Extreme Values and Their Applications in Finance

Ruey S. Tsay
Booth School of Business
University of Chicago

RMI, NUS, October 2009
Classification of Financial Risk:

1. Credit risk
2. Market risk
3. Operational risk

We focus on the market risk, because
- more data are available
- easier to understand
- the ideas applicable to other types of risk, e.g, RiskMetrics ⇒ CreditMetrics.
What is Value at Risk (VaR)?

- a measure of minimum loss of an asset or
- amount a financial position could decline

in a given period, associated with a given probability.

**A formal definition:**

- a future time period: \( \Delta t = \ell \)
- change in value \( \Delta V(\ell) \) or loss \( L(\ell) \)
- CDF of the loss function \( L(\ell): F_{\ell}(x) \)
- given (tail) probability: \( p \)

\[
p = P[L(\ell) \geq \text{VaR}] = 1 - P[L(\ell) \leq \text{VaR}] = 1 - F_{\ell}(\text{VaR}).
\]
A probabilistic view

Quantile: For a continuous distribution, the $q$th quantile $x_q$ is defined as $q = F_\ell(x_q)$. VaR is the $(1 - p)$th quantile of loss dist.

**Factors affect VaR:**

1. the tail probability $p$.  
2. the time horizon $\ell$.  
3. the CDF $F_\ell(x)$ of the loss function  
4. the mark-to-market value of the position.

Shall use log returns because log returns $\approx$ percentage changes.  
$\text{VaR} = \text{Value} \times (\text{VaR of log return})$. 
Remark on use of VaR

- The ideal situation is that we know the distribution of the loss function.
- VaR is just a percentile of the loss distribution. It cannot describe fully the potential loss.
- In addition, VaR is a point estimate and contains uncertainty. Should not overlook its uncertainty in applications.
Consider the log return $r_t$. For a long position, loss occurs when $r_t$ is negative. For a short position, loss occurs when $r_t$ is positive.

Since our discussion focuses on the upper tail of the loss function, we shall use negative return, i.e. $-r_t$, in data analysis for a long position.
Methods available for computing VaR

1. RiskMetrics
2. Econometric modeling
3. Empirical quantile
4. Traditional extreme value theory (EVT)
5. EVT based on exceedance over a high threshold (POT)
6. Conditional POT
Empirical demonstration

Data used in illustrations:
Daily log returns of IBM stock
- span: July 3, 62 to Dec. 31, 98.
- size: 9190 points
Position: long on the stock for $10 million.
<table>
<thead>
<tr>
<th>Year</th>
<th>Log Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>-0.2</td>
</tr>
<tr>
<td>1980</td>
<td>-0.1</td>
</tr>
<tr>
<td>1990</td>
<td>0.0</td>
</tr>
<tr>
<td>2000</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Figure: Daily log returns of IBM stock from July 1962 to December 1998

Ruey S. Tsay
Booth School of Business
University of Chicago

Extreme Values and Their Applications in Finance
Developed by J.P. Morgan

- \( r_t \) given \( F_{t-1} \): \( N(0, \sigma^2_t) \)
- \( \sigma^2_t \) follows the special IGARCH(1,1) model

\[
\sigma^2_t = \alpha \sigma^2_{t-1} + (1 - \alpha) r^2_{t-1}, \quad 1 > \alpha > 0.
\]

- VaR = 2.326\( \sigma_t \) if \( p = 0.01 \).
- \( k \)-horizon: \( \text{VaR}[k] = \sqrt{k} \text{VaR} \)
  
  The **square root of time rule**. It depends on two key assumptions. Normality & zero mean.

- Pros: simplicity and transparency
- Cons: model is not adequate
Illustration: IBM data

A special IGARCH(1,1) model:

\[
\begin{align*}
    r_t &= a_t, \quad a_t = \sigma_t \epsilon_t, \\
    \sigma_t^2 &= 0.9396 \sigma_{t-1}^2 + (1 - 0.9396) a_{t-1}^2
\end{align*}
\]

Because \( r_{9190} = 0.0128 \) and \( \hat{\sigma}_{9190}^2 = 0.0003472 \), \( \hat{\sigma}_{9190}^2(1) = 0.000336 \).

For \( p = 0.01 \), VaR of \( r_t = 2.326 \times \sqrt{0.000336} = 0.04265 \), and VaR = $426,500.
Econometric models

Setup:

- $r_t = \mu_t + a_t$ given $F_{t-1}$
- $\mu_t$: a mean equation (Ch. 2)
- $\sigma^2_t$: a volatility model (Ch. 3 or 4)
- Pros: sound theoretical justifications.
- Cons: a bit complicated & improvement might be limited.

