EDUCATION COMMITTEE

OF THE

SOCIETY OF ACTUARIES

ADVANCED SHORT-TERM ACTUARIAL MATHEMATICS STUDY NOTE

CHAPTER 5 OF *QUANTITATIVE ENTERPRISE RISK MANAGEMENT*

by

Mary R. Hardy
and
David Saunders

ASTAM-22-23 **Printed in the USA**

5
Extreme Value Theory

5.1 Summary

In this chapter, we present some key results from **extreme value theory** (EVT) and illustrate how EVT can be used to supplement traditional statistical analysis. We use EVT when we are concerned about the impact of very rare, very large losses. Because they are rare, we are unlikely to have much data, but using EVT we can infer the extreme tail behaviour of most distributions.

There are two different, but related, types of models for extreme value analysis. The first considers the distribution of the maximum value in a random sample of losses. These are called the **block maxima** models. The second comes from analysing the rare, very large losses, defined as the losses exceeding some high threshold. These are the **points over threshold** models. We present the key results for both of these, and show how they are connected. We derive formulas for the Value at Risk (VaR) and Expected Shortfall risk measures using EVT that are useful when the loss distribution is fat-tailed, and the risk measure parameter α is close to 1.0. We use examples throughout to highlight the potential uses in practical applications.

We present several key theorems in this chapter without proofs, as the mathematics required is beyond the scope of this text.

5.2 Introduction

In Chapter 4, we looked at distributions for loss frequency and severity, appropriate for a wide range of quantifiable risks. We saw that, given sufficient data, we can model the loss frequency and severity separately, and, assuming independence of frequency and severity, we can construct a distribution function for aggregate losses. Generally, the estimation part of the frequency and severity analysis would use conventional statistical methods, such as maximum likelihood. These methods give a good overall fit to a distribution from a reasonably sized sample of data. However, the main weight in the fitting process will, implicitly, focus on the centre of the distribution. This may create a fit that is less satisfactory in the extreme right tail of the loss distribution, where we are concerned with the very rare, but potentially disastrous extreme losses. Even if a distribution appears to fit the data satisfactorily overall, it may not adequately model the part of the distribution in the extreme tails, beyond the range of the available data. In these cases, we can use EVT to supplement the traditional analysis.

Some examples of risk management cases which are suited to EVT include the following:

- An insurer might model the claims exceeding some extreme threshold to assess the mitigation benefits of a reinsurance strategy.
- An investment bank might model the potential for extreme stock price movements as part of its risk management operations.
- A company with exposure to currency risk might use an extreme value approach to guide its purchase of currency derivatives.
- Ocean engineers model extreme weather conditions to design ocean structures that can withstand, for example, a 1-in-500-year storm event.

5.3 Distributions of Block Maxima

5.3.1 Block Maxima and the Maximum Domain of Attraction

Suppose we have an i.i.d. sample of n values, $X_1, X_2, \ldots, X_n$, with common distribution function $F(x)$. Let M_n denote the maximum of the sample, that is, $M_n = \max\{X_1, X_2, \ldots\}$. The distribution function for M_n is

$$\begin{aligned} F_n(m) &= \Pr[M_n \le m] \\ &= \Pr[X_1 \le m] \Pr[X_2 \le m] \Pr[X_3 \le m] \cdots \Pr[X_n \le m] \\ &= (F(m))^n. \end{aligned}$$

If we consider the limit of this distribution as the block size increases, there are only two possibilities:

$$\begin{aligned} \lim_{n\to\infty} F_n(m) &= 0 \qquad \text{if and only if } F(m) < 1, \\ \lim_{n\to\infty} F_n(m) &= 1 \qquad \text{if and only if } F(m) = 1. \end{aligned}$$

This is not particularly helpful. However, the first important result of EVT tells us that if we *normalize* the block maximum – that is, if we consider the random variable

$$M_n^* = \frac{M_n - d_n}{c_n},$$

where $c_n > 0$ and d_n are deterministic functions of n (involving the parameters of underlying distribution) – then in many cases we can find a limiting distribution for M_∞^*. That is, for some M_n we can find a distribution function $H(x)$, where $H(x)$ is not degenerate – meaning that the random variable M_∞^* is not a constant – and where

$$\lim_{n\to\infty} \Pr[M_n^* \le x] = \lim_{n\to\infty} F_n(c_n x + d_n) = H(x).$$

In this case, we say that the distribution $F(x)$ is in the **maximum domain of attraction** (MDA) of H.

5.3.2 The Generalized Extreme Value Distribution

There are three important distributions for limits of normalized block maxima: the Gumbel, Fréchet, and Weibull distributions.

The Gumbel Distribution

$$F(x) = \exp\left\{-\exp\left(-\frac{x-\mu}{\theta}\right)\right\}, \quad \theta > 0. \tag{5.1}$$

The Fréchet Distribution

$$F(x) = \exp\left\{-\left(\frac{x-\mu}{\theta}\right)^{-\alpha}\right\}, \quad x > \mu;\ \alpha > 0;\ \theta > 0. \tag{5.2}$$

The Weibull EV Distribution

$$F(x) = \exp\left\{-\left(\frac{\mu-x}{\theta}\right)^{\tau}\right\}, \quad x < \mu;\ \tau > 0;\ \theta > 0. \tag{5.3}$$

Note that this version of the Weibull is different from the specification in Chapter 4, but it is related by a sign change.

We can express these three distributions as variants of a single distribution, which is called the **generalized extreme value** (GEV) **distribution** with cumulative distribution function $H_\xi(x)$, where ξ is the **shape parameter**:

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-\frac{1}{\xi}}\right) & \xi \ne 0,\ \xi x > -1, \\ \exp\left(-e^{-x}\right) & \xi = 0. \end{cases} \tag{5.4}$$

- If $\xi = 0$ this gives the Gumbel distribution, with $\mu = 0$, $\theta = 1$.
- If $\xi > 0$ this gives the Fréchet distribution, with $\mu = -1/\xi$, $\theta = 1/\xi$, $\alpha = 1/\xi$.
- If $\xi < 0$, this gives the Weibull EV distribution, with $\mu = -1/\xi$, $\theta = -1/\xi$, $\tau = -1/\xi$.

Note that as $\xi \to 0^+$, or $\xi \to 0^-$, we find that the Gumbel distribution is the limiting case of both the Fréchet and Weibull EV distributions.

The GEV can be adjusted for scale and location, to give $H_{\xi,\mu,\theta}$ where

$$H_{\xi,\mu,\theta}(x) = \begin{cases} \exp\left(-(1+\xi(x-\mu)/\theta)^{-\frac{1}{\xi}}\right) & \xi \neq 0,\ (1+\xi\,(x-\mu)/\theta) > 0, \\ \exp\left(-e^{-(x-\mu)/\theta}\right) & \xi = 0, \end{cases} \tag{5.5}$$

where μ is a location parameter, and θ is a scale parameter.

The importance of the GEV distribution is apparent from the following theorem, which says that if a distribution has a non-degenerate limiting distribution for M_n^* (and most of the distributions that we use fall into this category), then the limiting distribution must be the GEV distribution (Figure 5.1).

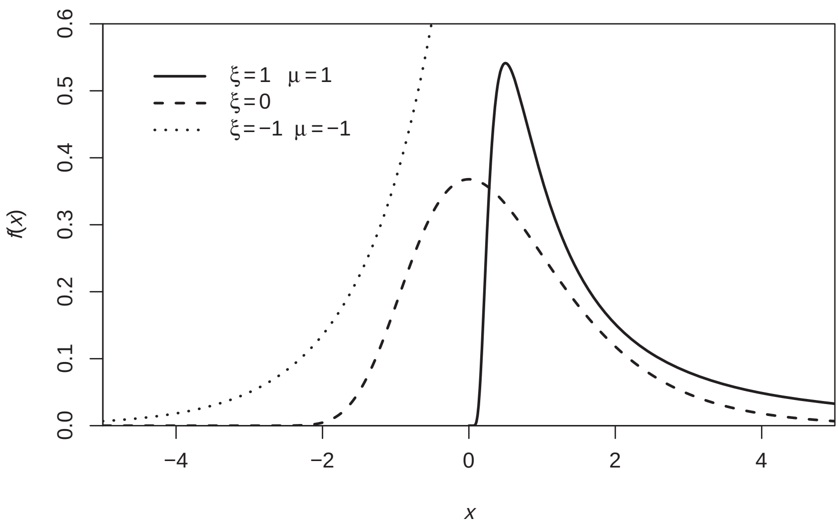


Figure 5.1 Generalized extreme value probability density functions; $\theta = 1$.

Theorem 5.1 ***The Fisher–Tippett–Gnedenko Theorem***
If a distribution F lies in the MDA of a non-degenerate distribution H, then H must be the GEV distribution, H_ξ.

