Financial Risk Forecasting
Chapter 9
Extreme Value Theory

Jon Danielsson ©2020
London School of Economics

To accompany
Financial Risk Forecasting
www.financialriskforecasting.com
Published by Wiley 2011
Version 1.0, August 2015
The focus of this chapter is on

- Basic introduction to extreme value theory (EVT)
- Asset returns and fat tails
- Applying EVT
- Aggregation and convolution
- Time dependence
Notation

\( \nu \)  Tail index
\( \xi = 1/\nu \)  Shape parameter
\( M_T \)  Maximum of \( X \)
\( C_T \)  Number of observations in the tail
\( u \)  Threshold value
\( \psi \)  Extremal index
Extreme Value Theory
Types of tails

• In this book, we follow the convention of EVT being presented in terms of the *upper tails* (i.e. *positive observations*)

• In most risk analysis we are concerned with the *negative observations* in the lower tails, hence to follow the convention, we can *pre-multiply returns by -1*

• Note, the upper and lower tails do not need to have the same thickness or shape
Extreme value distributions

• In most risk applications, we do not need to focus on the entire distribution
• The main result of EVT states that the tails of all distributions fall into one of three categories, regardless of the overall shape of the distribution
  - See next slide for the three distributions
• Note, this is true given the distribution of an asset return does not change over time
**Weibull** Thin tails where the distribution has a finite endpoint (e.g. the distribution of mortality and insurance/re-insurance claims)

**Gumbel** Tails decline exponentially (e.g. the normal and log-normal distributions)

**Fréchet** Tails decline by a *power law*; such tails are know as “fat tails” (e.g. the Student-t and Pareto distributions)
Extreme value distributions

Weibull
Extreme value distributions

- Weibull
- Gumbel
Extreme value distributions

- Weibull
- Gumbel
- Frechet
Fréchet distribution

- From the last slide, the Weibull clearly has a finite endpoint.
- And the Fréchet tail is thicker than the Gumbel’s.
- In most applications in finance, we know that returns are fat tailed.
- Hence we limit our attention to the Fréchet case.
Generalized extreme value distribution

- The Fisher and Tippett (1928) and Gnedenko (1943) theorems are the fundamental results in EVT.
- The theorems state that the maximum of a sample of properly normalized IID random variables converges in distribution to one of the three possible distributions: the Weibull, Gumbel or the Fréchet.
Generalized extreme value distribution

- The Fisher and Tippett (1928) and Gnedenko (1943) theorems are the fundamental results in EVT.
- The theorems state that the maximum of a sample of properly normalized IID random variables converges in distribution to one of the three possible distributions: the Weibull, Gumbel or the Fréchet.
- An alternative way of stating this is in terms of the maximum domain of attraction (MDA).
- MDA is the set of limiting distributions for the properly normalized maxima as the sample size goes to infinity.
Fisher-Tippett and Gnedenko theorems

- Let $X_1, X_2, \ldots, X_T$ denote IID random variables (RVs) and the term $M_T$ indicate maxima in sample of size $T$
- The *standardized distribution* of maxima, $M_T$, is

$$\lim_{T \to \infty} \Pr \left\{ \frac{M_T - a_T}{b_T} \leq x \right\} = H(x)$$

where the constants $a_T$ and $b_T > 0$ exist and are defined as $a_T = T \mathbb{E}(X_1)$ and $b_T = \sqrt{\text{Var}(X_1)}$
Fisher-Tippett and Gnedenko theorems

- Then the limiting distribution, $H(.)$, of the maxima as the \textit{generalized extreme value (GEV)} distribution is

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

- Depending on the value of $\xi$, $H_\xi(.)$ becomes one of the three distributions:
  - if $\xi > 0$, $H_\xi(.)$ is the Fréchet
  - if $\xi < 0$, $H_\xi(.)$ is the Weibull
  - if $\xi = 0$, $H_\xi(.)$ is the Gumbel
Asset Returns and Fat Tails
Fat tails

• The term “fat tails” can have several meanings, the most common being “extreme outcomes occur more frequently than predicted by normal distribution”

• While such a statement might make intuitive sense, it has little mathematical rigor as stated

• The most frequent definition one may encounter is Kurtosis, but it is not always accurate at indicating the presence of fat tails ($\kappa > 3$)

• This is because kurtosis is more concerned with the sides of the distribution rather than the heaviness of tails
The formal definition of fat tails comes from *regular variation*.

