Extreme Value Theory (EVT)

[Nematrian website page: ExtremeValueTheory, © Nematrian 2015]

Abstract

The pages that follow provide a brief introduction to Extreme Value Theory (EVT). EVT is a well-established branch of statistics that has been employed in insurance problems for many years but has only more recently been applied in the risk management field.

1. Introduction
2. Caveats
3. Main Block maxima results and the Fisher-Tippett, Gnedenko theorem
4. Block maxima results – examples of different limiting behaviour

1. Introduction

[ExtremeValueTheory1]

Extreme Value Theory (EVT) attempts to provide a complete characterisation of the tail behaviour of all types of probability distributions, arguing that this behaviour can in practice in the limit only take a small number of possible forms. When applied to single return series or loss series, it appears to offer a conceptually appealing approach to analysing extreme events and to calculating risk measures such as Value-at-Risk involving high severity low frequency confidence levels. It suggests that we can identify likelihoods of very extreme events 'merely' by using the following prescription:

(1) Identify the apparent type of tail behaviour being exhibited by the variable in question.

(2) Estimate the (small number of) parameters that then characterise the tail behaviour.

(3) Estimate the likelihood of occurrence however far into the distributional tail, by inserting the desired quantile or confidence level into the tail distribution estimated in step 2

EVT provides a set of limiting results that potentially enable one to analyse unusual events. It involves two broad sets of results, one applying to 'block maxima' (or 'block minima') and one applying to 'threshold exceedances'.

When EVT applies, the more traditional 'block maxima' results provide information on the distribution of the maximum value of the series in given blocks, e.g. daily losses over a 25 business day period (here the 25 business day period is the discrete 'block' of data). The distribution of the maxima converges to one of three different limiting forms that in aggregate can be represented by different parameterisations of the generalised extreme value (GEV) family of probability distribution.

The newer 'threshold exceedances' results provide an indication of the likelihood of outcomes exceeding a given threshold level. If the threshold is pushed out into the tail of the distribution then if EVT applies these likelihoods converge asymptotically to random variables from a simple family, the generalised Pareto distribution (GPD). The distributions resulting from the 'threshold exceedances' results are also termed 'peaks-over-thresholds' distributions. Our focus in most of these pages will be on the 'threshold exceedances' results as they are less wasteful of the data than the 'block maxima' results.

2. Caveats

[ExtremeValueTheory2]

In practice life is not as simple as is suggested in the Introduction. A particularly important issue is that extrapolation into the tail of a probability distribution isn't challenging because it is difficult to identify possible probability distributions that might fit the observed data. Instead the challenge is that the range of answers that can plausibly be obtained can be very wide. Extrapolation of any sort (including, as here, extrapolation into the tail of a distribution) is an intrinsically uncertain exercise, much less reliable than interpolation, as is explained in Press et al (2007).

Three other important caveats are relevant when EVT is applied to financial data:

(a) EVT relies on the tail of the distribution in question actually converging in some suitable sense. This generally occurs for smooth distributions commonly used by statisticians like the normal distribution, the Student's *t* distribution, the Pareto distribution etc. However, these sorts of distributions are very 'well behaved' in a mathematical sense and also form an infinitesimal proportion of the totality of possible distributions that might apply. So it is by no means obvious that convergence of the sort required for EVT to apply will actually take place in practice. It is relatively straightforward to construct distributions where convergence doesn't occur, although whether they are plausible is again a matter of opinion. Fundamentally, extrapolation involves exercise of judgement, and what one person thinks is reasonable someone else may think is not.

(b) EVT is usually developed from a univariate, i.e. single series, perspective. Some important financial problems, in particular portfolio construction, are intrinsically multivariate in nature. For example, most practical portfolio construction problems require selection between asset categories, so require an understanding of the joint behaviour of different return series. It is possible to develop a multivariate extreme value theory (including results for multivariate maxima and multivariate threshold exceedances), but the mathematics is quite complicated, perhaps best analysed using copulas, and is not very easily aligned to the portfolio construction problem.

(c) The conceptual appeal of EVT may encourage researchers to leap in with the technique without first trying to understand what might be causing the observed tail behaviour. An important point here is that financial data often exhibits time-varying volatility (also known as volatility clustering). If this point is given insufficient weight then EVT results, even if theoretically applicable, may be easily misinterpreted or misapplied.

3. Main Block maxima results and the Fisher-Tippett, Gnedenko theorem

[ExtremeValueTheory3]

As noted in the Introduction, 'block maxima' results are the more traditional variant of EVT but are also less useful for risk management purposes as they are less directly relevant to the task of, say, estimating VaR's at extreme threshold levels. However, it is conventional to discuss these first, so we also develop EVT in this manner.

Suppose we are interested in statistics applicable to a set of portfolio losses measured over time. We assume that these losses are random variables. These losses will be a series X_t, say. We will first assume that the losses are independent and identically distributed ('i.i.d') but later we will relax this assumption. We will also assume that the X_i are continuous random variables.

The role of the generalised extreme value (GEV) distribution in the theory of extremes
Is analogous to the role the normal distribution plays within the Central Limit Theorem (CLT). With the CLT we have to *normalise* the data for a limiting distribution to appear. Specifically, if $X_1, X_2, \ldots$ are iid with a finite variance and if we write $S_n = X_1 + \cdots + X_n$ then the CLT indicates that appropriately normalised sums, $Z_n = (S_n - a_n)/b_n$ converge in distribution to the standard normal, i.e. the $N(0,1)$, distribution as n tends to infinity. By 'normalise' we here mean a sequence of normalising constants not dependent on any particular X_i but dependent merely on n and on the parameters characterising the distribution from which they are all drawn. For the CLT the normalising constants (a_n) and (b_n) are defined by $a_n = nE(X_1)$ and $b_n = \sqrt{var(X_1)}$. In mathematical notation we have:

$$\lim_{n\to\infty} P\left(\frac{S_n - a_n}{b_n} \le x\right) = N(x) \quad x \in \mathbb{R}$$

where $N(x)$ is the cumulative distribution function of the unit normal distribution.