This classic result of EVT has some analogy to the **central limit theorem**. Consider $S_n = \sum_{j=1}^{n} X_j$ for i.i.d. X_j, with common mean $\mu > 0$ and common variance $\sigma^2 > 0$. There is, technically, no limiting distribution for S_n as n tends to infinity, as both the mean and variance of S_n will also tend to infinity. However, if we normalize the random variables, and consider

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

then the central limit theorem tells us that $\lim_{n\to\infty} Z_n \sim N(0,1)$. That means that a single distribution – the standard normal distribution – is the limiting distribution of the normalized sum of any random sample of i.i.d. variables, regardless of the distribution of the individual variables, provided that they have a finite variance. In EVT, instead of considering the mean, we start by considering the maximum value in a sample, and it turns out that, like the mean, there is a limiting distribution, the GEV, that applies to the normalized sample maximum (in many cases) regardless of the distribution of the individual variables.

Example 5.1 For the exponential distribution with mean 1, you are given that $c_n = 1$ and $d_n = \log n$. Show that this distribution lies in the Gumbel MDA.

Solution 5.1 We have $F(x) = 1 - e^{-x}$. Then

$$\begin{aligned}\Pr\left[\frac{M_n - d_n}{c_n} \le x\right] &= \Pr[M_n \le c_n\, x + d_n] \\ &= (F(c_n\, x + d_n))^n = (1 - \exp(-(c_n x + d_n)))^n .\end{aligned}$$

Now let $c_n = 1$ and $d_n = \log n$, so that

$$\begin{aligned}\Pr\left[\frac{M_n - d_n}{c_n} \le x\right] &= \left(1 - e^{-x-\log n}\right)^n \\ &= \left(1 - \frac{e^{-x}}{n}\right)^n .\end{aligned}$$

On the right-hand side we have a term of the form $(1 + k/n)^n$ (where, in this case, $k = -e^{-x}$), and the limit of this expression as $n \to \infty$, is e^k. Hence

$$\lim_{n\to\infty}\left(1 - \frac{e^{-x}}{n}\right)^n = e^{-e^{-x}},$$

which shows that the exponential distribution is in the MDA of the H_0, or Gumbel distribution.

Example 5.2 For the exponential distribution with mean 1, calculate the probability that $M_n \le 5$ for $n = 10$ and for $n = 100$, using the exponential distribution function directly, and using the limiting extreme value distribution.

Solution 5.2 We have $F(5) = 0.993262$, so

$$\Pr[M_{10} \le 5] = (F(5))^{10} = 0.93463,$$
$$\Pr[M_{100} \le 5] = (F(5))^{100} = 0.50861.$$

Using the Gumbel distribution, we have $c_n = 1$, $d_n = \log n$, so

$$\Pr[M_n \le 5] = \Pr\left[\frac{M_n - d_n}{c_n} \le \frac{5 - d_n}{c_n}\right] = \Pr[M_n - \log n \le 5 - \log n],$$

and since the limiting distribution is Gumbel, for large n this probability is approximately

$$H_0(5 - \log n) = e^{-e^{-(5-\log n)}} = e^{-ne^{-5}},$$

which gives

$$\Pr[M_{10} \le 5] \approx 0.93484, \qquad \Pr[M_{100} \le 5] \approx 0.50977. \qquad \square$$

This does not appear to be all that useful at this stage, as we need to know the underlying distribution to know the normalizing functions, and if we know the underlying distribution, we can calculate the required probabilities directly. However, we shall see in Section 5.4 that we can derive some very useful information about tail risk without knowing the full distribution of the underlying random variable. The key parameter for risk management purposes will be the ξ parameter in $H_\xi(x)$.

An obvious question arising from the Fisher–Tippett–Gnedenko theorem is how we find the normalizing functions. One approach is to use the following result, which we state without proof (see Embrechts et al. (2013) for more details).

Theorem 5.2 *Consider a loss X with distribution function $F(x)$ and survival function $S(x) = 1 - F(x)$. Then $F(x)$ is in the MDA of H_ξ if and only if*

$$\lim_{n\to\infty} nS(c_n x + d_n) = -\log H_\xi(x) = \begin{cases} (1+\xi x)^{-\frac{1}{\xi}} & \text{for } \xi \neq 0, \\ e^{-x} & \text{for } \xi = 0. \end{cases} \qquad (5.6)$$

$\square$

In some cases, we can use this result to determine c_n and d_n. One example is the Pareto distribution, as we demonstrate next.

Example 5.3 Determine the limiting distribution for a maximum of Pareto(α, λ) random values, and find the normalizing sequences c_n and d_n.

Solution 5.3 The Pareto(α, λ) distribution function is

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\alpha},$$

so

$$\begin{aligned} nS(c_n x + d_n) &= n\left(\frac{\lambda}{\lambda + d_n + c_n x}\right)^{\alpha} \\ &= n\left(1 + \frac{d_n}{\lambda} + \frac{c_n}{\lambda}x\right)^{-\alpha} && (5.7) \\ &= \left(n^{-1/\alpha} + \frac{n^{-1/\alpha} d_n}{\lambda} + \frac{n^{-1/\alpha} c_n}{\lambda}x\right)^{-\alpha}. && (5.8) \end{aligned}$$

The form of $nS(c_n x + d_n)$ in equation (5.8) is similar to the Fréchet form of $-\log H(x)$ in equation (5.6), where $\alpha > 0$, indicating that $\xi = \frac{1}{\alpha} > 0$.

To match the Fréchet form of $-\log H_\xi(x)$, we need (at least asymptotically)

$$n^{-1/\alpha} + \frac{n^{-1/\alpha} d_n}{\lambda} = 1, \quad \text{and} \quad \frac{n^{-1/\alpha} c_n}{\lambda} = \xi = \frac{1}{\alpha},$$

$$\Rightarrow d_n = (n^{1/\alpha} - 1)\lambda, \quad \text{and} \quad c_n = \frac{\lambda n^{1/\alpha}}{\alpha}.$$

In summary, the Pareto distribution is in the MDA of the Fréchet distribution with shape parameter $\xi = 1/\alpha$ and normalizing functions $c_n = \frac{\lambda n^{1/\alpha}}{\alpha}$ and $d_n = (n^{1/\alpha} - 1)\lambda$.

□

5.3.3 Notes on the Generalized Extreme Value (GEV) Distributions

The Fréchet Distribution

- The Fréchet distribution is the GEV distribution, with $\xi > 0$.
- The Fréchet distribution is bounded below, with $x > \mu - \frac{\theta}{\xi}$, for μ and θ as in equation (5.5).
- The Fréchet distribution is fat-tailed and is relatively popular for use in managing extreme risks in finance and insurance.

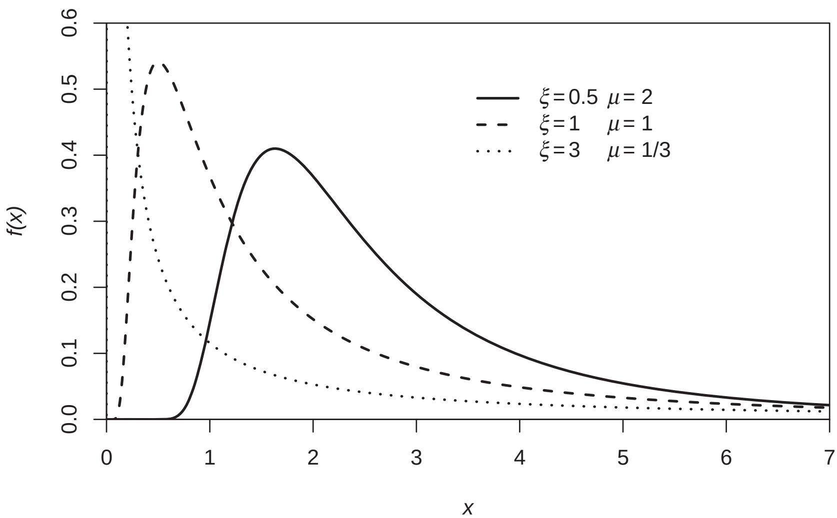


Figure 5.2 Fréchet probability density functions; $\theta = 1$.

- The distribution function is often expressed in terms of $\alpha = \frac{1}{\xi}$. This is called the **tail index** of the distribution.
- Larger values of ξ (and hence smaller values of α) indicate a fatter-tailed distribution.
- The Fréchet distribution is in the Fréchet MDA.
- For any distribution in the MDA of the Fréchet distribution, there are only a finite number of moments, up to $\alpha = 1/\xi$. That is, for a positive integer k, if $k < \frac{1}{\xi}$ we have $\mathrm{E}[X^k] < \infty$, but for $k \geq \frac{1}{\xi}$, we have $\mathrm{E}[X^k] = \infty$. This means that a distribution with an infinite number of moments (even if it is fat-tailed) cannot be in the MDA of the Fréchet distribution.
- Distributions in the Frechet MDA include the Pareto, Student's t, and Burr distributions.
- In Figure 5.2, we show density functions for $\xi = 0.5, 1,$ and 3. The scale parameter is $\theta = 1$ and the location parameter is set at $1/\xi$ to give a distribution bounded below at 0 in each case.
- Another characterization of the Frechet MDA uses the following definition:

Definition 5.3 A positive function L is **slowly varying at** ∞ if for any $t > 0$,

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1.$$

Examples of slowly varying functions include $\log(x)$ and $(k + x^{-1})^{-\alpha}$.