**Regular variation** A random variable, $X$, with distribution $F(.)$ has fat tails if it varies regularly at infinity; that is there exists a positive constant $\lambda$ such that:

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\lambda}, \quad \forall x > 0, \lambda > 0$$
Tail distributions

• In the fat-tailed case, the tail distribution is Fréchet:

\[ H(x) = \exp(-x^{-\nu}) \]

**Lemma** A random variable \( X \) has regular variation at infinity (i.e. has fat tails) if and only if its distribution function \( F \) satisfies the following condition:

\[ 1 - F(x) = \Pr\{X > x\} = Ax^{-\nu} + o(x^{-\nu}) \]

for positive constant \( A \), when \( x \to \infty \)
Tail distributions

- The expression $o(x^{-\iota})$ is the *remainder term* of the Taylor-expansion of $Pr\{X > x\}$, it consists of terms of the type $Cx^{-j}$ for constant $C$ and $j > \iota$

- As $x \to \infty$, the tails are asymptotically Pareto-distributed:
  \[ F(x) \approx 1 - Ax^{-\iota} \]

  where $A > 0$; $\iota > 0$; and $\forall x > A^{1/\iota}$
Normal and fat distributions

Normal and Student-\(t\) densities
Normal and fat distributions

Normal and Student-\( t \) densities
Normal and fat distributions

Pareto tails

---

Normal
Normal and fat distributions

Pareto tails

- Normal
- $\tau=2$
Normal and fat distributions

Pareto tails

- Normal
- $\alpha=2$
- $\alpha=4$
Normal and fat distributions

Pareto tails

- Normal
- \( \iota = 2 \)
- \( \iota = 4 \)
- \( \iota = 6 \)
Normal and fat distributions

- The definition demonstrates that fat tails are defined by how rapidly the tails of the distribution decline as we approach infinity.
- As the tails become thicker, we detect increasingly large observations that impact the calculation of moments:

\[ E(X^m) = \int x^m f(x) \, dx \]

- If \( E(X^m) \) exists for all positive \( m \), such as for the normal distribution, the definition of regular variation implies that moments \( m \geq \nu \) are not defined for fat-tailed data.
Applying EVT
Implementing EVT in practice

Two main approaches:

1. Block maxima
2. Peaks over thresholds (POT)
Block maxima approach

- This approach follows directly from the regular variation definition where we estimate the GEV by dividing the sample into blocks and using the maxima in each block for estimation.
- The procedure is rather wasteful of data and a relatively large sample is needed for accurate estimate.
Peaks over thresholds approach

• This approach is generally preferred and forms the basis of our approach below
• It is based on models for all large observations that exceed a high threshold and hence makes better use of data on extreme values
• There are two common approaches to POT:
  1. Fully parametric models (e.g. the Generalized Pareto distribution or GPD)
  2. Semi-parametric models (e.g. the Hill estimator)
### Generalized Pareto distribution

1. Consider a random variable $X$, fix a threshold $u$ and focus on the **positive part of $X - u$**;
2. The distribution $F_u(x)$ is
   \[ F_u(x) = \Pr(X - u \leq x | X > u) \]
3. If $u$ is VaR, then $F_u(x)$ is the probability that we exceed VaR by a particular amount (a shortfall) given that VaR is violated;
4. Key result is that as $u \to \infty$, $F_u(x)$ converges to the GPD, $G_{\xi, \beta}(x)$.
• The GPD $G_{\xi, \beta}(x)$ is

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

where $\beta > 0$ is the scale parameter; $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\frac{\beta}{\xi}$ when $\xi < 0$

• We therefore need to estimate both shape($\xi$) and scale($\beta$) parameters when applying GDP

• Recall, for certain values of $\xi$ the shape parameters, $G_{\xi, \beta}(.)$ becomes one of the three distributions
GEV and GPD

- The GEV is the limiting distribution of normalized maxima, whereas the GPD is the limiting distribution of normalized data beyond some high threshold.
- Note, the tail index is the same for both GPD and GEV distributions.
- The parameters of GEV can be estimated from the log-likelihood function of GPD.
**VaR under GPD**