Block maxima results focus on suitably normalised maxima of discrete sets of X_t. So, suppose each block consists of n elements (so the j'th block involves elements numbered $n(j-1)+1$ to nj (if the first entry in the series is numbered entry 1). We calculate $M_n = max\left(X_{n(j-1)+1}, \ldots, X_{nj}\right)$ and we are interested in the distributional form of M_n (appropriately normalised) as $n \to \infty$. If the available observed data involves m such blocks, i.e. is of length mn, say, then we will have only m different (independent) values of M_n. In some loose sense only 'one' data point from each block drives M_n and any information implicit in the remainder is thrown away by focusing merely on these maxima (although of course all in some underlying sense influence M_n). Thus the approach appears likely to make relatively inefficient use of the available data when applied to real life data series, if n is large.

Suppose the cumulative distribution function of each X_i is $F(x)$, then because they are i.i.d. we will have $P(M_n(x) \le x) = F^n(x)$

The main block maxima EVT result is then as follows:

(1) Suppose that there are real sequences of numbers (d_n) and (c_n), where $c_n > 0$ for all n such that:

$$\lim_{n\to\infty} P\left(\frac{M_n - d_n}{c_n} \le x\right) = \lim_{n\to\infty} F^n(c_n x + d_n) = H(x)$$

for some non-degenerate $H(x)$ (by non-degenerate we mean that the limiting distribution is not concentrated onto a single point).

(2) F is then said to be *in the maximum domain of attraction of* H, written $F \in MDA(H)$.

(3) The Fisher-Tippett, Gnedenko Theorem states that if $F \in MDA(H)$ for some non-degenerate distribution function H then H (when appropriately standardised) must represent a *generalised extreme value* (GEV) distribution, H_ξ, for some value of ξ. Such a distribution has a distribution function:

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

where $1 + \xi x > 0$.

A (non-standardised) three-parameter family is obtained by defining $H_{\xi,\mu,\sigma} = H_\xi((x-\mu)/\sigma)$ for a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$. It is always possible to choose (d_n) and (c_n) so that the resulting distribution takes the standard form. Some commentators replace ξ by $1/\alpha$ to make the link with Pareto distributions clearer (see threshold exceedance results). If $\alpha \equiv 1/\xi$ is positive then it is known as the *tail index*, for reasons set out below.

The GEV is 'generalised' in the sense that it subsumes three types of distribution which are known by other names, i.e.:

Value of ξ	**Distributional type**	**Distributional form**
$\xi = 0$	Gumbel	$\exp\left(-\exp(-x)\right)$ for $-\infty < x < \infty$
$\xi > 0$	Fréchet	$\exp\left(-(1+\xi x)^{-1/\xi}\right)$ for $1+\xi x > 0$, otherwise 0
$\xi < 0$	Weibull	$\exp\left(-(1-\xi x)^{1/\xi}\right)$ for $1-\xi x > 0$, otherwise 1

The Weibull distribution is a short-tailed distribution with a so-called finite *right endpoint*, $x_F = sup\{x \in \mathbb{R}, x < 1\}$. The Gumbel and Fréchet distributions have infinite right end points, but the decay in the tail of the Fréchet distribution is much slower than for the Gumbel distribution.

4. Block maxima results – examples of different limiting behaviour

[ExtremeValueTheory4]

The limiting behaviours of many traditional distributions are well-known. For example, block maxima values for the normal, log-normal and exponential distributions converge to the Gumbel distribution. Those for the Student's *t*, inverse gamma, *F* and Pareto distributions converge to the Fréchet distribution.

Two examples where it is easy to derive the limiting distributions are the exponential and the Pareto distribution:

Exponential

Suppose the underlying distribution is an exponential distribution with distribution function $F(x) = 1 - \exp(-\beta x)$ for $\beta > 0$ and $x \geq 0$. Then by choosing normalising sequences $c_n = 1/\beta$ and $d_n = \ln n/\beta$ we have:

$$F^n(c_n x + d_n) = \left(1 - \frac{1}{n}\exp(-x)\right)^n, \quad x \geq -\ln n$$

$$\therefore \lim_{n\to\infty} F^n(c_n x + d_n) = \exp(-\exp(-x)), \quad x \in \mathbb{R}$$

So $F \in MDA(H_0)$.

Pareto

Suppose the underlying distribution is a Pareto distribution $Pa(\alpha, \kappa)$ with distribution function $F(x) = 1 - (\kappa/(\kappa + \alpha))^{\alpha}$ for $\alpha > 0, \kappa > 0$ and $x \geq 0$. Then by choosing the normalising sequences $c_n = \kappa n^{1/\alpha}/\alpha$ and $d_n = \kappa n^{1/\alpha} - \kappa$ we have:

$$F^n(c_n x + d_n) = \left(1 - \frac{1}{n}\left(1 + \frac{x}{\alpha}\right)^{-\alpha}\right)^n, \ \ 1 + \frac{x}{\alpha} \geq n^{1/\alpha}$$
$$\therefore \lim_{n\to\infty} F^n(c_n x + d_n) = \exp\left(1 - \left(1 + \frac{x}{\alpha}\right)^{-\alpha}\right), \ \ 1 + \frac{x}{\alpha} > 0$$

So $F \in MDA\left(H_{1/\alpha}\right)$.