A distribution function F is in the MDA of the Fréchet distribution with parameter ξ if and only if

$$S(x) = x^{-1/\xi} L(x), \tag{5.9}$$

where $L(x)$ is a function which is slowly varying at ∞.

The Gumbel Distribution

- The Gumbel distribution is the GEV distribution with $\xi = 0$.
- The distribution is unbounded.
- Many distributions are in the MDA of the Gumbel distribution, ranging from quite thin-tailed distributions, such as the normal and exponential, to quite fat-tailed, such as the gamma, lognormal, and inverse Gaussian. Note that even though the Gumbel distribution is not bounded, the distributions in its MDA may be bounded (e.g., the lognormal random variable is strictly positive).
- Distributions in the Gumbel MDA have infinite number of moments – that is, $E[X^k] < \infty$ for any $k = 1, 2, 3, \ldots$
- The Gumbel distribution is in the MDA of the Gumbel distribution.
- All Gumbel distributions have the same shape, as the shape parameter, ξ, is fixed. The location and scale may vary.
- The Gumbel density function is illustrated in Figure 5.3.

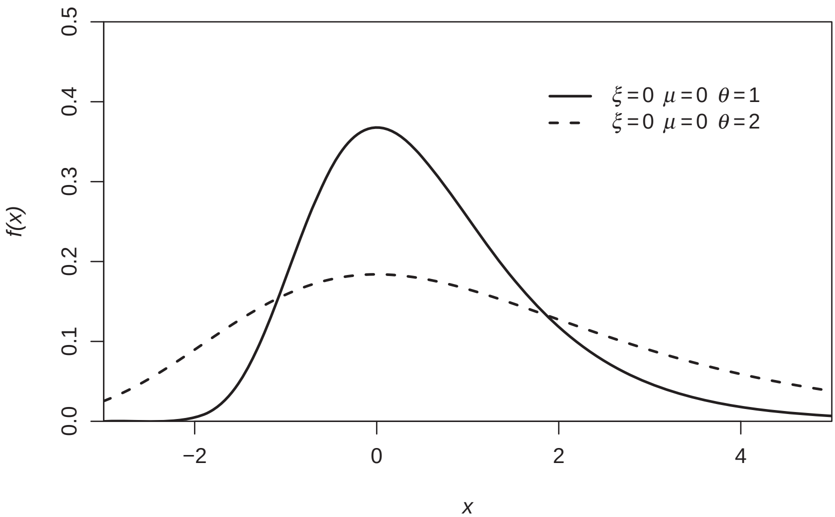


Figure 5.3 Gumbel probability density functions

The Weibull EV Distribution

- The Weibull EV distribution is the GEV distribution with $\xi < 0$.
- The Weibull EV distribution is bounded above at $\mu - 1/\xi$. It is, therefore, less useful for most large loss modelling, but can be useful for maxima of bounded losses.
- The Weibull EV distribution is related to the Weibull distribution by a sign change – that is, if Y is a Weibull EV-distributed random variable, with $\mu = 0, \theta = 1$, then $1 + \xi Y$ is a Weibull distributed random variable.
- The Weibull EV distribution, beta distribution, and uniform distribution are in the MDA of the Weibull EV distribution.
- In Figure 5.4, we show density functions for $\xi = -1, -0.5$, and -0.25. The scale parameter is $\theta = 1$, and the location parameter is set at $1/\xi$ to give a distribution bounded above at 0 in each case.

5.3.4 Estimating the GEV Parameters

For a distribution lying in the MDA of the GEV distribution, the Fisher–Tippett–Gnedenko theorem tells us that $(M_n - d_n)/c_n$ approximately follows the GEV distribution, with some parameters μ', θ', and ξ, say. In this case, the non-normalized random variable M_n also approximately follows the GEV

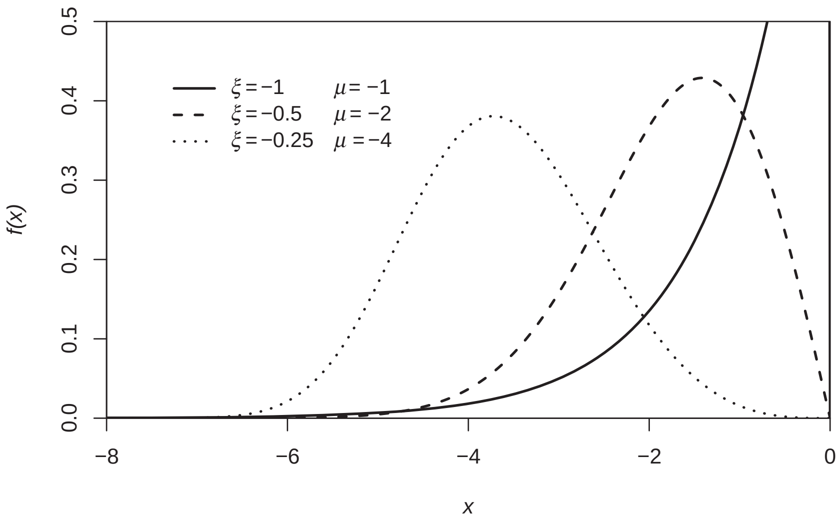


Figure 5.4 Weibull EV probability density functions; $\theta = 1$.

distribution, with adjusted location and scale parameters, μ and θ, but with the same shape parameter ξ. We can, therefore, estimate ξ, which is the key parameter for classifying the extreme value distribution, by fitting the GEV to block maxima of our data, using maximum likelihood estimation.

Assume we have a set of data, which we divide into k blocks, each with n values. This gives us k sample points of the n-block maxima. Let m_j denote the maximum of the jth block, for $j = 1, 2, \ldots, k$. The division of the data into blocks is natural for data arising as a time series, such as stock market data or flood levels. If the data has no natural ordering, then it may be randomly partitioned into blocks.

By differentiating the GEV distribution function $H_\xi(x)$, we find the GEV density function $h_\xi(x)$,

$$h_\xi(x) = (1+\xi x)^{-(1+\frac{1}{\xi})}\, e^{-(1+\xi x)^{-\frac{1}{\xi}}}, \tag{5.10}$$

or, for the more general distribution, including scale and location parameters,

$$h_{\xi,\,\mu,\,\theta}(y) = \frac{1}{\theta}\left(1+\xi\left(\frac{y-\mu}{\theta}\right)\right)^{-(1+\frac{1}{\xi})} e^{-(1+\xi\left(\frac{y-\mu}{\theta}\right))^{-\frac{1}{\xi}}}. \tag{5.11}$$

Assuming independence of the block maxima, then for the maximum likelihood estimators, we find the parameters which maximize the sum of the log densities of the sample. That is, we maximize $l(\xi, \mu, \theta)$ where

$$l(\xi, \mu, \theta) = \sum_{j=1}^{k} \log(h_{\xi,\,\mu,\,\theta}(m_j)) \tag{5.12}$$

$$= -k\log(\theta) - \left(1+\frac{1}{\xi}\right)\sum_{j=1}^{k}\log\left(1+\xi\left(\frac{m_j-\mu}{\theta}\right)\right)$$

$$- \sum_{j=1}^{k}\left(1+\xi\left(\frac{m_j-\mu}{\theta}\right)\right)^{-\frac{1}{\xi}}. \tag{5.13}$$

In practice, we must balance conflicting data requirements. With a fixed sample size, we must decide how large the blocks will be. If they are very large, we can be more confident that the maxima lie in the tail of the distribution, suitable for applying the GEV distribution function, but the number of data points may be small. On the other hand, if we use smaller blocks to get a larger sample of block maxima, we may not be sufficiently near to the tail of the distribution to find accurate parameter estimates.

The result may be that even with relatively large data sets, the standard error of the parameter estimates is large. This is particularly true for the ξ parameter,

Table 5.1. *Annual maxima of monthly losses on the S&P/TSX composite index, 1956–2012, per $100 invested, in chronological order (by rows)*

5.88	10.25	0.85	5.06	4.68	2.37	8.52	3.35
0.66	6.68	7.47	5.72	4.25	11.03	10.09	6.47
3.57	11.55	10.52	5.74	7.05	5.81	5.31	10.18
19.41	14.04	10.14	5.53	3.80	6.53	4.83	25.52
2.70	6.75	8.40	3.37	4.44	3.31	6.89	1.97
3.63	4.89	22.45	6.29	8.77	14.23	7.76	3.02
3.97	5.82	3.63	6.42	18.23	6.52	3.78	9.06

which is problematic, since it is the parameter that indicates which form of the GEV distribution is appropriate.