The VaR in the GPD case is:

$$\text{VaR}(p) = u + \frac{\beta}{\xi} \left[ \left( \frac{1 - p}{F(u)} \right)^{-\xi} - 1 \right]$$
Hill method

- Alternatively, we could use the semi-parametric Hill estimator for the tail index in distribution $F(x) \approx 1 - Ax^{-\hat{\nu}}$:

  $$\hat{\xi} = \frac{1}{\hat{\nu}} = \frac{1}{C_T} \sum_{i=1}^{C_T} \log \frac{x(i)}{u}$$

  where $x(i)$ is the notation of sorted data, e.g. maxima is denoted as $x(1)$

- As $T \to \infty$, $C_T \to \infty$ and $C_T/T \to 0$

- Note that the Hill estimator is sensitive to the choice of threshold, $u$
Which method to choose?

- **GPD**, as the name suggests, is more general and can be applied to all three types of tails
- **Hill method** on the other hand is in the maximum domain of attraction (MDA) of the Fréchet distribution
- Hence Hill method is only valid for fat-tailed data
Risk analysis

• After estimation of the tail index, the next step is to apply a risk measure
• The problem is finding \( \text{VaR}(p) \) such that

\[
\Pr [X \leq -\text{VaR}(p)] = F_X (-\text{VaR}(p)) = p
\]

where \( F_X(u) \) is the probability of being in the tail, that is the returns exceeding the threshold \( u \)
Risk analysis

- Let $G$ be the distribution of $X$ since we are in the left tail (i.e. $X \leq -u$). By the Pareto assumption we have:

$$G(-\text{VaR}(p)) = \left( \frac{\text{VaR}(p)}{u} \right)^{-\xi}$$

- And by the definition of conditional probability:

$$G(-\text{VaR}(p)) = \frac{p}{F_X(u)}$$
VaR estimator

- Equating the previous two relationship, we obtain:

$$\text{VaR}(p) = u \left( \frac{F_X(u)}{p} \right)^{\frac{1}{\xi}}$$

- $F_X(u)$ can be estimated by the proportion of data beyond the threshold $u$, $C_T/T$

- The VaR estimator is therefore:

$$\hat{\text{VaR}}(p) = u \left( \frac{C_T/T}{p} \right)^{\frac{1}{\xi}}$$
EVT often applied inappropriately

- EVT should only be applied in the tails
- The closer to the centre of the distribution, the more inaccurate the estimates are
- However, there are no rules to define when the estimates become inaccurate, it depends on the underlying distribution of the data
- In some cases, it may be accurate up to 1% or even 5%, while in other cases it is not reliable even up to 0.1%
Finding the threshold

- Actual implementation of EVT is relatively simple and delivers good estimates where EVT holds
- The sample size $T$ and the choice of probability level $p$ depends on the underlying distribution of the data
- As a *rule of thumb*: $T \geq 1000$ and $p \leq 0.4\%$
- For applications with smaller sample sizes or less extreme probability levels, other techniques should be used
  - Such as HS or fat-tailed GARCH
• It can be challenging to estimate EVT parameters given the effective sample size is small
• This relates to choosing the number of observations in the tail, \( C_T \)
• We have 2 conflicting directions:
  1. By lowering \( C_T \), we can reduce the estimation bias
  2. On the other hand, by increasing \( C_T \), we can reduce the estimation variance
Optimal threshold $C^*_T$
Optimal threshold \( C_T^* \)

- If the underlying distribution is known, then deriving the optimal threshold is easy, but in such a case EVT is superfluous.

- Most common approach to determine the optimal threshold is the \textit{eyeball method} where we look for a region where the tail index seems to be stable.

- More formal methods are based on minimizing the mean squared error (MSE) of the Hill estimator, but such methods are not easy to implement.
Application to the S&P 500 index

Returns from 1975 to 2015 – 10,000 observations
Distribution of S&P 500 returns

Empirical distribution
Distribution of S&P 500 returns

Tails truncated

![Graph showing the distribution of S&P 500 returns with empirical CDF and normal CDF, highlighting the tails being truncated.](image)
Hill plot for daily S&P 500 returns

From 1975 to 2015
Hill plot for daily S&P 500 returns

From 1975 to 2015

Optimal region
Upper and lower tails

The lower tail

- Empirical CDF
- EVT CDF
- Normal CDF
Upper and lower tails

The upper tail

![Graph showing empirical CDF, EVT CDF, and Normal CDF for returns.](image-url)
Aggregation and Convolution
Aggregation of outcomes

• The act of adding up observations across time is known as *time aggregation*

• And the act of adding up observations across assets/portfolios is termed *convolution*
**Theorem** Let $X_1$ and $X_2$ be two independent random variables with distribution functions satisfying

$$1 - F_i(x) = \Pr\{X_i > x\} \approx A_i x^{-\lambda_i} \quad i = 1, 2$$

when $x \to \infty$. Note, $A_i$ is a constant.

