on $[a_n,\infty)$. Thus the heavy tail feature of $F(x)$ is transferred to the limit distribution for the maximum.

The converse is also true. That is to say, if (4) holds for a distribution $F(x)$, then (2) is implied. To give a heuristic explanation for this, consider (5) which holds that $F^n\approx G$ or $F\approx G^{1/n}$. Upon differentiation we get an expression for the density

$$f(x)\approx\frac{1}{n}G(x)^{1/n-1}g(x)=\frac{1}{n}G(x)^{1/n}\alpha a_n{}^{\alpha}x^{-\alpha-1}.$$

One shows that the $n^{-1}G(x)^{1/n}\alpha a_n{}^{\alpha}$ part is in fact a slowly varying function, i.e. respects the limit in (3). Thus the density $f(x)$ factors into a power part $x^{-\alpha-1}$ and a slowly varying part, just like the density $g(x)$ for the maximum. Hence, the tail of the return density $f(x)$ necessarily has a Pareto like tail shape.

Thus whether we focus on the distribution of the maximum M_n in large samples, or on the distribution of very large x (extreme VaR levels), are two sides of the same coin. The two-way street relation between (2) and (4) is important when we discuss estimation. To summarize, the above shows that if we are only interested in the tail behaviour and the occurrence of extreme values, we do not need to model a specific distribution $F(x)$ of the returns. Instead we can proceed by estimating the 'Pareto factor' $ax^{-\alpha}$ for large values of x.

In case the distribution does not exhibit the heavy tail feature, the EVT implies that one of two other limit distributions may apply. Which case applies depends on whether the distribution has finite endpoints, or has exponential like tails. See Reiss and Thomas (2001) or Embrechts, Klueppelberg and Mikosch (1997) for further details.

In risk management applications one usually calculates the VaR for several time horizons (e.g. for internal use and regulatory purposes). This would in principle require re-estimation of the Pareto factor at the different frequencies. However, due to an important additivity property of distributions which satisfy (2), such repeated estimation is unnecessary. Recall that the sum of two consecutive daily log-returns is equal to the two-day log-return. Fortunately, if two random variables are heavy tailed, then the distribution of their sum is also heavy tailed. Specifically, Feller (1971) showed that if X_1, X_2 are independent and satisfy (2), we have that

$$\Pr\{X_1 + X_2 > x\} \approx 2ax^{-\alpha} \qquad x > 0, \quad \alpha > 0 \tag{7}$$

where $X_1 + X_2$ is the two-day logarithmic return. Hence, once the scale factor a and the tail index α have been estimated, one can rescale linearly the probabilities for the desired time horizon.

Conversely, if the probability level is kept constant, one can adjust the VaR level for the time horizon by inverting (7). This yields the alpha-root of time rule.

Proposition (The α-root rule): The extreme returns estimates over T-days are equal to the one-day extreme returns estimates multiplied by the alpha root of the considered time horizon T.

Compare this result $T^{1/\alpha}$ to the case of the normal distribution when the scaling has to be done by the square root $T^{1/2}$ of the time horizon T. For many financial data one finds that $\alpha > 2$ so that the scaling factor is smaller than in case of normality, i.e. using the normal scaling factor induces overly conservative capital levels at longer horizons (while using the normal model to calculate VaR levels at short horizons may be imprudent).

4. Statistics of Extremes

For applications the norming constant and the tail index need to be estimated to obtain the Pareto factor. One possible estimation procedure is to create sub-sample maxima, and apply maximum likelihood to (5); see e.g. Longin (2000). Since there can be multiple extreme realizations in a single subsample, an efficient use of the data is to exploit the tight connection between (4) and (2), and use all realizations above a certain high threshold s, say, to estimate the tail part of the unknown density $f(x)$; see e.g. Hols and De Vries (1991) or Danielsson and De Vries (2000) for this approach.

In this semi-parametric set-up the tail index α is commonly estimated by means of Hill's procedure; see Box 1.

BOX 1 The Hill estimator

The Hill estimator is motivated by the maximum likelihood estimator for the power coefficient of the Pareto density $h(x)$ in (6). First note that the conditional Pareto density reads $h(x \mid x > s) = \alpha(x/s)^{-\alpha-1}s^{-1}$. Taking logarithms and differentiating with respect to α yields

$$\partial \log h(x \mid x > s)/\partial\alpha = 1/\alpha - \log(x/s).$$

The Hill estimator is found by equating this first order condition to zero, replacing x with realizations $X_i > s$, and to sum over these elements. Solving for $1/\alpha$ gives

$$\frac{1}{\hat{\alpha}} = \frac{1}{k}\sum_{i=1}^{k}\log\frac{X_{(i)}}{s},$$

where the $X_{(i)}$ are the largest descending order statistics $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(k)} \geq s \geq X_{(k+1)} \geq \ldots \geq X_{(n)}$, pertaining to the sample X_1, X_2, …, X_n, and where n is the number of observations. The number k is the number of extreme returns above s, where s is the point from where the Pareto approximation applies.

