Estimation of the marginal expected shortfall

Juan Juan Cai
Tilburg University, NL

Laurens de Haan
Erasmus University Rotterdam, NL
University of Lisbon, PT

John H. J. Einmahl
Tilburg University, NL

Chen Zhou
De Nederlandsche Bank
Erasmus University Rotterdam, NL
Expected shortfall of an asset $X$ at probability level $p$ is

$$E\left(-X \mid X \leq -F_x^{-}\left(p\right)\right)$$

where $F_x(x) := P\{X \leq x\}$ and $F_x^{-}$ the inverse function of $F_x$. 
A bank holds a portfolio \( R = \sum_i y_i R_i \)

Expected shortfall at probability level \( p \)

\[
-E(R|R < -\text{VaR}_p)
\]

Can be decomposed as

\[
-\sum_i y_i E(R_i|R < -\text{VaR}_p)
\]

The sensitivity to the i-th asset is

\[
-E(R_i|R < -\text{VaR}_p)
\]

(is marginal expected shortfall in this case)
More generally:

Consider a random vector \((X,Y)\)

Marginal expected shortfall (MES) of \(X\) at level \(p\) is

\[
E(X \mid Y > F_Y^{-}(1-p))
\]

(these are losses hence “\(Y\) big” is bad).

All these are risk measures i.e. characteristics that are indicative of the risk a bank occurs under stress conditions.
We are interested in **MES** under exceptional stress conditions of the kind that have occurred very rarely or even not at all. This is the kind of situation where extreme value can help.

We want to estimate $E(X \mid Y > F_Y^{-1}(1 - p))$ for small $p$ on the basis of i.i.d. observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ and we want to prove that the estimator has good properties.
When we say that we want to study a situation that has hardly ever occurred, this means that we need to consider the case \( p \leq \frac{1}{n} \) i.e., when a non-parametric estimator is impossible, since we need to extrapolate.

On the other hand we want to obtain a limit result, as \( n \) (the number of observations) goes to infinity. Since the inequality \( p \leq \frac{1}{n} \) is essential, we then have to assume \( p = p_n \) and \( np_n = O(1) \) as \( n \to \infty \).
Note that a parametric model in this situation is also not realistic:

The model is generally chosen to fit well in the central part of the distribution but we are interested in the (far) tail where the model may not be valid.

Hence it is better to “let the tail speak for itself”.

This is the semi-parametric approach of extreme-value theory.
Notation: (t big and \( p \) small, \( t = \frac{1}{p} \))

\[
U_1(t) := F_X^{-}\left(1 - \frac{1}{t}\right)
\]

\[
U_2(t) := F_Y^{-}\left(1 - \frac{1}{t}\right)
\]

\[
\theta_p := E\left(X \mid Y > U_2\left(\frac{1}{p}\right)\right)
\]

MES
\[ \theta_p = E \left( X \mid Y > U_2 \left( \frac{1}{p} \right) \right) = \frac{\int_0^\infty P \left( X > x, Y > U_2 \left( \frac{1}{p} \right) \right) dx}{P \left\{ Y > U_2 \left( \frac{1}{p} \right) \right\} } \]

\[ = \frac{1}{p} \int_0^\infty P \left\{ X > x, Y > U_2 \left( \frac{1}{p} \right) \right\} dx \]

\[ = \frac{1}{p} U_1 \left( \frac{1}{p} \right) \int_0^\infty P \left\{ X > xU_1 \left( \frac{1}{p} \right), Y > U_2 \left( \frac{1}{p} \right) \right\} dx \text{ i.e.,} \]

\[ \frac{\theta_p}{U_1 \left( \frac{1}{p} \right)} = \frac{1}{p} \int_0^\infty P \left\{ X > xU_1 \left( \frac{1}{p} \right), Y > U_2 \left( \frac{1}{p} \right) \right\} dx \]
We consider the limit of this as $p \downarrow 0$.

**Conditions (1):** First note (take $x = 1$ upstairs)

$$P\left\{X > U_1\left(\frac{1}{p}\right), \ Y > U_2\left(\frac{1}{p}\right)\right\}$$

$$= P\{1 - F_1(X) < p, \ 1 - F_2(Y) < p\}$$

where $F_1$ and $F_2$ are the distribution functions of $X$ and $Y$.