Solving for the maximum likelihood estimates can be done very conveniently with suitable software packages.

Example 5.4 Table 5.1 shows the block maxima for monthly losses (in %) on the S&P/TSX composite index, over a period of 56 years, from 1956–2012. The block size is 12 months. Assume the values are independent and identically distributed:

(a) Find the maximum likelihood estimates of the GEV distribution parameters for this data.
(b) Find the MLE for the GEV distribution parameters using a block size of 24 months.

Solution 5.4 We used the `fit.GEV` function from the R package *QRM* (Pfaff and McNeil, 2020). This generates parameter estimates and standard errors.

(a) We find, using the annual data, that the MLE parameter estimates are as follows (standard errors in parentheses):

$$\mu = 5.009(0.450), \quad \theta = 3.012(0.346), \quad \text{and } \xi = 0.1575(0.098).$$

This indicates that the Fréchet distribution is the most appropriate, although the Gumbel distribution is also possible.

(b) Using 24-month blocks, we find the following estimates and standard errors:

$$\mu = 7.144(0.60), \quad \theta = 2.655(0.53), \quad \text{and } \xi = 0.389(0.213),$$

which, again, points to the Fréchet distribution, possibly with a heavier tail, but does not rule out the Gumbel distribution. Note that we expect the μ and θ parameters to change, as they are functions of the block size. For large n though, the value of ξ should be stable for different block sizes.

As we expect, the smaller sample size of block maxima (using 24-month blocks) leads to larger standard errors. □

An interest in the tail behaviour of a random variable does not always mean an interest in block maxima; the block maximum may not be very extreme in some cases, and in others, a block may have several values that would be considered extreme. In the example above, using 24-month blocks, we lose the fifth-largest value from the 12-month blocks (14.04), because it is adjacent to an even larger maximum. In the next section, we select extreme values differently, by considering all the values falling beyond some threshold.

5.4 Distribution of Excess Losses Over a Threshold

In Example 5.4, we saw that much of the data is discarded and some non-extreme data points are incorporated into the estimation. This seems, intuitively, to be an inefficient way to model tail behaviour. We now expand the analysis, to consider the distribution of losses which are in the tail of the underlying distribution, where we define the tail by setting a threshold level corresponding to a very large loss. We then consider the distribution of exceedances, or excess losses, which are the differences between the loss values and the threshold. That is, for an underlying loss random variable X, and threshold $d > 0$, the **excess loss** is $Y_d = X - d|X > d$. This method is often called the **peaks** or **points over threshold** (POT) approach.

From this definition, given that X is continuous, and that $\Pr[X > d] > 0$, the excess loss random variable Y_d is also continuous, and $Y_d \geq 0$. We can derive the distribution and density function for Y_d in terms of the functions for X, as follows.

5.4.1 The Excess Loss Random Variable

The distribution function of the excess loss $Y_d = X - d|X > d$, is denoted $F_d(y)$, and the survival function is $S_d(y) = 1 - F_d(y)$. These are related to the underlying distribution function and survival functions, F_X and S_X, as

$$
\begin{aligned}
F_d(y) = \Pr[Y_d \leq y] &= \Pr\left[X - d \leq y |\, X > d\right] \\
&= \frac{F_X(y+d) - F_X(d)}{1 - F_X(d)} \\
&= \frac{S_X(d) - S_X(y+d)}{S_X(d)} \\
&= 1 - \frac{S_X(y+d)}{S_X(d)} \\
\Rightarrow S_d(y) &= \frac{S_X(y+d)}{S_X(d)}.
\end{aligned}
$$

Similarly, the probability density function for Y_d is

$$
f_d(y) = \frac{f_X(y+d)}{S_X(d)}.
$$

The expected value of Y, as a function of the excess threshold d, is called the **mean excess loss** (MEL) of X and is denoted $e(d)$; that is,

$$
e(d) = \mathrm{E}\left[X - d | X > d\right]. \tag{5.14}
$$

For a continuous random variable, X, we have

$$
\begin{aligned}
e(d) = \mathrm{E}\left[X - d | X > d\right] &= \frac{1}{S_X(d)} \left(\int_d^{\infty} (y-d) f_X(y)\, dy \right) \qquad (5.15) \\
&= \frac{1}{S_X(d)} \left(\int_d^{\infty} y\, f_X(y)\, dy - d \int_d^{\infty} f_X(y) dy \right) \\
&= \frac{1}{S_X(d)} \left(\int_d^{\infty} y\, f_X(y)\, dy - d\, S_X(x) \right) \\
&= \frac{\mathrm{E}[X] - \mathrm{E}[X \wedge d]}{S_X(d)}. \qquad (5.16)
\end{aligned}
$$

Similarly to the GEV distribution, which provided the asymptotic distribution for block maxima, the general asymptotic distribution for excess losses is the **generalized Pareto distribution** (GPD). This distribution is closely related to the GEV distribution.

5.4.2 The Generalized Pareto Distribution (GPD)

The distribution function of the **generalized Pareto distribution** (GPD) is, for $\beta > 0$,

$$
G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi\, x/\beta)^{-1/\xi} & \xi \neq 0, \\ 1 - e^{-x/\beta} & \xi = 0, \end{cases} \tag{5.17}
$$

where $x \geq 0$ for $\xi \geq 0$, and $0 \leq x \leq -\beta/\xi$ for $\xi < 0$. In the GPD distribution, β is a scale parameter and ξ is the shape parameter. There is no location parameter as the distribution is fixed, with a lower bound at zero.

Notes

1. When $\xi > 0$ the GPD is a regular Pareto distribution, as in Section 4.4.4, with parameters $a = 1/\xi, \theta = \beta/\xi$.
2. For $0 < \xi < 1$, the mean of $X \sim \text{GPD}_{\xi,\beta}$ is
$$\mathrm{E}[X] = \frac{\beta}{1-\xi}.$$
For $\xi \geq 1$ the mean does not exist.
3. When $\xi > 0$, and for integer k, the kth moment of the GPD exists only for $k < 1/\xi$.
4. When $\xi = 0$ the GPD is the exponential distribution with mean β.
5. When $\xi < 0$, the distribution is a generalized beta distribution, which is left and right bounded. The left bound is zero, the right bound is $-\beta/\xi$.
6. $G_{\xi,\beta}(y) = 1 + \log H_\xi(y/\beta)$ where H_ξ is the GEV distribution function, from equation (5.4).

Example 5.5 Show that, if $Y \sim \text{GPD}_{\xi,\beta}$, for $\xi \geq 0$, $\beta > 0$, and $Z = Y - k|Y > k$ then
$$Z \sim \text{GPD}_{\xi,\beta^*} \text{ where } \beta^* = \beta + \xi\, k,$$
and hence derive the MEL function for the $\text{GPD}_{\xi,\beta}$ distribution.

Solution 5.5 Consider the survival function of Z. We assume first that $\xi > 0$, and show that the survival function for Z is GPD with ξ unchanged, and a new β^*.

$$\begin{aligned}
\Pr[Z > t] = \Pr[Y - k > t|Y > k] &= \Pr[Y > t + k|Y > k] \\
&= \frac{1 - G_{\xi,\beta}(t+k)}{1 - G_{\xi,\beta}(k)} \\
&= \frac{\left(1 + \frac{\xi}{\beta}(t+k)\right)^{-\frac{1}{\xi}}}{\left(1 + \frac{\xi}{\beta}(k)\right)^{-\frac{1}{\xi}}} \\
&= \left(1 + \frac{\xi\, t}{\beta + \xi k}\right)^{-\frac{1}{\xi}} \\
&= 1 - G_{\xi,\beta^*}(t),
\end{aligned}$$

where $\beta^* = \beta + \xi\, k$ as required.

If $\xi = 0$, then Y has an exponential distribution, with mean β and the survival function for Z is

$$\Pr[Z > t] = \Pr[Y > t + k | Y > k] = \frac{e^{-\frac{t+k}{\beta}}}{e^{-\frac{k}{\beta}}} = e^{-\frac{t}{\beta}},$$

which shows that the distribution of Z is also exponential with mean β (this is known as the **memoryless property** of the exponential distribution).

The mean excess loss function for the $\text{GPD}_{\xi,\beta}$ distribution is the expected value of the random variable Z (i.e. $\beta^*/(1-\xi)$), expressed as a function of the excess point k. That is,

$$e(k) = \begin{cases} \dfrac{\beta + \xi k}{1 - \xi} & \text{for } 0 < \xi < 1, \\ \beta & \text{for } \xi = 0. \end{cases} \tag{5.18}$$

For $\xi \geq 1$, $e(k) = \infty$ since in this case the GPD distribution does not have a finite first moment. $\square$

The important point to note from this example is that for the GPD distribution, the MEL function is a straight line. If $\xi \in (0,1)$, it has slope $\xi/(1-\xi)$, and if $\xi = 0$, the MEL function is flat.

The relationship between the GPD and GEV distributions is captured in the following theorem.