Once the tail index has been estimated, we can directly estimate large quantiles (returns x), i.e. there is no need to first estimate the scaling constant a in the Pareto factor. This is explained further in Box 2.

BOX 2 The Quantile Estimator

Large quantile estimates can be obtained as follows. Consider two tail probabilities $p = 1-F(x_p)$ and $t = 1-F(x_t)$. For example, take $p < 1/n < t \leq k/n$, where n is the number of observations and k is the number of extreme returns we use. Let x_p and x_t denote the quantiles corresponding to the probability levels p and t. Then, by the Pareto approximation we deduce that $p \approx a\, x_p^{-\alpha}$ and $t \approx a\, x_t^{-\alpha}$. Combining these two expressions yields $x_p \approx x_t (t/p)^{1/\alpha}$. Let $t \approx m/n$, m integer, be the empirical distribution function at t. Replace x_t by X_{m+1}, which is the m+1 largest positive return. This provides the following quantile estimator

$$\hat{x}_p = X_{m+1}\left(\frac{m}{np}\right)^{1/\hat{\alpha}},$$

where $\hat{\alpha}$ is the Hill estimator of the tail index α; see de Haan et. al. (1994).

We now illustrate these techniques on the daily AEX returns as described in Section 1. Using the Hill estimator we find that for the positive returns $\alpha = 3.0$ (k = 119, n = 4849), and for the negative returns $\alpha = 2.6$ (k = 153). These values confirm the typical results from academic research that the tail index is so low that only the first few moments are bounded, and that the tail index for the positive returns is larger than the index for the negative returns.

Extreme return estimates are presented in Table 2, using $\alpha = 3$. The risks of loss (or gain) are expressed as events that occur once per so many years, and are given in the first column. The associated return estimates appear in columns two (gains) and three (losses). For example, the 25 Y Pr. Level means that on average once per 25 years there is a one-day positive return higher than 12%. The estimates for the two-day returns are obtained by using the alpha-root of time rule, i.e. by multiplication of the one-day return estimates with the factor $2^{1/3}$. Other methods typically underestimate the loss (and gain) levels reported in this table and hence their relevance for risk management.

Table 2: AEX index estimated returns

Daily extreme returns estimates		
Pr. Level	Positive	(–) Negative
25 Y	12.0%	12.5%
20 Y	11.0%	11.5%
15 Y	10.5%	10.5%
Two-day extreme returns estimates		
25 Y	15.1%	15.7%
20 Y	13.9%	14.5%
15 Y	13.0%	13.2%

5. Conclusions

In this chapter we have presented the basic concepts of extreme value theory and the statistics of extremes. Particularly, we have shown how this theory is beneficial to the investigation of the occurrence of large and as of yet unseen market movements.

The way in which we have presented the theory relied on the assumption that the returns are independently distributed. Empirical research has shown that even though autocorrelation in the returns is very low, there are clusters of volatility (dependence in the second moment). Fortunately, the theory of extremes also holds for dependent variables, see McNeil and Frey (2000) who study conditional VaR. Nevertheless, the alpha-root of time rule has to be adapted when the data are dependent. Apart from time dependency, the cross sectional dependency is also of interest for the co-dependence between different asset markets in view of possible systemic risk. The multivariate dependency can be captured via a copula function, see chapter **XXX** of this book.

The long and short of all this is that it is essential for risk managers to peek beyond the sample, for which EVT offers a reliable and coherent method.

---

[i] An introductory textbook to EVT with many applications is Reiss and Thomas (2001). A rigorous treatise is Embrechts, Klueppelberg and Mikosch (1997).

[ii] As simple example, suppose that *F(x)* is the normal distribution with mean zero and variance one: $F(x)=\Pr\{X\leq x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-x^2/2}dx$. Then $[F(x)]^n$ can not be directly solved already for n larger than three.

Bibliography

Cambell, J. Y., A. W. Lo and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton: Princeton University Press).

Danielsson, J., and C.G. de Vries, 2000, Value-at-Risk and Extreme Returns", *Annales D'Economie et de Statistique,* 60, pp. 239-270.

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

Feller, W., 1971, *An Introduction to Probability Theory and its Applications*, Second Edition, Volume II (New York: John Wiley & Sons).

Haan, de L., D. W. Jansen, K. Koedijk and C. G. de Vries, 1994, "Safety First Portfolio Selection, Extreme Value Theory and Long Run Asset Risks", in J. Galambos, J. Lechner, E. Simiu and N. Macri (eds), *Extreme Value Theory and Applications*, pp. 471-487.

Hols, M., and C.G. de Vries, 1991, "The Limiting Distribution of Extremal Exchange Rate Yields", *Journal of Applied Econometrics,* 6, pp. 287-302.

Longin, F. M., 2000, "From VaR to Stress Testing: The Extreme Value Approach", *Journal of Banking and Finance*, 24, pp. and 1097-1130.

McNeill, A., and R. Frey, 2000, "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach", *Journal of Empirical Finance,* 7, pp. 271-300.

Reiss, R. D., and M. Thomas, 2001, *Statistical Analysis of Extreme Values*, Second Edition (Basle: Birkhauser).