Extreme Value Theory and its Applications to Financial Risk Management

Tsevas, G. and Panaretos, John

1998

HERCMA '98
E. A. Lipitakis (Editor)
1998, LEA
Printed in Hellas

EXTREME VALUE THEORY AND ITS APPLICATIONS TO FINANCIAL RISK MANAGEMENT

G. Tsevas and J. Panaretos
Department of Statistics,
Athens University of Economics and Business,
76 Patission St., 104 34 Athens, Greece,
e-mail for correspondence: tsevas@stat-athens.aueb.gr;

Abstract

The phenomenon of high volatility in financial markets stemming from the increased complexity of financial instruments traded, as well as the evidence of losses due to natural and man-made catastrophes, highlight the need for sophisticated risk management practices. The analysis concerning the statistical distribution of extreme events (e.g. stock market crashes), is considered to be important for modern risk management. In this review paper, an introduction to the basic results of Extreme Value Theory (EVT) is made. More specifically, the methodological basis of EVT for quantile estimation is introduced. Moreover, EVT methods for estimating conditional probabilities concerning tail events, given that we incur a loss beyond a certain threshold u, are presented. Finally, the application of the theory is demonstrated by considering an example using equity return data.

Keywords: *central limit theorem, standard extreme value distributions, quantiles, mean excess function, Value-At-Risk, shortfall distribution, peaks over threshold-method.*

1. Introduction

The presence of high volatility in financial markets and in general the huge financial costs arising from the existence of extreme events, have resulted in the establishment of Risk Management as a key operation in every financial institution or corporation. The banking and insurance industry is continuously under the development of new products that will manage financial risk. Methods such as the estimation of Value-at-Risk (VAR) and the so-called shortfall have been introduced in the world of finance.

Given the significance of extreme events, Extreme Value Theory (EVT) yields methods for quantifying such events and their consequences in a statistically optimal way. In general, EVT is a subject whose motivations match the following points: a) Risk Management is interested in methods leading to estimation of tail probabilities concerning tail events and quantiles of profit-loss distributions and b) Financial data exhibit fat tails.

Following this small introduction on the importance of EVT in the new risky environment of finance, we present an outline of our review paper. Firstly, section 2 serves as an introductory part where the relation between EVT and the Central Limit Theorem (CLT), as well as some main results of EVT, are presented. Secondly, in section 3 we introduce the analysis on tail and quantile estimation. More analytically, this section starts with an introduction on the statistical inference of the p-quantile and continues with a description of two main graphical techniques for statistical data fitting. Moreover, a subsection is devoted on the VAR-measure and the "shortfall" d.f., whereas a description of the Peaks over Threshold (POT) model and method closes section 3. Finally, section 4 presents a description of an application of EVT on equity return data, whereas in section 5 some conclusions are drawn.

2. Basic extreme value theory

In Embrechts, Bassi and Kafetzaki (1997), a comparison is made between the CLT and the main results from Extreme Value Theory (EVT). The reasons for this comparison is that CLT forms the mathematical background on which EVT is based.

More specifically, according to CLT, we suppose that $X_1,\dots, X_n$ are independent and identically distributed (i.i.d.) random variables with distribution function (d.f.) F. If S_n is the sum of $X_1,\dots, X_n$, then the CLT solves the following problems:

- The finding of constants $\alpha_n>0$ and $b_n \in \Re$ so that, for given F it holds:

$$(S_n-b_n)/\alpha_n \xrightarrow{d} G, \ n\to\infty, \tag{1}$$

  where: G is a non-degenerate d.f.
- The characterization of the d.f. G in (1) above.
- The solution of the *domain of attraction problem* or the finding of all the d.f.'s F that satisfy (1) given a possible d.f. G and appropriate sequences (α_n) and (b_n) (i.e. $F\in D(G)$).