Illustration: IBM data

**Case 1: Gaussian**

\[
r_t = -0.00066 - 0.0247r_{t-2} + a_t, \quad a_t = \sigma_t \epsilon_t
\]

\[
\sigma_t^2 = 0.00000389 + 0.0799a_{t-1}^2 + 0.9073\sigma_t^2.
\]

From \(r_{9189} = 0.00201, r_{9190} = 0.0128\) and \(\sigma_{9190}^2 = 0.00033455\), we have

\[
\hat{r}_{9190}(1) = -0.00071 \quad \text{and} \quad \hat{\sigma}_{9190}^2(1) = 0.0003211.
\]

If \(p = 0.01\), then the quantile is

\[
-0.00071 + 2.3262 \times \sqrt{0.0003211} = 0.0409738.
\]

\[\text{VaR} = \$409,738.\]
Case 2: Student-\(t_5\)

\[
    r_t = -0.0003 - 0.0335r_{t-2} + a_t, \quad a_t = \sigma_t \epsilon_t
\]

\[
    \sigma_t^2 = 0.000003 + 0.0559a_{t-1}^2 + 0.9350\sigma_{t-1}^2.
\]

From the data, \(r_{9189} = 0.00201\), \(r_{9190} = 0.0128\) and \(\sigma_{9190}^2 = 0.000349\), we have

\[
    \hat{r}_{9190}(1) = -0.000367 \quad \text{and} \quad \hat{\sigma}_{9190}^2(1) = 0.0003386.
\]

If \(p = 0.01\), the quantile is

\[
    -0.000367 + \left(3.3649/\sqrt{5/3}\right)\sqrt{0.0003386} = 0.0475943.
\]

\(\text{VaR} = \$475,943\).
Effects of heavy-tails seen with $p = 0.01$.

Multiple step-ahead forecasts are needed.

**Example 7.3 (continued).** 15-day horizon.

$\hat{r}_{9190}[15] = -0.00998$ and $\sigma_t[15] = 0.0047948$.

If $p = 0.05$, the quantile is $-0.00998 + 1.6449\sqrt{0.0047948} = 0.1039191$.

15-day VaR = $10,000,000 \times 0.1039191 = $1,039,191.

RiskMetrics: VaR = $287,700 \times \sqrt{15} = $1,114,257.
Empirical quantile

Sample of log returns: \( \{r_t \mid t = 1, \ldots, n\} \).
Order statistics:

\[ r(1) \leq r(2) \leq \cdots \leq r(n) \]

\( r(i) \) as the \( i \)th order statistic of the sample.
\( r(1) \) is the sample minimum
\( r(n) \) the sample maximum.

**Idea:** Use the empirical quantile to estimate the theoretical quantile of \( r_t \).
For a given tail probability \( p \), what is the empirical quantile?
If \( n(1 - p) = \ell \) is an integer, then it is \( r(\ell) \).
If \( n(1 - p) \) is not an integer, find the two neighboring integers \( \ell_1 < n(1 - p) < \ell_2 \) and use interpolation.
Let \( p_i = \ell_i/[n(1 - p)] \). The quantile is

\[
\hat{x}_p = \frac{p_2 - p}{p_2 - p_1} r(\ell_1) + \frac{p - p_1}{p_2 - p_1} r(\ell_2).
\]
Illustration: IBM data

\( n = 9190. \) If \( 1 - p = 0.05, \) then \( n(1 - p) = 8730.5. \)
5% quantile is \( (r_{(8730)} + r_{(8731)})/2 = 0.021603. \)
\( \text{VaR} = \$216,030. \)

If \( 1 - p = 0.01, \) then \( n(1 - p) = 9098.1 \) and the 1% quantile is
\[
\hat{x}_{0.99} = \frac{p_2 - 0.99}{p_2 - p_1} r_{(9098)} + \frac{0.99 - p_1}{p_2 - p_1} r_{(9099)} \\
= \frac{0.00001}{0.00011} (3.627) + \frac{0.0001}{0.00011} (3.657) \\
\approx 3.6303.
\]

\( \text{VaR} \) is \$363,030. \)
Focus on the upper tail behavior of $r_t$ (loss function). For iid random sample,

$$P(r_{(n)} \leq x) = P(r_1 \leq x, r_2 \leq x, \ldots, r_n \leq x)$$

$$= \prod_{i=1}^{n} P(r_i \leq x) \quad \text{(independence)}$$

$$= [P(r_1 \leq x)]^n \quad \text{(iden. dist.)}$$

$$= [F(x)]^n,$$

where $F(x)$ is the marginal CDF of $r_t$.