Then, the distribution function $F$ of the variable $X = X_1 + X_2$ in the positive tail can be approximated by 2 cases.
Case 1  When $\iota_1 = \iota_2$ we say that the random variables are first-order similar and we set $\iota = \iota_1 = \iota_2$ and $F$ satisfies

$$1 - F(x) = \Pr\{X > x\} \approx (A_1 + A_2)x^{-\iota}$$

Case 2  When $\iota_1 \neq \iota_2$ we set $\iota = \min(\iota_1, \iota_2)$ and $F$ satisfies

$$1 - F(x) = \Pr\{X > x\} \approx Ax^{-\iota}$$

where $A$ is the corresponding constant
• As a consequence, if two random variables are *identically distributed*, the distribution function of the sum (Case 1) will be given by

\[ \Pr\{X_1 + X_2 > x\} \approx 2Ax^{-t} \]

• Hence the probability doubles when we combine two observations from different days

• But if one observations comes from a fatter tailed distribution than the other, then only the heavier tail matters (Case 2)
Time scaling

**Theorem (de Vries 1998)** Suppose $X$ has finite variance with a tail index $\nu > 2$. At a constant risk level $\rho$, increasing the investment horizon from 1 to $T$ periods increases the VaR by a factor:

$$T^{1/\nu}$$

Note, EVT distributions retain the same tail index for longer period returns.
• Recall from chapter 4, under Basel Accords, financial institutions are required to calculate VaR for a 10-day holding periods
• The rules allow the 10-day VaR to be calculated by scaling the one-day VaR by $\sqrt{10}$
• The theorem shows that the scaling parameter is slower than the square-root-of-time adjustment
• Intuitively, as extreme values are more rare, they should aggregate at a slower rate than the normal distribution
• For example, if $\lambda = 4$, $10^{1/\lambda} = 1.78$, which is less than $\sqrt{10} = 3.16$
### VaR and the time aggregation of fat tail distributions

<table>
<thead>
<tr>
<th>Risk level</th>
<th>5%</th>
<th>1%</th>
<th>0.5%</th>
<th>0.1%</th>
<th>0.05%</th>
<th>0.005%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extreme value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Day</td>
<td>0.9</td>
<td>1.5</td>
<td>1.7</td>
<td>2.5</td>
<td>3.0</td>
<td>5.1</td>
</tr>
<tr>
<td>10 Day</td>
<td>1.6</td>
<td>2.5</td>
<td>3.0</td>
<td>4.3</td>
<td>5.1</td>
<td>8.9</td>
</tr>
<tr>
<td><strong>Normal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Day</td>
<td>1.0</td>
<td>1.4</td>
<td>1.6</td>
<td>1.9</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td>10 Day</td>
<td>3.2</td>
<td>4.5</td>
<td>4.9</td>
<td>5.9</td>
<td>6.3</td>
<td>7.5</td>
</tr>
</tbody>
</table>
• For one-day horizons, we see that in general EVT VaR is higher than VaR under normality, especially for more extreme risk levels
• This is balanced by the fact that 10-day EVT VaR is less than the normal VaR
• This seems to suggest that the square-root-of-time rule may be sufficiently prudent for longer horizons
• It is important to keep in mind that \( \sqrt{t} \) root rule (de Vries) only holds \textit{asymptotically}
Time Dependence
Time dependence

• Recall the assumption of IID returns in the section on EVT, which suggests that EVT may not be relevant for financial data
• Fortunately, we do not need an IID assumption, since EVT estimators are consistent and unbiased even in the presence of higher moment dependence
• We can explicitly model extreme dependence using the extremal index
Example

- Let us consider extreme dependence in a MA(1) process:

\[ Y_t = X_t + \alpha X_{t-1} \quad |\alpha| < 1 \]

- Let \( X_t \) and \( X_{t-1} \) be IID such that \( \Pr\{X_t > x\} \to Ax^{-\iota} \) as \( x \to \infty \). Then by Feller’s theorem

\[ \Pr\{Y_t \geq x\} \approx (1 + \alpha^t)Ax^{-\iota} \quad \text{as} \quad x \to \infty \]

- Dependence enters “linearly” by means of the coefficient \( \alpha_t \). But the tail shape is unchanged.