EVT starts appearing when S_n is replaced by $M_n = X_{1,n} = \max(X_1,\dots, X_n)$ where M_n represents the maximum of a sequence of n (i.i.d) r.v's. The area of study for EVT is the limiting distribution of M_n, appropriately scaled. Related to the above is the following theorem on the limit laws for maxima that is due to Fisher-Tippett (1928) whose result forms the basis of classical extreme value theory:

Theorem 2.1. *Suppose that $X_1, \dots, X_n$ are i.i.d. r.v.'s with d.f. F. If there exist suitable norming constants $\alpha_n>0$ and $b_n\in\Re$ so that:*

$$(M_n - b_n)/\alpha_n \xrightarrow{d} G, \ \ n\to\infty, \tag{2}$$

where G is a non-degenerate d.f., then G(x) belongs to the type of one of the following three d.f.'s:

1. Gumbel: $$\Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \Re \tag{3}$$

2. Frechet: $$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \quad \text{for } \alpha > 0 \tag{4}$$

3. Weibull: $$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \text{for } \alpha > 0 \tag{5}$$

The above three d.f.'s are called *standard extreme value d.f.'s* and the corresponding r.v.'s are called *standard extremal r.v.'s*.

However, a one-parameter representation of the three standard cases in one family of d.f.'s is widely accepted as the *standard* representation and is referred to as the *generalized extreme value distribution (GEV)* $\mathrm{H}_\xi(x)$. That is :

$$\mathrm{H}_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\} & \text{if } \xi \neq 0 \\ \exp\{-\exp\{-x\}\} & \text{if } \xi = 0 \end{cases}, \quad \text{where } 1+\xi x > 0 \tag{6}$$

Thus, the case $\xi = -\alpha^{-1} < 0$ corresponds to the Weibull distribution, $\xi = 0$ to the Gumbel distribution and $\xi = \alpha^{-1} > 0$ to the Frechet distribution.

However, we can also refer to the related location-scale family $H_{\xi;\,\mu,\,\psi}$ as GEV by replacing the argument x above by $(x-\mu)/\psi$ for $\mu \in \Re$, $\psi > 0$. Thus, we have:

$$H_\vartheta(x) = H_{\xi;\mu,\psi}(x) = \exp\left\{-(1+\xi\frac{x-\mu}{\psi})^{-1/\xi}\right\}, \quad 1+\xi\frac{x-\mu}{\psi} > 0 \tag{7}$$

where the parameter $\theta = (\xi, \mu, \psi) \in \Re\times\Re\times\Re_+$ consists of a shape parameter ξ, location parameter μ and scale parameter ψ.

3. Tail and Quantile Estimation

The use of extreme value theory in finance and insurance applications can be more easily justified by the fact that it forms the methodological basis on which tail and quantile estimation is based.

3.1. An introduction of the statistical inference of the p-quantile

Suppose F is a d.f. of a real-valued r.v. X. Given a threshold probability $p \in (0, 1)$, the p-quantile x_p is defined as:

$$x_p = F^{\leftarrow}(p) = \inf\{x \in \Re: F(x) \geq 0\} \tag{8}$$

For the statistical inference of the p-quantile x_p we suppose that the data $X_1, ..., X_n$ are i.i.d. with distribution function F. If we take the *order statistics* of the data $X_1, ..., X_n$, an obvious estimator of the p-quantile x_p for k = 1, 2,..., n is the following:

$$\hat{x}_{p,n} = F_n^{\leftarrow}(p) = X_{k,n}, \quad \text{for } 1-(k/n) < p \leq 1-(k-1)/n, \tag{9}$$

where $X_{k,n}$ is the kth largest observation

$F_n^{\leftarrow}(p)$ is the *quantile function* of the empirical d.f. F_n given by:

$F_n(x)=(\#\{i: 1 \leq i \leq n \text{ and } X_i \leq x\}) / n, \ x \in \Re$

Using the CLT we can show that for a distribution function F with continuous density f with $f(x_p) \neq 0$ and k = k(n), so that $n-k = np + o(n^{1/2})$ with $o(n^{1/2})$ a negligible quantity, it holds:

$$\hat{x}_{p,n} \sim AN(x_p, (p(1-p))/(nf^2(x_p))), \tag{10}$$

which means that the estimated p-quantile is asymptotically normally distributed. The asymptotic distribution of this estimator allows for estimation of p close to 1 given that sufficient data are available. As a result, an important point is the choice of k = k(n).