Consider $P[r_{(n)} \leq x] = [1 - (1 - F(x))]^n = [1 - \frac{1}{n}n(1 - F(x))]^n.$
A properly normalized \( r(n) \) assumes a special distribution:

\[
F_*(x) = \begin{cases} 
\exp[-(1 + \xi x)^{-1/\xi}] & \text{if } \xi \neq 0 \\
\exp[-\exp(x)] & \text{if } \xi = 0
\end{cases}
\]

for \( x < -1/\xi \) if \( \xi < 0 \) and for \( x > -1/\xi \) if \( \xi > 0 \).

\( \xi \): the \textit{shape parameter}

\( \alpha = 1/\xi \): tail index of the distribution.
Classification of distributions

- **Type I**: $\xi = 0$, the Gumbel family. The CDF is
  \[
  F_*(x) = \exp[-\exp(x)], \quad -\infty < x < \infty. \tag{1}
  \]

- **Type II**: $\xi > 0$, the Fréchet family. The CDF is
  \[
  F_*(x) = \begin{cases} 
  \exp[-(1 + \xi x)^{-1/\xi}] & \text{if } x > -1/\xi \\
  0 & \text{otherwise.} 
  \end{cases} \tag{2}
  \]

- **Type III**: $\xi < 0$, the Weibull family. The CDF here is
  \[
  F_*(x) = \begin{cases} 
  \exp[-(1 + \xi x)^{-1/\xi}] & \text{if } x < -1/\xi \\
  1 & \text{otherwise.} 
  \end{cases}
  \]

The pdf of the normalized minimum is

\[
  f_*(x) = \begin{cases} 
  (1 + \xi x)^{-1/\xi-1} \exp[-(1 + \xi x)^{-1/\xi}] & \text{if } \xi \neq 0 \\
  \exp[x - \exp(x)] & \text{if } \xi = 0
  \end{cases}
  \]

where $-\infty < x < \infty$ for $\xi = 0$, $x > -1/\xi$ for $\xi < 0$ and $x < -1/\xi$ for $\xi < 0$. 

Ruey S. Tsay
Booth School of Business
University of Chicago

Extreme Values and Their Applications in Finance
Traditional estimation

How to apply the EVT distribution to risk assessment?
If we know the three parameters, we can compute the quantiles!

Divide the sample into non-overlapping subsamples.
Suppose there are $T$ data points, we divide the data as

$$\{r_1, \cdots, r_n \mid r_{n+1}, \cdots, r_{2n} \mid r_{2n+1}, \cdots, r_{3n} \mid \cdots \mid r_{(g-1)n+1}, \cdots, r_{ng}\},$$

$n$: size of subgroup (assumed to be sufficiently large).

Idea: find the maximum of each subgroup. These maxima are the data used to estimate the three parameters.
Several estimation methods available. We use maximum likelihood estimates.
Illustration: IBM data

<table>
<thead>
<tr>
<th>$n$</th>
<th>$g$</th>
<th>Scale $\alpha_n$</th>
<th>Location $\beta_n$</th>
<th>Shape Par. $\xi_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td></td>
<td>Negative returns (long position)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>437</td>
<td>0.823(0.035)</td>
<td>1.902(0.044)</td>
<td>0.197(0.036)</td>
</tr>
<tr>
<td>63</td>
<td>145</td>
<td>0.945(0.077)</td>
<td>2.583(0.090)</td>
<td>0.335(0.076)</td>
</tr>
<tr>
<td>126</td>
<td>72</td>
<td>1.147(0.131)</td>
<td>3.141(0.153)</td>
<td>0.330(0.101)</td>
</tr>
<tr>
<td>252</td>
<td>36</td>
<td>1.542(0.242)</td>
<td>3.761(0.285)</td>
<td>0.322(0.127)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td>Returns (Short position)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>437</td>
<td>0.931(0.039)</td>
<td>2.184(0.050)</td>
<td>0.168(0.036)</td>
</tr>
<tr>
<td>63</td>
<td>145</td>
<td>1.157(0.087)</td>
<td>3.012(0.108)</td>
<td>0.217(0.066)</td>
</tr>
<tr>
<td>126</td>
<td>72</td>
<td>1.292(0.158)</td>
<td>3.471(0.181)</td>
<td>0.349(0.130)</td>
</tr>
<tr>
<td>252</td>
<td>36</td>
<td>1.624(0.271)</td>
<td>4.475(0.325)</td>
<td>0.264(0.186)</td>
</tr>
</tbody>
</table>