This is a copula.
We impose conditions on the copula as $p \downarrow 0$:

Suppose there exists a positive function $R(x, y)$ (the dependence function in the tail) such that for all $0 \leq x, y \leq \infty$, $x \lor y > 0$, $x \land y < \infty$

$$\lim_{p \downarrow 0} \frac{1}{p} P \left\{ X > U_1 \left( \frac{x}{p} \right), \ Y > U_2 \left( \frac{y}{p} \right) \right\} = R(x, y) \quad \text{i.e.,}$$

$$\lim_{p \downarrow 0} \frac{1}{p} P \left\{ 1 - F_1(X) < \frac{p}{x}, \ 1 - F_2(Y) < \frac{p}{y} \right\} = R(x, y) \cdot$$
This condition indicates and specifies dependence specifically in the tail.

(usual condition in extreme value theory)

(2): Compare: in the definition of $\theta_p$, we have

$$P \left\{ X > x U_1 \left( \frac{1}{p} \right), \ Y > U_2 \left( \frac{1}{p} \right) \right\}$$

and in the condition we have (for $y = 1$)

$$P \left\{ X > U_1 \left( \frac{x}{p} \right), \ Y > U_2 \left( \frac{1}{p} \right) \right\}.$$
In order to connect the two we impose a second condition, on the tail of $X$: for $x > 0$

$$\lim_{t \to \infty} \frac{P\{X > tx\}}{P\{X > t\}} = x^{-\frac{1}{\gamma_1}}.$$

Where $\gamma_1$ is a positive parameter.

This second condition implies a similar condition for the quantile function $U_1(t) = F^{-1}\left(1 - \frac{1}{t}\right)$ namely

$$\lim_{t \to \infty} \frac{U_1(tx)}{U_1(t)} = x^{\gamma_1} \quad (x > 0).$$
We say that $P\{X > t\}$ is “regularly varying at infinity” with index $-1/\gamma_1$ ($\in R.V_{-1/\gamma_1}$) and $U_1$ is also regularly varying, with index $\gamma_1$.

(usual condition is extreme value theory)

These two conditions are the basic conditions of one-dimensional extreme value theory.
Examples:

Student distribution, Cauchy distribution.

It is quite generally accepted that most financial data satisfy this condition.

**Sufficient condition:** \(1 - F(t) = ct^{\frac{-1}{n}} + \text{lower order powers}\).

Under these conditions we get the first result:
\[
\lim_{p \downarrow 0} \frac{\theta_p}{U_1\left(\frac{1}{p}\right)} = \lim_{p \downarrow 0} \frac{E\left(\frac{1}{p} \Big| X > U_2\left(\frac{1}{p}\right)\right)}{U_1\left(\frac{1}{p}\right)} = \int_0^\infty R\left(\frac{1}{x^{1/n}}, 1\right)dx
\]

Hence \( \theta_p \) goes to infinity as \( p \downarrow 0 \) at the same rate as \( U_1\left(\frac{1}{p}\right) \), the value-at-risk for \( X \).

Now we go to statistics and look at how to estimate \( \theta_p \).
We do that in stages:

First we estimate \( \theta_{k/n} \) where \( k = k(n) \to \infty, k(n)/n \to 0 \) as \( n \to \infty \).

Clearly we can estimate \( \theta_{k/n} \) non-parametrically (it is just inside the sample).

The second stage will be the extrapolation from \( \theta_{k/n} \) to \( \theta_p \) with \( p \leq 1/n \).

For the time being we suppose that \( X \) is a positive random variable.
Recall \[ \theta_{\frac{k}{n}} = E\left( X \mid Y > U_2\left( \frac{n}{k} \right) \right) \]

**First step:** replace quantile \( U_2(\frac{n}{k}) \) by corresponding sample quantile \( Y_{n-k,n} \) (\( k \)-th order statistic from above).