Furthermore, using the i.i.d. condition on $X_1, ..., X_n$, the following confidence interval can be obtained for i < j:

$$P(X_{j,n} \leq x_p < X_{i,n}) = \sum_{r=i}^{j-1} \binom{n}{r} p^{n-r}(1-p)^r \tag{11}$$

However, exact confidence intervals can be produced for n not too large. Moreover, p has to be such that x_p lies within the range of the data so as to obtain exact confidence intervals.

For quantile estimation for p close to 1 extra (parametric) conditions on F must be assumed and therefore our data must be statistically fitted in order to see whether these conditions hold. Standard methods for data fitting include likelihood techniques, goodness-of-fit procedures and Bayesian methods. Two main graphical techniques for data fitting are the following:

- ***QQ-(or quantile) plot:*** This technique represents a graphical goodness-of-fit procedure.

From the QQ-plot we can get a quick assessment of the plausibility of certain distributional model assumptions. The main use of it is when we want to test whether a sample $X_1, ... , X_n$ comes from a certain distribution. More specifically, we take the ordered sample and we suppose that a parametric model $F(.;\theta)$ is fitted to $X_1, ..., X_n$. The next step involves the construction of a plot from the points:

$$\{(X_{k,n}, F^{\leftarrow}(p_{k,n})) : k=1,...,n\}$$

where $X_{k,n}$ is the kth largest observation and $(p_{k,n})$ a plotting sequence which is often taken as: $p_{k,n} = (n-k+1)/(n+1)$, or some variant of this.

A good fit results when a straight line is observed on the plot. In other words, if a certain distribution provides a good fit to our data, then this QQ-plot should look roughly linear. Model deviations (like heavier tails, skewness, outliers, ...) can be easily observed by the plot.

- ***Mean-excess plot:*** This second graphical technique that particularly enables discrimination in the tails, is based on the quantity e(u) known as the *mean excess function (mef)*:

$$e(u) = E(X-u \mid X>u), \quad u \geq 0 \tag{12}$$

and its empirical counterpart given by

$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)^+}{\sum_{i=1}^{n} I_{\{X_i > u\}}}, \tag{13}$$

where $y^+ = \max(y, 0)$ and I_A denotes the indicator function of A which means that

$I_A(x) = 0$ for $x \notin A$ and $I_A(x) = 1$ for $x \in A$., where $A=\{X_i > u\}$.

By plotting $\{(u, e_n(u)) : u \geq 0\}$ we can distinguish between short and long-tailed distributions. More specifically, long-tailed d.f.'s exhibit an upwards sloping behavior, exponential-type d.f.'s are characterized by a constant mean excess plot, whereas the plot for short-tailed data shows a decreasing pattern (decreases to 0).

Given this first approach on tail and quantile estimation, its use in finance is highlighted in the next section.

3.2. On the Value-at-Risk (VAR) measure and beyond

A definition of VAR should precede our analysis. A technical and clearly stated definition for the VAR of a portfolio is the following: *VAR is an amount of money such that a portfolio will lose less than that amount over a specified period with a specified probability.*

The VAR of a portfolio can be translated as a possible extreme loss that results from holding the portfolio for a fixed period using as a measure of risk the volatility over the last 100 trading days (P&L-distribution). In other words, if X is a random variable representing the portfolio losses, we are interested to calculate, for a given (small) confidence level α, the level (α-quantile) u_α so that $P(X > u_\alpha) = 1-F(u_\alpha) = \alpha$, assuming that we have absolute loss data.