Theorem 5.4 ***The Pickands–Balkema–De Haan Theorem***
Let F denote the distribution function of a random variable X which is bounded above at $x^{sup} \leq \infty$. Then

$$F \in MDA\left(H_\xi\right) \iff \lim_{d \to x^{sup}} \sup_{0 \leq x \leq x^{sup}} \left| F_d(x) - G_{\xi,\beta}(x) \right| = 0, \tag{5.19}$$

for some function β.

What this theorem tells us is that every distribution in the MDA of H_ξ will have a distribution for excess losses that converges to the $G_{\xi,\beta}$ distribution, as the threshold tends to the maximum loss. In practice, this means that for a sufficiently high threshold d, the excess loss random variable Y_d will (approximately) follow the GPD with the same shape parameter ξ as the GEV, and with a scale parameter β. Since the left tail of Y_d is fixed at 0, there is no need for a location parameter.

In order to use the theorem, we need to identify the threshold beyond which we can assume that excess losses are close to being GPD.

One approach is to examine the empirical mean excess loss function for the data, as a function of the threshold d. Given an ordered sample,

$x_{(1)}, x_{(2)}, \ldots x_{(n)}$, we define the empirical mean excess loss function $\hat{e}(x_{(j)})$, for $j = 1, 2, \ldots, n-1$, as

$$\hat{e}(x_{(j)}) = \sum_{k=j+1}^{n} \frac{x_{(k)}}{n-j} - x_{(j)}.$$

That is, for each ordered sample value $x_{(j)}$, we take the mean of the observations which are larger than $x_{(j)}$, and subtract $x_{(j)}$ to get an estimator of the MEL evaluated at $x_{(j)}$.

We know that the GPD mean excess loss function is a straight line, even if the mean excess loss function of the underlying distribution is not. Hence, the region where the empirical mean excess loss function becomes approximately linear indicates where the GPD becomes a good approximation to the distribution of excess losses.

Once we have chosen a suitable threshold, d^*, say, we consider the reduced sample of values of $X_{(j)} - d^*$ for all $X_{(j)} > d^*$, and fit the GPD to the reduced sample.

If the mean excess loss function has positive gradient, then we fit the tail data to the Pareto distribution. If the mean excess loss function appears flat, then we fit the tail data to the exponential distribution.

In practice, identifying an appropriate threshold can be challenging. As we get up to the largest data points, the empirical MEL tends to be very volatile. See Section 5.5 for an illustration.

5.4.3 Risk Measures for the GPD

Suppose we have a loss distribution function, F_X, and that for some threshold d, the excess loss random variable $X - d | X > d$ may be assumed to follow the GPD with parameters ξ and β.

We can estimate the VaR, assuming it lies above d, using the GPD for the excess distribution above d.

For example, for $\xi > 0$, consider first the survival function for some $x > d$. Let $S_X(x)$ denote the survival function of the original distribution, and let $S_d(y)$ denote the GPD survival function for the excess loss distribution.

Then

$$\begin{aligned}
\Pr[X > x] &= \Pr[X > d]\ \Pr[X > x | X > d] \\
&= \Pr[X > d]\ \Pr[X - d > x - d | X > d] \\
&= S_X(d)\ S_d(x-d) \qquad (5.20) \\
&= S_X(d) \left(1 + \frac{\xi}{\beta}(x-d)\right)^{-1/\xi}. \qquad (5.21)
\end{aligned}$$

So we can treat the survival function as a combination of the underlying survival function up to the threshold d, and the GPD survival function beyond d.

If we replace x with the α-VaR, $Q_\alpha(X)$, assuming $Q_\alpha(X) > d$, we have

$$\begin{aligned}
\Pr[X > Q_\alpha(X)] &= 1-\alpha = \Pr[X > d]\ \Pr[X > Q_\alpha(X)|X > d] \\
\Rightarrow 1-\alpha &= S_X(d)\left(1+\frac{\xi(Q_\alpha - d)}{\beta}\right)^{-\frac{1}{\xi}} \\
\Rightarrow Q_\alpha &= d + \frac{\beta}{\xi}\left(\left(\frac{S_X(d)}{1-\alpha}\right)^{\xi} - 1\right). \qquad (5.22)
\end{aligned}$$

Equation (5.22) shows that the α-VaR can be calculated using the GPD parameters β and ξ, together with $S_X(d)$, which is the survival probability at the threshold d from the original distribution. In practice, we may not know the distribution for the underlying X, which means that we must approximate $S_X(d)$. The usual (and intuitive) non-parametric estimator is the proportion of the sample which is greater than d. For example, suppose the threshold d is selected at the 950th smallest sample value from a sample size of 1,000; that is, $d = x_{(950)}$. Then there are 50 values greater than d which form the sample for estimating the GPD parameters. The empirical probability that $X > d$ is $\hat{S}(d) = 50/1{,}000$. So, in general, assuming that d is selected such that j sample values exceed d, out of a sample of n values,

$$S_X(d) \approx \frac{\text{Number of values of } x_i > d}{\text{Total number of values of } x_i} = \frac{j}{n}. \qquad (5.23)$$

We can also use the GPD to evaluate extreme Expected Shortfall risk measures. We assume, for convenience, that losses are continuous (at least in the tail), in which case the α-Expected Shortfall of the loss is related to the α-VaR (Q_α), and the mean excess loss (MEL) function, $e_X(d)$, as follows:

$$\begin{aligned}
\mathrm{ES}_\alpha &= E[X|X > Q_\alpha] \\
&= Q_\alpha + E[X - Q_\alpha|X > Q_\alpha] \\
&= Q_\alpha + e_X(Q_\alpha). \qquad (5.24)
\end{aligned}$$

Suppose that for a continuous loss random variable X, and for a given threshold d, the distribution of $Y_d = X - d|X > d$ is GPD with parameters ξ, β. Suppose also that we are interested in the α-Expected Shortfall of X, where $Q_\alpha > d$. If $\xi \geq 1$, the Expected Shortfall does not exist (moments higher than the $(1/\xi)$th are infinite for the GPD).

If $\xi < 1$, then we consider

$$X - Q_\alpha|X > Q_\alpha = Y_d - (Q_\alpha - d)|Y_d > (Q_\alpha - d).$$

Let $k = Q_\alpha - d$, then from Example 5.5, we know that $Z = Y_d - k|Y_d > k$ follows a GPD with parameters ξ and β^* where

$$\beta^* = \beta + \xi\, k = \beta + \xi\,(Q_\alpha - d).$$

We will use E$[Z]$ for the Expected Shortfall, as

$$e_X(Q_\alpha) = \mathrm{E}[Z] = \frac{\beta + \xi(Q_\alpha - d)}{1 - \xi}$$

so

$$\begin{aligned} \mathrm{ES}_\alpha &= Q_\alpha + e_X(Q_\alpha) \\ &= Q_\alpha + \frac{\beta + \xi(Q_\alpha - d)}{1 - \xi} \\ \Rightarrow \mathrm{ES}_\alpha &= \frac{1}{1-\xi}\big(Q_\alpha + \beta - \xi\, d\big). \end{aligned} \tag{5.25}$$

Note that we have assumed that the α parameter for the risk measure is sufficiently far into the tail of the distribution that the GPD distribution is appropriate for the random variable $X - Q_\alpha | X > Q_\alpha$.

Example 5.6 An analyst is estimating risk measures for severity data for auto insurance policies. She has a sample of 200 values, and has set the GPD threshold at $d = 1.0$ (in $ millions). The parameters of the GPD are $\xi = 0.80$ and $\beta = 0.65$.

The 24 values from the sample which exceed the threshold are given in Table 5.2, in descending order:

(a) Calculate the 95%, 99%, and 99.9% quantiles, using the GPD.
(b) Calculate the 95%, 99%, and 99.9% Expected Shortfalls, using the GPD.

Solution 5.6

(a) First, we check that the empirical 95% quantile lies above $d = 1.0$. We see from the data that $Q_{95\%} \approx X_{(191)} = 2.3 > d$ as required.

Table 5.2. *Data points exceeding threshold,* $d = 1.0$*, from a sample of 200, for Example 5.6*

11.33	6.17	4.67	4.41	4.20	3.31	2.97	2.65	2.58	2.29	2.12	1.76
1.35	1.34	1.28	1.27	1.25	1.15	1.13	1.10	1.09	1.07	1.02	1.01

Next, we estimate $S_X(1.0)$. From the sample, we have 24 values exceeding the threshold from a total sample of 200 data values, so $S_X(d) \approx 24/200 = 0.12$.