Use command `gev` of the package `evir` in R.
Use a two-step procedure, because of the division into subgroup. VaR for $r_t$:

$$\text{VaR} = \begin{cases} 
\beta_n - \frac{\alpha_n}{\xi_n} \left(1 - \left[-n \ln(1 - p)\right]^{-\xi_n}\right) & \text{if } \xi_n \neq 0 \\
\beta_n + \alpha_n \ln\left[-n \ln(1 - p)\right] & \text{if } \xi_n = 0.
\end{cases}$$
Focus on $\xi \neq 0$.

Let $p^*$ be a small upper tail prob and $r_n^*$ be the $(1 - p^*)$th quantile of the GEV for the subperiod maximum. Then,

$$1 - p^* = \exp \left[ - \left( 1 + \frac{\xi_n(r_n^* - \beta_n)}{\alpha_n} \right)^{-1/\xi_n} \right].$$

From which,

$$\ln(1 - p^*) = - \left[ 1 + \frac{\xi_n(r_n^* - \beta_n)}{\alpha_n} \right]^{-1/\xi_n},$$

$$r_n^* = \beta_n - \frac{\alpha_n}{\xi_n} \left\{ 1 - \left[ - \ln(1 - p^*) \right]^{-\xi_n} \right\}.$$
Connection subgroup maximum $r_{n,i}$ to returns $r_t$.

$$1 - p^* = P(r_{n,i} \leq r_n^*) = [P(r_t \leq r_n^*)]^n.$$  

For a given upper tail probability $p$ of $r_t$, the above relation gives $1 - p^* = (1 - p)^n$ so that $\ln(1 - p^*) = n \ln(1 - p)$. Consequently,

$$\text{VaR} = \beta_n - \frac{\alpha_n}{\xi_n} \left\{ 1 - \left[ -n \ln(1 - p) \right]^{-\xi_n} \right\}.$$
If \( n = 63 \) (quarterly maximum), then \( \hat{\alpha}_n = 0.945 \), \( \hat{\beta}_n = 2.583 \), and \( \hat{\xi}_n = 0.335 \). If \( p = 0.01 \), the VaR is

\[
\text{VaR} = 2.583 - \frac{0.945}{0.335} \left\{ 1 - [-63 \ln(1 - 0.01)]^{-0.335} \right\} = 3.04969
\]

VaR is $304,969.
If \( p = 0.05 \), then VaR is $166,641.
For \( n = 21 \), the results are:
\text{VaR} = $340,013 \text{ for } p = 0.01;
\text{VaR} = $184,127 \text{ for } p = 0.05.
Discussion

- Results depend on the choice of $n$
- VaR seems low, but it might be due to the choice of $p$.
  - If $p = 0.001$, then
    - $\text{VaR} = \$546,641$ for the Gaussian AR(2)-GARCH(1,1) model
    - $\text{VaR} = \$666,590$ for the extreme value theory with $n = 21$. 
Position = $10 million.

Higher tail probability. If $p = 0.05$, then

1. $302,500 for the RiskMetrics,
2. $287,200 for an AR(2)-GARCH(1,1) with normal
3. $283,520 for an AR(2)-GARCH(1,1) with $t_5$
4. $216,030 using the empirical quantile, and
5. $184,127 for EVT with $n = 21$. 
$p = 0.01$, then

1. $426,500$ for the RiskMetrics,
2. $409,738$ for an AR(2)-GARCH(1,1) model,
3. $475,943$ for an AR(2)-GARCH(1,1) model with $t_5$
4. $365,800$ for empirical quantile, and
5. $340,013$ for EVT with $n = 21$.