The obvious estimator of \( \theta \) is then

\[
\hat{\theta}_{\frac{k}{n}} := \frac{1}{n} \sum_{i=1}^{n} X_i 1_{\{y_i > y_{n-k,n}\}} = \frac{1}{k} \sum_{i=1}^{n} X_i 1_{\{y_i > y_{n-k,n}\}}.
\]
First result:

Under some strengthening of our conditions (relating to $R$ and to the sequence $k(n)$)

$$\sqrt{k}\left(\frac{\hat{\theta}_n^k}{\theta_n^k} - 1\right) \xrightarrow{d} \Theta,$$

a normal random variable that we describe now.
Background of limit result is our assumption

\[ \lim_{p \downarrow 0} \frac{1}{p} P \left\{ 1 - F_1(X) < \frac{p}{x}, 1 - F_2(Y) < \frac{p}{y} \right\} = R(x, y) . \]

Now define \( V := 1 - F_1(X) \)

\( W := 1 - F_2(Y) . \)

\( V \) and \( W \) have a uniform distribution, their joint distribution is a \textit{copula}. 
Now consider the i.i.d. r.v.’s

\[(V_i, W_i) = 1 - F_1(X_i), 1 - F_2(Y_i) \quad (i \leq n).\]

Empirical distribution function:

\[
\frac{1}{n} \sum_{i=1}^{n} 1_{\{V_i \leq x, W_i \leq y\}}
\]

We consider the lower tail of \((V_i, W_i)\) i.e., the higher tail for \((X_i, Y_i)\).

That is why we replace \((x, y)\) by \(\left(\frac{1}{x}, \frac{1}{y}\right)\) and for \(x, y > 0\) define the tail version

\[
T_n(x, y) := \frac{1}{k} \sum_{i=1}^{n} 1_{\left\{\frac{V_i}{nx} \leq \frac{k}{nx}, \frac{W_i}{ny} \leq \frac{k}{ny}\right\}}
\]
Now $T_n(X, Y)$ is close to its mean which is

$$\frac{n}{k} P \left\{ 1 - F_1(X) \leq \frac{k}{nx}, 1 - F_2(Y) \leq \frac{k}{ny} \right\}$$

and this is close to $R(x, y)$.

“Hence” $T_n(x, y) \xrightarrow{p} R(x, y)$ and – even better –

$$\sqrt{k} \left( T_n(x, y) - R(x, y) \right)$$

converges in distribution to a mean zero Gaussian process $W_R(x, y)$ (in $D$- space).
This stochastic process $W_R(x,y)$ has independent increments that is,

$$E W_R(x_1, y_1)W_R(x_2, y_2) = R(x_1 \wedge x_2, y_1 \wedge y_2)$$

and in particular

$$Var W_R(x,y) = R(x,y).$$
Formulated in a different way:

- Index the process by intervals:
  \[ \tilde{W}_R ((0,x) \times (0,y)) := W_R (x, y) \]

Then for two intervals \( I_1 \) and \( I_2 \)

\[ E \tilde{W}_R (I_1) \tilde{W}_R (I_2) = R (I_1 \cap I_2) . \]

(abuse of notation)
Hence $W_R$ is the direct analogue of Brownian motion in 2-dimensional space.
How do we use this convergence for $\hat{\theta}_{k/n}$?

$$\int_{0}^{\infty} T_n(x,1) \, dx^{-\gamma_1} = \frac{1}{k} \sum_{i=1}^{n} \int_{0}^{\infty} 1_{X_i > U_1\left(\frac{n}{k}\right), Y_i > U_2\left(\frac{n}{k}\right)} \, dx^{-\gamma_1}$$

$$\approx \frac{1}{k} \sum_{i=1}^{n} \int_{0}^{\infty} 1_{X_i > x^{-\gamma_1}U_1\left(\frac{n}{k}\right), Y_i > U_2\left(\frac{n}{k}\right)} \, dx^{-\gamma_1}$$

$$= \frac{1}{k} \sum_{i=1}^{n} \int_{0}^{\infty} 1_{X_i > x \ U_1\left(\frac{n}{k}\right), Y_i > U_2\left(\frac{n}{k}\right)} \, dx$$
\[
\hat{\Theta}_{k/n} = \frac{1}{k} \sum_{i=1}^{n} \frac{X_i}{U_1\left(\frac{n}{k}\right)} 1\{Y_i > U_2\left(\frac{n}{k}\right)\}
\]