It is obvious that for sufficiently small values of α, this quantile estimation will typically be outside (or at the edge of) the range of available data. Consequently, extrapolation beyond the range of the data is required. On the other hand, having determined the level u_α, another area of interest concerns the estimation of the potential losses above u_α. Therefore, we need to estimate the conditional probability d.f., also known as the excess d.f. $F_{u_\alpha}(x)$:

$$F_{u_\alpha}(x) = P(X-u_\alpha \leq x \mid X > u_\alpha), \quad x \geq 0, \tag{14}$$

An estimate of this conditional probability will (in the case of sufficient data) involve the losses $X_{1,n}, \ldots, X_{k-1,n}, X_{k,n}$ above some (large) loss $X_{k+1,n}$. If our data are insufficient, a suitable model or approximation for (14) has to be found. This approximation is the *(conditional) mean excess loss* given that a loss above u_α (VAR) has occurred. This quantity, which is known as "shortfall" or "beyond VAR" within the finance industry, is given by the following equation:

$$e(u_\alpha) = E(X - u_\alpha \mid X > u_\alpha) \tag{15}$$

Given this background, the main result concerning the estimation of conditional (also known as *overshoot*) probability d.f.'s (14) assuming that our losses are denoted as positive (one-sided distribution), is presented by the following theorem:

Theorem 3.2.1. *Suppose that F is a d.f. with excess distributions F_u, $u \geq 0$. Then, for $\xi \in \Re$, $F \in MDA(H_\xi)$ if and only if there exists a positive measurable function $\beta(u)$ so that:*

$$\lim_{u \uparrow x_F} \sup_{0 \leq x \leq x_F - u} \left| \bar{F}_u(x) - \bar{G}_{\xi,\beta(u)}(x) \right| = 0, \quad \text{where} \tag{16}$$

$\bar{F}_u(x) = 1 - F_u(x)$ *is the tail of F,*

$x_F = \sup\{x \in \Re : F(x) < 1\}$ is the right-endpoint of a df F,

$G_{\xi,\beta}(x)$ is the generalized Pareto distribution (GPD) with parameters $\xi \in \Re$, $\beta > 0$ and is defined as :

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + (\xi x / \beta)\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp(-x/\beta), & \xi = 0 \end{cases} \tag{17}$$

This theorem answers the crucial question of estimating beyond VAR (or shortfall). In other words, under reasonable conditions on F, there exists a canonical class of d.f.'s (GP d.f.'s) approximating the excess d.f. in (14). The so-called ***threshold u*** has to be taken large ($u \uparrow \infty$). Therefore, from this approximation an upper-tail estimator can be obtained.

For x, u $\geq$ 0, we have:

$$\bar{F}(u + x) = \bar{F}_u(x)\bar{F}(u) \approx \bar{G}_{\xi,\beta(u)}(x)\bar{F}(u) \tag{18}$$

$$\hat{\bar{F}}(u + x) = \bar{G}_{\hat{\xi},\hat{\beta}}(x)\frac{N_u}{n}, \tag{19}$$

where $\hat{\xi}, \hat{\beta}$ are estimators of ξ and β that can be obtained through various methods

$N_u = \#\{i : 1 \leq i \leq n,\ X_i > u\}$: the number of exceedances.

From the estimate of $\overline{F}$ above (19) we can obtain the quantile estimator

$$\hat{x}_p = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\xi}} - 1\right) \tag{20}$$

3.3. Peaks Over Threshold (POT) model and method

1. ***POT model:*** The model assumes that excesses of an i.i.d. sequence over a high threshold u occur at the times of a Poisson process and the corresponding excesses over u are independent and follow a GPD. Moreover, the model assumes that excesses and exceedance times are independent of each other.
2. ***POT method:*** The method is described by the following steps:

Step 1. The data are $X_1, \ldots, X_n$, and for a given threshold u, denote N_u the number of exceedances of the level u.

