A.XI. Introduction to Extreme Value Theory

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Outline

Introduction

EVT

$$= \text{Extreme Value Theory}$$
$$= \text{probability theory of extreme events}$$

No matter the shape of the center of the distribution, the shape of the tail takes always very special forms when we are far enough in the tail.

Distribution of Maxima: Fisher-Tippet Theorem

The Fisher-Tippet theorem specifies the form of the limit distribution for centered and normalized maxima.

Let $X_1, ..., X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables and $M_n = \max(X_1, ..., X_n)$. For adequate choices of $c_n > 0$ and $d_n \in \Re$ such that

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} H$$

for some non degenerate distribution function H:

$\Rightarrow H$ belongs to one of the three families of extreme value distribution functions: Frechet, Weibull, Gumbell

CDF of Distributions of Maxima

Their cumulative distribution functions are given by:

Frechet:

$$F(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left(-x^{-\alpha}\right), & x > 0, \end{cases} \qquad \alpha > 0.$$

Weibull:

$$F(x) = \begin{cases} \exp\left(-(-x)^{\alpha}\right), & x \leq 0, \\ 1, & x > 0 \end{cases} \qquad \alpha > 0.$$

Gumbel:

$$F(x) = \exp\left(-e^{-x}\right), \qquad x \in \Re.$$

These extreme value distributions represent the limit laws for the *normalized maxima* of i.i.d. random variables.

Distribution of Excesses over High Thresholds

The theorem of Pickands-Balkema-De Haan specifies the form of the limit distribution of *scaled excesses* over high thresholds.

The limiting distribution is the Generalized Pareto Distribution (GPD).

The Generalized Pareto Distribution (GPD)

The generalised Pareto distribution (GPD)

$$Z \sim GPD(\xi, u, \sigma)$$

has cumulative distribution function:

$$F_{\xi,u,\sigma}(z) = 1 - \left(1 + \xi\left(\frac{z-u}{\sigma}\right)\right)^{-1/\xi}, \quad \text{for} \quad \xi \neq 0,$$

with a support equal to the strictly half line if $\xi > 0$.

The parameter ξ determines the shape of the distribution, and in particular the decay of the right tail (heavy tailed behavior).

The parameter σ is a scale parameter, and the location parameter u is the threshold value.

The Generalized Pareto Distribution (GPD)

The mean and variance are equal to:

$$\begin{aligned} E[Z] &= u + \frac{\sigma}{1-\xi}, \\ V[Z] &= \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}, \qquad 0 < \xi < 1/2. \end{aligned}$$

The excesses $Z - u$ above u will be distributed as

$$Z - u \sim GPD(\xi, 0, \sigma).$$

VaR and ES Estimation Using the GPD

Main Idea: Fit parameters of GPD to data above a given threshold u, and compute VaR and ES using associated explicit forms.

The parameters ξ and σ can be estimated either by maximum likelihood or by a method of moments.

Although asymptotically the most efficient method, maximum likelihood does not display its efficiency over the method of moments even in samples as large as 500.

Therefore the method of moments may be favored on the basis of the simplicity of the estimators:

$$\widehat{\xi} = \frac{1}{2}\left(1 - \frac{(\bar{m}-u)^2}{S^2}\right), \qquad \widehat{\sigma} = \frac{\bar{m}-u}{2}\left(\frac{(\bar{m}-u)^2}{S^2} + 1\right)$$

where $\bar{m}$ is the empirical mean and S^2 is the empirical variance.

VaR and ES Estimation Using the GPD

Both estimators are asymptotically normally distributed as:

$$\sqrt{T}\begin{pmatrix} \hat{\xi}-\xi \\ \hat{\sigma}-\sigma \end{pmatrix} \sim N(0,V),$$

with

$$V = \frac{(1-\xi)^2}{(1-2\xi)(1-3\xi)(1-4\xi)} \times \begin{pmatrix} (1-2\xi)^2(1-\xi+6\xi^2) & \sigma(1-2\xi)(1-4\xi+12\xi^2) \\ \sigma(1-2\xi)(1-4\xi+12\xi^2) & 2\sigma^2(1-6\xi+12\xi^2) \end{pmatrix}.$$

A consistent estimate $\hat{V}$ of V can be obtained by substituting $\hat{\xi}, \hat{\sigma}$ for ξ, σ.

The POT Method

The goal in the peaks-over-threshold (POT) method is to fit the GPD using the estimators $\hat{\xi}, \hat{\sigma}$ to the data above a chosen threshold u, and then to deduce the VaR and ES from the fitted distribution.

Indeed let us assume that the losses Y conditionally to be above u are distributed as $GPD(\xi, u, \sigma)$, or equivalently the excesses $Y - u$ conditionally to Y being above u are distributed as $GPD(\xi, 0, \sigma)$. Then it can be shown that for $y > u$:

$$\begin{aligned} P[Y \leq y] &= (1 - P[Y \leq u])\, F_{\xi,u,\sigma}(y) + P[Y \leq u] \\ &= F_{\xi,\tilde{\mu},\tilde{\sigma}}(y), \end{aligned}$$

with $\tilde{\sigma} = \sigma\,(1 - P[Y \leq u])^{\xi}$, and

$$\tilde{\mu} = u - \frac{\tilde{\sigma}\left((1 - P[Y \leq u])^{-\xi} - 1\right)}{\xi}.$$

The POT Method

So the probability of getting a loss below the value y (itself larger than the threshold u) is given by the c.d.f. of a generalized pareto distribution with the same shape parameter ξ, but modified location and scale parameters $\tilde{\mu}, \tilde{\sigma}$.

Estimates of ξ and σ are obtained from computing the method of moments estimators on the data above the threshold u, while $P[Y \leq u]$ is estimated by the empirical distribution evaluated at u (proportion of data below the threshold).

Once the parameters of the GPD have been estimated, it is sufficient to plug estimated parameter values in the explicit formula for the VaR.

The POT Method

Inverting the equality defining the VaR:

$$P[Y \leq VaR] = F_{\xi,\tilde{\mu},\tilde{\sigma}}(VaR) = p,$$

we obtain

$$VaR = \tilde{\mu} + \frac{\tilde{\sigma}}{\xi}\left((1-p)^{-\xi} - 1\right).$$

Since we have postulated that the excesses $Y - u$ conditionally that Y is above u are $GPD(\xi, 0, \sigma)$, ES takes the explicit form:

$$\begin{aligned} ES &= E[Y \,|\, Y > VaR] \\ &= VaR + E[(Y-u) - (VaR - u) \,|\, (Y-u) > (VaR - u)] \\ &= VaR\left(\frac{1}{1-\xi} + \frac{\sigma - \xi u}{(1-\xi) VaR}\right). \end{aligned}$$

The POT Method

The choice of the threshold u should balance bias and variance.

If u is chosen very large, we have very few data and the estimation will be imprecise (large variance).

If the threshold is chosen too small, estimates can be biased since the theorem of EVT justifying the approach may not apply.

Computation of the asymptotic variances of the estimated values $\hat{\xi}, \hat{\sigma}$ could help to assess this tradeoff. Empirical quantiles may also help set the threshold.

It can be chosen for example equal to the 90% or 95% quantile of the empirical loss distribution.

Finally we may also examine the stability of the estimates or of the loglikelihood.