“\approx” since: \( U_1 \in \text{R.V.} \Rightarrow \frac{X_{n-k,n}^p}{U_1\left(\frac{n}{k}\right)} \rightarrow 1 \).
Hence
\[
\frac{\sqrt{k}}{U_1 \left( \frac{n}{k} \right)} \left( \hat{\theta}_{\frac{k}{n}} - \theta_{\frac{k}{n}} \right) \approx \sqrt{k} \int_0^\infty (T_n(x,1) - R(x,1)) \, dx^{-\gamma_1}
\]
and we get
\[
\sqrt{k} \left( \frac{\hat{\theta}_{\frac{k}{n}}}{\theta_{\frac{k}{n}}} - 1 \right)^d \rightarrow (\gamma_1 - 1) W_R(\infty,1) + \left( \int_0^\infty R(s,1) \, ds^{-\gamma_1} \right)^{-1} \int_0^\infty W_R(s,1) \, ds^{-\gamma_1}
\]
a mean zero normally distributed random variable.
Last step: extrapolation from

\[ \theta_{\frac{k}{n}} \] (inside the sample)

to

\[ \theta_p \] (outside the sample).

Again we use the reasoning typical for extreme value theory.
Consider our first (non-statistical) result again:

\[
\lim_{p \downarrow 0} \frac{E\left( X \mid Y > U_2 \left( \frac{1}{p} \right) \right)}{U_1 \left( \frac{1}{p} \right)} = \int_0^\infty R \left( x^{-\frac{1}{n}}, 1 \right) dx
\]

In particular this holds for \( p = \frac{k}{n} \) i.e.

\[
\lim_{n \to \infty} \frac{E\left( X \mid Y > U_2 \left( \frac{n}{k} \right) \right)}{U_1 \left( \frac{n}{k} \right)} = \int_0^\infty R \left( x^{-\frac{1}{n}}, 1 \right) dx .
\]
Combine the two:

\[ \theta_p = E \left( X \left| Y > U_2 \left( \frac{1}{p} \right) \right. \right) \]

\[ \sim \frac{U_1 \left( \frac{1}{p} \right)}{U_1 \left( \frac{n}{k} \right)} \cdot E \left( X \left| Y > U_2 \left( \frac{n}{k} \right) \right. \right) = \frac{U_1 \left( \frac{1}{p} \right)}{U_1 \left( \frac{n}{k} \right)} \cdot \theta_{\frac{k}{n}}. \]
This leads to an estimate for $\theta_p$

$$\hat{\theta}_p := \frac{U_1\left(\frac{1}{p}\right)}{\hat{\theta}_{\frac{k}{n}}}$$

$$U_1\left(\frac{n}{k}\right)$$

Here $\hat{\theta}_{\frac{k}{n}}$ is the estimator we discussed before and

$$U_1\left(\frac{n}{k}\right) = X_{n-k,n}$$

as before.
It remains to define and to study $U_{1}\left(\frac{1}{p}\right)$ with

$$p = p_{n} \leq \frac{1}{n} \text{ as } n \to \infty.$$ 

Now $U_{1}\left(\frac{1}{p}\right)$ is a one-dimensional object (only connected with $X$, not $Y$). Such quantile is beyond the scope of the sample.

Recall our condition: $P\{X > t\} \in R.V.$

which implies
\[
\lim_{t \to \infty} \frac{U_1(tx)}{U_1(t)} = x^{\gamma_1}.
\]

Hence for large \( t \) and (say) \( x > 1 \)

\[
U_1(tx) \approx U_1(t) \cdot x^{\gamma_1}
\]

We use this relation with

\[ t \quad \text{replaced by} \quad \frac{n}{k} \]

\[ tx \quad \text{replaced by} \quad \frac{1}{p} \]
Then \( x = k/(np) \). We get
\[
U_1\left(\frac{1}{p}\right) \approx U_1\left(\frac{n}{k}\right)\left(\frac{k}{np}\right)^{\gamma_1}.
\]

This suggests the estimator for \( U_1\left(\frac{1}{p}\right) \):
\[
\hat{U}_1\left(\frac{1}{p}\right) = U_1\left(\frac{n}{k}\right)\left(\frac{k}{np}\right)^{\hat{\gamma}_1} = X_{n-k,n}\left(\frac{k}{np}\right)^{\hat{\gamma}_1}
\]
where \( \hat{\gamma}_1 \) is an estimator for \( \gamma_1 \).
Since $\gamma_1 > 0$ we use the well-known Hill estimator:

$$\hat{\gamma}_1 := \frac{1}{k_1} \sum_{i=0}^{k_1-1} \log X_{n-i,n} - \log X_{n-k_1,n}.$$