Then using equation (5.22) we have

$$Q_\alpha = d + \frac{\beta}{\xi}\left(\left(\frac{S_X(d)}{1-\alpha}\right)^{\xi} - 1\right)$$

$$\Rightarrow Q_{95\%} = 1.0 + \frac{0.65}{0.80}\left(\left(\frac{0.12}{0.05}\right)^{0.8} - 1\right) = 1.82,$$

$$Q_{99\%} = 1.0 + \frac{0.65}{0.80}\left(\left(\frac{0.12}{0.01}\right)^{0.8} - 1\right) = 6.12,$$

$$Q_{99.9\%} = 1.0 + \frac{0.65}{0.80}\left(\left(\frac{0.12}{0.001}\right)^{0.8} - 1\right) = 37.6.$$

(b) From equation (5.25) we have

$$\mathrm{ES}_\alpha = \frac{1}{1-\xi}\left(Q_\alpha + \beta - \xi\, d\right)$$

$$\Rightarrow \mathrm{ES}_{95\%} = \frac{1}{0.2}\,(1.82 + 0.65 - 0.8(1.0)) = 8.35,$$

$$\mathrm{ES}_{99\%} = \frac{1}{0.2}\,(6.12 + 0.65 - 0.8(1.0)) = 29.85,$$

$$\mathrm{ES}_{99.9\%} = \frac{1}{0.2}\,(37.6 + 0.65 - 0.8(1.0)) = 187.25. \qquad \square$$

5.4.4 The Hill Estimator

An alternative approach to the empirical mean loss function for estimating d, β, and ξ is the **Hill estimator**, which estimates the tail index, $\alpha = 1/\xi$, for $\xi > 0$.

The Hill estimator uses the fact that for distributions in the MDA of the Fréchet distribution, the survival function can be written

$$S(x) = L(x)\, x^{-1/\xi} = L(x)\, x^{-\alpha},$$

where $L(x)$ is slowly varying at infinity. We then find that the mean excess loss function of the log of the loss data (we assume the losses are > 0) converges to $\xi = 1/\alpha$.

Suppose we have an ordered sample of loss data, $x_{(1)}, \ldots, x_{(n)}$. The Hill estimator is

$$\hat{\alpha}_j^H = \left(\sum_{k=j+1}^{n} \frac{\log(x_{(k)})}{n-j+1} - \log(x_{(j)}) \right)^{-1},$$

which is a slight variant of the empirical MEL function above, applied to the logs of the sample values.

Since different values of j will give different estimators, it is customary to plot values for a range of j, towards the higher end of the sample.

We select a threshold at the $(n-j)$th sample value, that is, at $d = x_{(n-j)}$, which means that the probability that $X > d$ is estimated to be $\hat{S}(d) = j/n$. The Hill estimator for the survival function for $x > x_{(n-j)}$ is then

$$\hat{S}^H(x) = \frac{j}{n} \left(\frac{x}{x_{(n-j)}} \right)^{-\hat{\alpha}_j^H}.$$

McNeil et al. (2015) show that this is similar to the estimate derived from equation (5.21), replacing $S_X(d)$ with the empirical estimate j/n, but without the scale parameter β.

5.5 Example: US Hurricane Losses, 1940–2012

In this section, we explore the use of extreme value distributions in the analysis of data relating to losses in the United States of America between 1940 and 2012, from 179 hurricanes and tropical storms.[1] The data is adjusted to 2017 values and losses are expressed in $millions.

The data is shown in Figure 5.5.

This analysis would be important, for example, for an insurer or reinsurer with exposure to hurricane losses. Suppose we are interested in estimating the 99% VaR or Expected Shortfall of the hurricane severity distribution. The five largest losses in the sample (in 2017 $millions) are

$L_{(175)}$	$L_{(176)}$	$L_{(177)}$	$L_{(178)}$	$L_{(179)}$
54,660	65,900	71,790	79,110	91,130

The empirical estimate of the 99% VaR from the data is approximately 81,514 (unsmoothed). The empirical estimate of the 99% Expected Shortfall is

[1] The data is derived from www.icatdamageestimator.com.

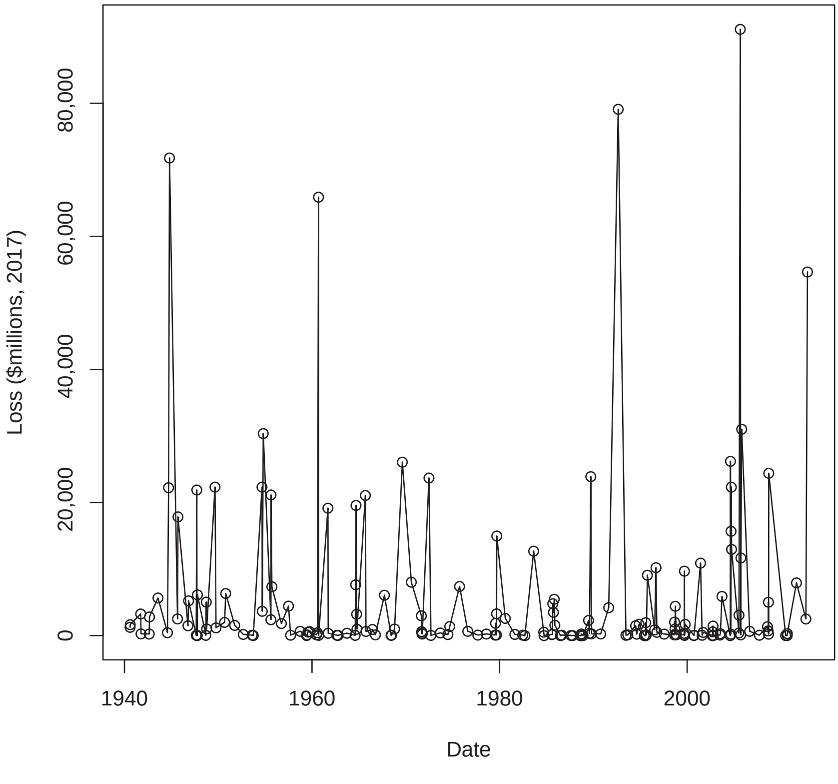


Figure 5.5 US hurricane loss data, 1940-2012, in 2017 values (millions)

the average of the values greater than 81,514, but there is only one such value, so the empirical Expected Shortfall is 91,130.

It is interesting to re-estimate the risk measures using the generalized Pareto distribution to model the tail of the severity distribution.

We first plot the empirical mean excess loss (MEL) function. The result is shown in Figure 5.6 The lower plot omits the final five values – the MEL typically fluctuates significantly in the tail. The increasing linear trend above a threshold of around 20,000 indicates that a GPD with positive ξ parameter should be the best fit.

Using the 'fit.GPD' function from the 'QRM' package in R, with a threshold of 20,000, gives maximum likelihood estimates for the GPD parameters (with associated standard errors) of