- This example suggest that time dependence has same effect as having an IID sample with fewer observations.
• Suppose we record each observation twice:

\[ Y_1 = X_1, \quad Y_2 = X_1, \quad Y_3 = X_2, \ldots \]

• And it increases the sample size to \( D = 2T \). Let us define \( M_D \equiv \max(Y_1, ..., Y_D) \). Evidently from Fisher-Tippett and Gnedenko theorem:

\[ \Pr\{M_D \leq x\} = F_T(x) = F_{D/2}^D(x) \]

supposing \( a_T = 0 \) and \( b_T = 1 \)

• The important result here is that dependence increases the probability that the maximum is below threshold \( x \)
Extremal index

**Extremal index** \( \psi \)  It is a measure of tail dependence and \( 0 < \psi \leq 1 \)

- If the data are *independent* then we get
  \[
  \Pr\{M_T \leq x\} \to e^{-x-\iota} \quad \text{as } T \to \infty 
  \]
  when \( a_T = 0 \) and \( b_T = 1 \)

- If the data are *dependent*, the limit distribution is
  \[
  \Pr\{M_D \leq x\} \to \left(e^{-x-\iota}\right)^\psi = e^{-\psi(x-\iota)}
  \]
• \( \frac{1}{\psi} \) is a measure of the \textit{cluster size} in large samples, for double-recorded data \( \psi = \frac{1}{2} \)

• For the MA(1) process in the previous example, we obtain the following

\[
\Pr \left\{ T^{-\frac{1}{1}} M_D \leq x \right\} \rightarrow \exp \left( -\frac{1}{1 + \alpha^t} x^{-t} \right)
\]

where \( \psi = \frac{1}{1+\alpha^t} \)
## Dependence in ARCH

- Consider the normal ARCH(1) process:

\[
Y_t = \sigma_t Z_t \\
\sigma_t^2 = \omega + \alpha Y_{t-1}^2 \\
Z_t \sim \mathcal{N}(0, 1)
\]

- Subsequent returns are uncorrelated but are not independent, since

\[
\text{Cov}(Y_t, Y_{t-1}) = 0 \\
\text{Cov}(Y_t^2, Y_{t-1}^2) \neq 0
\]
• Even when $Y_t$ is conditionally normally distributed, we noted in chapter 2 that the unconditional distribution of $Y$ is fat tailed

• de Haan et al. show that the unconditional distribution of $Y$ is given by

\[
\Gamma \left( \frac{\iota}{2} + \frac{1}{2} \right) = \sqrt{\pi} (2\alpha)^{-\iota/2}
\]
Extremal index for ARCH(1) – Example

- Extremal index for the ARCH(1) process can be solved using the previous equation.
- From the table below, we see that the higher the $\alpha$, the fatter the tails and the higher the level of clustering.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.10</th>
<th>0.50</th>
<th>0.90</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\iota$</td>
<td>26.48</td>
<td>4.73</td>
<td>2.30</td>
<td>2.02</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.99</td>
<td>0.72</td>
<td>0.46</td>
<td>0.42</td>
</tr>
</tbody>
</table>

- Similar results can be obtained for GARCH.
When does dependence matter?

- The importance of extreme dependence and the extremal index $\psi$ depends on the underlying applications.
- Dependence can be *ignored* if we are dealing with *unconditional probabilities*.
- And dependence *matters* when calculating *conditional probabilities*.
- For many stochastic processes, including GARCH, the time between tail events become increasingly independent.
Example – S&P 500 index extremes

From 1970 to 2015, 1% events
Example – S&P 500 index extremes

From 1970 to 2015, 0.1% events
Example – S&P 500 index extremes

0.1% events during the crisis

-10%
-5%
0%
5%
10%

Sep 08
Nov 08
Jan 09
Mar 09