Property of Hill’s estimator:

$$\sqrt{k_1} \left( \hat{\gamma}_1 - \gamma_1 \right)^d \rightarrow_{d} \gamma_1 N_1 \quad (N_1 \text{ standard normal})$$

($k_1$ may differ from $k$ but satisfies similar conditions)
Property of $X_{n-k_1,n}$:

$$\sqrt{k_1} \left( \frac{X_{n-k_1,n}}{U\left(\frac{n}{k_1}\right)} - 1 \right) \xrightarrow{d} N_0 \quad \text{(standard normal)}$$

($N_0$ and $N_1$ are independent).

Combine the two relations:
\begin{align*}
\frac{U_1\left(\frac{1}{p}\right)}{U_1\left(\frac{1}{p}\right)} &= \frac{X_{n-k_1,n} U_1\left(\frac{n}{k_1}\right)}{U_1\left(\frac{n}{k_1}\right) U_1\left(\frac{1}{p}\right)} \left(\frac{k_1}{np}\right)^{\hat{\gamma}_1} \\
\text{subject to } U_1 &\in \mathcal{R}, V.
\end{align*}

\begin{align*}
\approx & \quad \frac{X_{n-k_1,n}}{U_1\left(\frac{n}{k_1}\right)} \left(\frac{np}{k_1}\right)^{\hat{\gamma}_1} \left(\frac{k_1}{np}\right)^{\hat{\gamma}_1} = \frac{X_{n-k_1,n}}{U_1\left(\frac{n}{k_1}\right)} \left(\frac{k_1}{np}\right)^{\hat{\gamma}_1-\gamma_1} \\
\approx & \quad \left(1 + \frac{N_0}{\sqrt{k_1}}\right) \exp\left\{\sqrt{k_1} \left(\hat{\gamma}_1 - \gamma_1\right) \log \frac{k_1}{np} \right\}.
\end{align*}
Now assume that
\[
\log \frac{k_1}{np} \sqrt{k_1} \to 0 \quad (n \to \infty)
\]
(this means that \( p \) can not be too small).

Then (expansion of function “exp’’)
\[
\frac{U_1 \left( \frac{1}{p} \right)}{U_1 \left( \frac{1}{p} \right)} \approx \left( 1 + \frac{N_0}{\sqrt{k_1}} \right) \left( 1 + \sqrt{k_1} \left( \hat{\gamma} - \gamma \right) \right) \log \frac{k_1}{np} \sqrt{k_1}
\]

and hence
\[
\frac{\sqrt{k_1}}{\log \frac{k_1}{np}} \left( \frac{U_1 \left( \frac{1}{p} \right)}{U_1 \left( \frac{1}{p} \right)} - 1 \right) \xrightarrow{d} \gamma N_1
\]

(i.e. asymptotically normal).

Final result:
Conditions

➢ Suppose $\gamma_1 \in (0,1/2)$ and $X > 0$.

➢ Assume $d_n := \frac{k}{np} \geq 1$ and $\lim_{n \to \infty} \frac{\log d_n}{\sqrt{k_1}} = 0$.

Denote $r := \lim_{n \to \infty} \frac{\sqrt{k} \log d_n}{\sqrt{k_1}} \in [0, \infty]$. Then as $n \to \infty$,

$$\min\left(\sqrt{k}, \frac{\sqrt{k_1}}{\log d_n} \left(\frac{\hat{\Theta}}{\Theta} - 1\right)\right) \xrightarrow{d} \begin{cases} \Theta + r \gamma N_1, & \text{if } r \leq 1; \\ \frac{1}{r} - \Theta + \gamma N_1, & \text{if } r > 1. \end{cases}$$

Corner cases are $r = 0$ and $r = +\infty$. 
So far we assumed $X > 0$.