$$\hat{\xi} = 0.75(0.41), \qquad \hat{\beta} = 7{,}005(3{,}066).$$

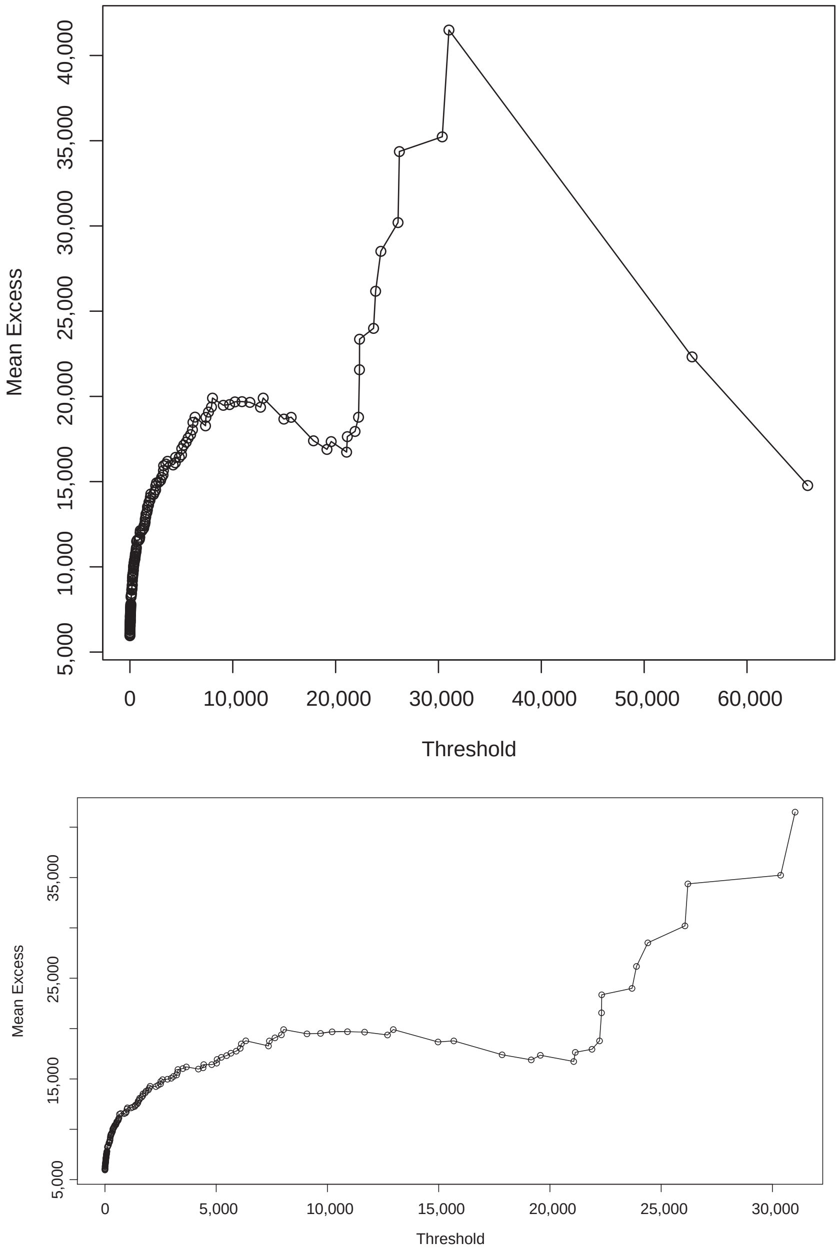


Figure 5.6 Empirical MEL for hurricane loss data

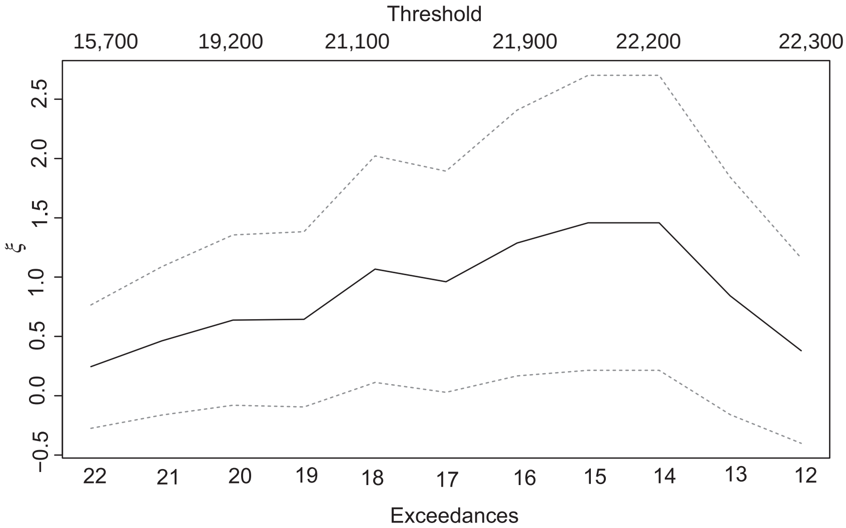


Figure 5.7 MLE estimators for ξ using different numbers of tail values, for the US hurricane loss data

Moving the threshold to 21,000 gives

$$\hat{\xi} = 1.2(0.53), \qquad \hat{\beta} = 3{,}850(1{,}960),$$

which demonstrates the difficulty of estimating the ξ parameter in practice. In Figure 5.7, we show the ξ estimates and 95% confidence intervals for the storm data, based on a range of the number of tail loss values to be included in the calculation.

In Figure 5.8, we show the estimates of ξ, with 95% confidence bands, using the Hill method. The range of thresholds is the same as in Figure 5.7, but the estimated values of ξ are quite different. The Hill estimator is less accurate than the maximum likelihood estimator used in Figure 5.7, and this sample does not generate enough tail values for the method to be suitable.

We use the fitted GPD, assuming a threshold of 20,000, to estimate the risk measures for the loss distribution. We use formula (5.22) for the estimated 99% VaR. That is

$$Q_\alpha = d + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{S_X(d)}\right)^{-\xi} - 1\right),$$

where $\alpha = 0.99$, $d = 20{,}000$, $\beta = 7{,}005$, $\xi = 0.75$. We estimate $S_X(d)$ using the empirical survival function at $d = 20{,}000$; that is, using the proportion of the data that is greater than 20,000. There are 179 values in the data set, and

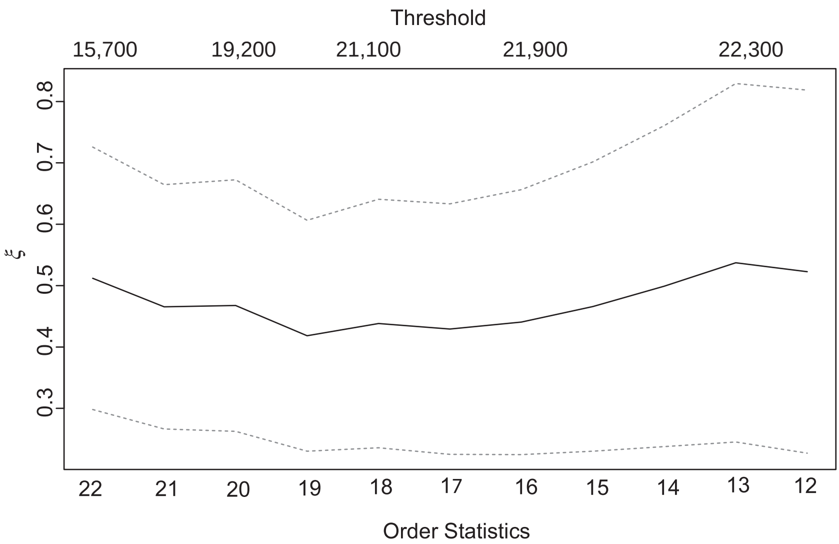


Figure 5.8 Hill estimators for ξ, using different threshold values, for the hurricane loss data, with 95% confidence interval

19 are greater than the threshold, so we use $S_X(d) = 19/179 = 0.1061$. This gives an estimate of $Q_{99\%} \approx 65{,}567$.

For the 99% Expected Shortfall we use formula (5.25),

$$\mathrm{ES}_\alpha = \frac{1}{1-\xi}\big(Q_\alpha + \beta - \xi\, d\big),$$

with the same parameters as used for Q_α. For $\alpha = 0.99$, this gives a 99% Expected Shortfall estimate of 230,291.

We see that the VaR estimate is fairly close to the empirical estimate. The Expected Shortfall, however, is considerably larger than the maximum observed loss; we use the GPD to extrapolate beyond the data, so our estimates are not bounded by the values observed in the past. We see the influence of the extrapolation in Table 5.3, which shows the empirical estimates of the VaR and Expected Shortfall for several different values of α, using smoothed empirical estimates for the VaR. We compare them with the estimates found using the GPD. As the risk measure moves further into the tail, the Expected Shortfall is highly influenced by the GPD. In the last row, we illustrate that for risk measures using α beyond the availability in the data, the GPD extrapolation can be used to give results consistent with the upper tail of the data.

Table 5.3. *Comparison of smoothed empirical and GPD estimates of risk measures for US hurricane loss data (millions)*

α	Empirical VaR	Empirical ES	GPD VaR	GPD ES
0.90	21,140	38,783	20,424	49,716
0.95	26,070	56,273	27,081	76,345
0.99	81,514	91,130	65,567	230,291
0.999	n/a	n/a	319,429	1,245,735

5.6 Notes and Further Reading

The two perspectives on EVT discussed in this chapter, block maxima and points over threshold (POT), both have practical applications, but the POT approach is much more useful in applied risk management, as it focuses on exactly the statistic that is measured with the Expected Shortfall risk measure. Both perspectives categorize the nature of the distribution in its extreme tail using the ξ parameter; in practice, the uncertainty in the estimated value of ξ can be a significant problem. Ultimately, the nature of extreme values means that the available data is slim and, therefore, the uncertainty is high. Nevertheless, EVT offers a valuable approach for evaluating far right tail risk measures of loss.

Embrechts et al. (2013) offers a deep exploration of EVT in risk management, with a focus on insurance and finance.

5.7 Exercises

Exercise 5.1 Describe the advantages of using EVT to calculate tail risk measures of a distribution, compared with a parametric model of the full loss distribution.

Exercise 5.2 Describe the trade-off involved in selecting block sizes for the block maxima approach to estimating ξ. Explain how the selection influences (i) the bias and (ii) the variance of the estimate.

Exercise 5.3 The normal distribution is in the MDA of the Gumbel extreme value distribution.

(a) Explain in words what this means.
(b) The normal distribution is symmetric, while the Gumbel distribution is positively skewed. Explain why this is not inconsistent.

Exercise 5.4 Show that, if X follows the GEV distribution with parameters $\xi < 0$, μ, and $\theta > 0$, then $Y = 1+\xi(x-\mu)/\theta$ follows the Weibull distribution defined in Section 4.4.2, and identify the parameters of the standard Weibull distribution.

Exercise 5.5 Consider an exponential distribution with distribution function

$$F(x) = 1 - e^{-x/\beta}.$$

(a) Show that F is in the maximum domain of attraction of the Gumbel distribution, using normalizing sequences $c_n = \beta$ and $d_n = \beta \log n$.
(b) Let M denote the maximum of 80 observations of an exponential distribution with mean 100. Calculate the probability that $M > 1{,}000$ using (i) the exponential distribution and (ii) the Gumbel distribution.
(c) Comment on your results.