Step 2. Set a value for u that is large enough so that the approximation (18) is valid. Here, we have to take into account that a value of u too high results in too few exceedances and consequently high variance estimators. On the other hand, for u too small, estimators become biased. One method to determine the optimal value of u is to look for a u-value above which the m.e.f. looks linear.

Step 3. Attempt to fit the GP distribution to the excesses over a level u. This process will involve estimation of the parameters ξ, β through estimation procedures such as the maximum likelihood method. Following this process, a tail fit (19) and the estimator for the quantile x_p (20) can be obtained.

Step 4. Repeat step 3 for various u-values. This involves refitting a model for each u-value.

4. A Description of an Application on Equity Return Data

Embrechts, Resnick and Samorodnitsky (1997), applied some of the techniques from EVT on daily equity return data from the automobile company BMW. The chosen period was from January 2, 1973 until July 23, 1996. They focused on the left tail (i.e. negative daily return values) of the d.f. F of their data. Their sample resulted in n = 2770 observations.

They first constructed the mean excess plot that showed an increasing behavior from u equal to 0.02 and above, clearly indicating fat (even Pareto type) tails. Using the maximum likelihood method, they estimated ξ and β as a function of u and k (the number of exceedances of u). For example, the estimate of the shape parameter ξ was found to have a value of 0.223 that corresponds to a Pareto tail with value $1/\xi = 4.484$. From these estimates, using (19), an estimate of the (conditional) excess d.f. $F_u(x)$ can be plotted. More specifically, they plotted the GPD-fit to the shifted d.f. F_u (x-u), $x \geq u$, on log-scale. From this plot, one can find the conditional probability of high excesses given that we have an exceedance of u=0.02.

They also wanted to estimate the tail of the unconditional d.f. $\bar{F}(u)$ which yields information on the frequency with which a given high level u is exceeded. Therefore, using the parameter estimates corresponding to u=0.02, the tail-fit for $\bar{F}(u)$ (as in (19)) on doubly logarithmic scale was plotted. In general, though they extended the GPD-fit to the left of u=0.02, only the range above this u-value is relevant. The fitting method is only designed for the tail. Below u (where typically data are abundant) one can use a smooth version of the empirical d.f. From this last plot, quantile estimates were deduced, together with 95% confidence intervals (c.i.'s). For example, the estimate for the 99.9% quantile $X_{0.999}$ (or $u_\alpha = u_{0.001}$) was found to be 0.081. The 95% c.i. for $X_{0.999}$ was found to range between 0.070 and 0.101.

5. Conclusion

In this review we attempted to summarise the most basic areas of EVT by presenting the mathematical analysis behind it. We also examined how EVT can be used in order to answer questions dealing with extreme cases within the finance industry.

References

Embrechts, P., Bassi, F. and Kafetzaki, M. (1997). *A survival kit on quantile estimation*, http://www. math.ethz.ch/~baltes/ftp/papers.html.

Embrechts, P., Resnick, S. and Samorodnitsky, G. (1997). *Living on the edge.* http://www. math.ethz.ch/~baltes/ftp/papers.html.

Embrechts, P., Resnick, S. and Samorodnitsky, G. (1998). *Extreme value theory as a risk management tool*, http://www. math.ethz.ch/~baltes/ftp/papers.html.

Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer-Verlag Berlin Heidelberg New York.

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.* 24, 180-190.

Gnedenko, B.V. and Kolmogorov, A.N. (1954). *Limit Theorems for Sums of Independent random Variables*. Addison-Wesley, Cambridge, Mass.

Jorion, P. (1997). *Value at risk: The new benchmark for controlling market risk.* McGraw-Hill Companies, Inc.

Tsevas, G. and Panaretos, J. (1998). *Extreme value theory and its applications to financial risk management*. Technical Report No 48. Department of Statistics. Athens University of Economics and Business.