For general $X \in \mathbb{R}$ we need some extra conditions:

1. Thinner left tail: $E\left|\min(X,0)\right|^\frac{1}{\gamma} < \infty$.

2. A further bound on $p = p_n$.

Then the left tail can be ignored.
Estimator in case $X \in \mathbb{R}$:

$$\hat{\theta}_p := \left( \frac{k}{np} \right)^{\frac{1}{\hat{\gamma}_1}} \frac{1}{k} \sum_{i=1}^{n} X_{i}^{1_{\{X_i > 0, Y_i > Y_{n-k,n}\}}}.$$ 

Has same behaviour as in case $X > 0$. 
**Simulation** setup:

- Transformed Cauchy distribution on $(0, \infty)^2$:
  
  Take $(Z_1, Z_2)$ standard Cauchy on $\mathbb{R}^2$ and define
  
  $$ (X, Y):= \left( \left| Z_1 \right|^\frac{2}{3}, \left| Z_2 \right| \right) $$

- Student-\(t_3\) distribution on $(0, \infty)^2$.

- With $(Z_1, Z_2)$ as before

  $$ (X, Y) = \left( \left( \max(0, Z_1) \right)^\frac{2}{3} + \left| \min(0, Z_1) \right|^\frac{1}{3}, \max(0, Z_2) + \left| \min(0, Z_2) \right|^\frac{1}{3} \right) $$
Table 1: Standardized mean and standard deviation of $\log \frac{\hat{\theta}_p}{\theta_p}$

<table>
<thead>
<tr>
<th></th>
<th>$n = 2,000$</th>
<th>$n = 5,000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1/2,000$</td>
<td>0.152 (1.027)</td>
<td>0.107 (1.054)</td>
</tr>
<tr>
<td>Transformed Cauchy distribution (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student-$t_3$ distribution</td>
<td>0.232 (0.929)</td>
<td>0.148 (0.964)</td>
</tr>
<tr>
<td>Transformed Cauchy distribution (2)</td>
<td>-0.147 (1.002)</td>
<td>-0.070 (1.002)</td>
</tr>
</tbody>
</table>

The numbers are the standardized mean of $\log \frac{\hat{\theta}_p}{\theta_p}$ and between brackets, the ratio of the sample standard deviation and the real standard deviation based on 500 estimates with $n = 2,000$ or 5,000 and $p = 1/n$. 
Application

Three investments banks:

Goldman Sachs (GS), Morgan Stanley (MS), and T. Rowe Price (TROW).


Data (Y): same for market index NYSE + AMES + Nasdaq.
Hill Estimator of $\gamma_1$

$\hat{\gamma}_1$

$k_1$

GS
MS
TROW
Table 2: MES of the three investment banks

<table>
<thead>
<tr>
<th>Bank</th>
<th>$\gamma_1$</th>
<th>$\hat{\theta}_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldman Sachs (GS)</td>
<td>0.386</td>
<td>0.301</td>
</tr>
<tr>
<td>Morgan Stanley (MS)</td>
<td>0.473</td>
<td>0.593</td>
</tr>
<tr>
<td>T. Rowe Price (TROW)</td>
<td>0.379</td>
<td>0.312</td>
</tr>
</tbody>
</table>

Here $\hat{\gamma}_1$ is computed by taking the average of the Hill estimates for $k_1 \in [70, 90]$. $\hat{\theta}_p$ is given as before, with $n = 2513$, $k = 50$ and $p = 1/n = 1/2513$. 
Interpretation table 2:

$$\hat{\theta}_p = 0.301 \text{ (Goldman Sachs)}$$

Hence in a once-per-decade market crisis the expected loss in log return terms is 30% (perhaps about 26% in equity prices)
References


The 8th Conference on Extreme Value Analysis
July 8-12, 2013
Fudan University, Shanghai, China
We are pleased to announce that the 8th Conference on Extreme Value Analysis will take place from July 8 to 12, 2013 at Fudan University, Shanghai, China.

Organizers: Deyuan Li, Liang Peng, Zhengjun Zhang, Ming Zheng

Email: eva2013sh@yahoo.com
Website: http://eva.fudan.edu.cn

Topics:
- Univariate, multivariate, infinite dimensional extreme value theory
- Order statistics and records
- Rare events and risk analysis
- Spatial/spatio-temporal extremes
- Heavy tails in actuarial sciences
- Other related applications