Exercise 5.6 Describe the trade-off involved in selecting a threshold for the points over threshold approach to estimating ξ. Explain how the selection influences (i) the bias and (ii) the variance of the estimate.

Exercise 5.7 You are given the following information about three random variables that are in the MDA of the GEV distribution. State with reasons whether $\xi < 0$, $\xi = 0$, or $\xi > 0$.

(a) The first random variable, X, is greater than 0, unbounded on the right side, and has finite kth moment for all $k = 1, 2, 3, \ldots$
(b) The second random variable, Y, is equal to $-X$.
(c) The third random variable has a finite number of moments.

Exercise 5.8 A company is modelling losses from cyberattacks.

The following table shows the largest 20 values of a sample of 1,000 observations of the losses, sorted in decreasing order:

196.1	148.2	79.8	35.8	27.1	22.9	21.8	20.9	16.7	15.8
15.5	13.7	13.0	12.5	11.5	10.3	9.7	9.4	8.3	8.0

(a) Estimate the 99% VaR of X.
(b) Estimate the 99% Expected Shortfall of X.
(c) Assume that the losses are from a distribution which is GPD for $Y = X - d|X > d$, where $d = 10$, with shape parameter $\gamma = 0$, and scale parameter $\beta = 30$.

(i) Show that

$$\Pr[X > x] = S_X(d)\left(e^{-(x-d)/\beta}\right) \quad \text{for } x > d.$$

(ii) Estimate the 99% VaR and 99% Expected Shortfall using the GPD for the tail probabilities.

(iii) Comment on the differences between your estimates in (a) and (b), and your estimates using the GPD.

Exercise 5.9 Consider the one-parameter GEV distribution

$$F(x) = e^{-x^{-1/\gamma}}, \qquad 0 < \gamma < 1.$$

Show that $\mathrm{E}[X] = \Gamma(1-\gamma)$, where $\Gamma(u) = \int_0^\infty t^{u-1}\, e^{-t}\, dt$.

Exercise 5.10 Consider the Pareto distribution with distribution function

$$F(x) = 1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha}.$$

(a) Show that this distribution is in the Fréchet MDA, with distribution function $e^{-y^{-\alpha}}$, using normalizing sequences

$$c_n = \theta\, n^{1/\alpha}, \quad \text{and } d_n = -\theta.$$

(b) Hence determine the normalizing sequences c_n^* and d_n^* such that the limiting distribution for the maximum is the GEV distribution

$$H(x) = e^{-(1+\gamma x)^{-1/\gamma}}.$$

(c) Given $\theta = 200$, $\alpha = 3.0$,

(i) Compare the exact probability that $M_n \leq m$ with the probability using the GEV distribution applied to normalized M_n, for $n = 25$ and $m = 500$.

(ii) Repeat (i) for $m = 500$, and $n = 50$, and $n = 100$.

(iii) Repeat (i) with $n = 25$, $m = 1{,}000$, and $m = 2{,}000$.

(iv) Comment on your results.

Exercise 5.11 You are given that the maximum of a sample of n independent variables has a Fréchet distribution, with parameters α and θ ($\mu = 0$).

Show that the maximum of a sample of $2n$ variables also has a Fréchet distribution, and determine the parameters.

Exercise 5.12 You are given that the monthly maximum for a financial series monitored daily has a Gumbel distribution, with location and scale parameters μ and θ.

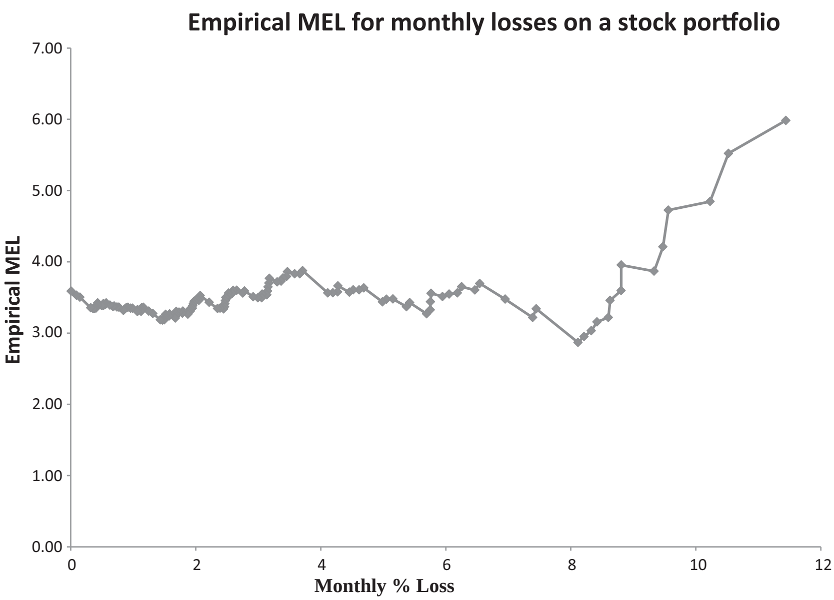
Show that the annual maximum also has a Gumbel distribution, and determine the parameters.

Exercise 5.13 You are given that the Fréchet distribution with parameter α is in the MDA of the Fréchet distribution, with the same parameter α. You are also given that for this distribution the normalizing sequence $d_n = 0$. What is the c_n normalizing sequence in this case?

A distribution with the property that it lies in the MDA of the same distribution, with the same parameter is called a **max stable** distribution.

Exercise 5.14 An investment firm is interested in fitting an extreme value distribution to portfolio losses. The data set comprises 300 values for the monthly percentage loss in portfolio value.

The analyst has assumed the data is a random sample of independent identically distributed observations. She has calculated the empirical MEL function, which is shown in the figure below.



You are also given the 20 largest values from the data, sorted from high to low:

24.3	18.4	15.6	11.4	10.5	10.2	9.6	9.5	9.3	8.8
8.8	8.6	8.6	8.4	8.3	8.2	8.1	7.4	7.4	6.9

(a) (i) Estimate the 98% Expected Shortfall from the data.
(ii) Calculate the empirical MEL function, $\hat{e}(x)$ at $x = 10.0$.
(b) The analyst decides to use the GPD with threshold $d = 8$. Explain why this choice appears reasonable.
(c) The MLE estimators for the GPD distribution, with $d = 8$, are $\hat{\xi} = 0.55$, $\hat{\beta} = 1.5$.

(i) Calculate the 98% Expected Shortfall using the GPD distribution.
(ii) Calculate the 99.9% Expected Shortfall using the GPD distribution.

(d) Comment on the difference between the Expected Shortfall estimates from (a)(i) and (c)(i). Which method would you recommend?

Exercise 5.15 A shipping company is reviewing its expected losses from events at sea, in order to determine how much insurance cover is needed.

The number of journeys each year, denoted N, has a Poisson distribution with expected value 100.

In 80% of journeys, there is no loss. In 19% of journeys, there is a minor loss, with severity following a lognormal distribution, with parameters $\mu = 7$, $\sigma = 1.6$. In the remaining 1% of journeys, there is a major loss, and the severity in these cases follows a Pareto distribution with parameters $a = 2.1$ and $\theta = 66{,}000$.

Let Y_j denote the loss from the jth voyage, and let I denote the type of loss involved, where $I \in \{\text{No Loss, Minor Loss, Major Loss}\}$.

You are given that Y_j are i.i.d. and are independent of N.

(a) Calculate the mean of the aggregate annual loss.
(b) The company is interested in insuring jumbo losses, defined as losses exceeding 50,000.

(i) Calculate $\Pr[Y_j > 50{,}000 | I = \text{Minor Loss}]$ and $\Pr[Y_j > 50{,}000 | I = \text{Major Loss}]$, and hence determine $\Pr[Y_j > 50{,}000]$.
(ii) Given that an individual loss exceeds 50,000, calculate the probability that the loss arose from a major loss event.

(c) You are given that the lognormal distribution lies in the Gumbel MDA, and the Pareto lies in the Fréchet MDA.

A consultant claims that Y_j will be in the Gumbel MDA, because there is a higher probability that Y_j comes from the lognormal distribution than from the Pareto distribution. Critique this claim.
(d) The company insures losses from major loss events only. The remaining uninsured loss will therefore be

$$Y_j^* = \begin{cases} Y_j & I \neq \text{Major Loss}, \\ \min(Y_j, 50{,}000) & I = \text{Major Loss}. \end{cases}$$

Show that the expected value of the annual aggregate uninsured loss is 102,700 to the nearest 100.

(e) The insurer offers the cover at a cost of 47,000. As an alternative, it offers a co-insurance policy with the same expected value of insured losses for only 37,000. Under a co-insurance policy, the insurer pays a fixed proportion of all losses.

(i) Determine the proportion of each loss retained by the shipping firm under the co-insurance.

(ii) Comment on how the shipping company might decide between the two policies.