If $p = 0.001$, then

1. $566,443$ for the RiskMetrics,
2. $546,641$ for an AR(2)-GARCH(1,1) model,
3. $836,341$ for an AR(2)-GARCH(1,1) model with $t_5$
4. $780,712$ for empirical quantile, and
5. $666,590$ for EVT with $n = 21$. 
Multi-period VaR with EVT

$$\text{VaR}(\ell) = \ell^{1/\alpha}\text{VaR} = \ell^\xi\text{VaR}$$

where $\alpha$ is the tail index and $\xi$ is the shape parameter. For IBM data with $p = 0.05$ and $n = 21$,

$$\text{VaR}(30) = (30)^{0.197}\text{VaR} = 1.954 \times $184,127 = $359,841.$$
Another measure of risk based on the subgroup maximum. 
$L_{n,g} = \text{the level that is exceeded in one out of every } g \text{ non-overlapping subperiods of length } n$. That is,

$$P(r_{n,i} > L_{n,g}) = \frac{1}{g}.$$ 

From the extreme value distribution of maximum,

$$L_{n,g} = \beta_n - \frac{\alpha_n}{\xi_n} \{1 - [\ln(1 - \frac{1}{g})]^{-\xi_n}\}.$$ 

The corresponding subperiod is called a stressed period.
For (negative) IBM log returns with $n = 21$ and $g = 12$, we have

$m_1 = \text{gev}(\text{nibm}, \text{block}=21)$

$\text{rl.21.12} = \text{rlevel.gem}(m_1, \text{k.blocks}=12)$

$\text{rl.21.12}$

[1] 4.1779 4.4820 4.8581 (⇐ 95% C.I.)
Based on exceedances over a high threshold

Idea: frequency of big returns and their magnitudes are important. Rare events occurred in cluster, not random.

Statistical theory:
Two-dimensional Poisson process
Two possible cases:
(A) Homogeneous: parameters are fixed over time
(B) Non-homogeneous case: parameters are time-varying, depending on some explanatory variables.
Figure: Negative daily log returns of IBM stock from July 1962 to December 1998

Extreme Values and Their Applications in Finance
Homogeneous case

Select a high threshold $\eta$. Consider the returns that exceed $\eta$. Let $t_i$ be the $i$th exceeding time and $r_t - \eta$ the exceedance. This approach considers the condition distribution of $x = r - \eta > 0$ given $r > \eta$.

$$P(r \leq x + \eta | r > \eta) = \frac{P(\eta \leq r \leq x + \eta)}{P(r > \eta)} = \frac{P(r \leq x + \eta) - P(r \leq \eta)}{1 - P(r \leq \eta)}.$$ 

Using $e^{-y} \approx 1 - y$ and GEV, we obtain

$$P(r \leq x + \eta | r > \eta) \approx 1 - \left(1 + \frac{\xi x}{\alpha + \xi(\eta - \beta)}\right)^{-1/\xi}. \quad (3)$$
The CDF in the form

$$G(x) = 1 - \left[1 + \frac{\xi x}{\psi(\eta)}\right]^{-1/\xi}, \quad \xi \neq 0,$$

where $\psi(\eta) > 0$, is called a GPD. Thus, the exceedance $x$ conditional on $r > \eta$ follows a GPD with parameter $\xi$ and $\psi(\eta) = \alpha + \xi(\eta - \beta)$. An important property of GPD: if $x + \eta_0 | r > \eta_0 \sim GPD(\xi, \psi(\eta_0))$, then $x + \eta | r > \eta \sim GPD(\xi, \psi(\eta))$ for any $\eta > \eta_0$, where $\psi(\eta) = \psi(\eta_0) + \xi(\eta - \eta_0)$. Tail index unchanged.
Mean excess function

The threshold plays an important role in the POT approach. If \( x = r - \eta_o > 0 \) given \( r > \eta_o \) is \( \text{GPD}(\xi, \psi(\eta_o)) \), then the mean excess over the threshold \( \eta_o \) is \((\xi < 1)\)

\[
E(r - \eta_o | r > \eta_o) = \frac{\psi(\eta_o)}{1 - \xi}.
\]

For any \( \eta > \eta_o \), define the mean excess function \( e(\eta) \) as

\[
e(\eta) = E(r - \eta | r > \eta) = \frac{\psi(\eta_o) + \xi(\eta - \eta_o)}{1 - \xi}.
\]

Thus, for \( y > 0 \),

\[
e(\eta_o + y) = E[r - (\eta_o + y) | r > \eta_o + y] = \frac{\psi(\eta_o) + \xi y}{1 - \xi}.
\]

The mean excess function is a linear function of \( y = \eta - \eta_o \).
A graphical method to infer the choice of threshold.

Define

\[ e_T(\eta) = \frac{1}{N_\eta} \sum_{i=1}^{N_\eta} (r_{t_i} - \eta), \]

where \( N_\eta \) is the number of returns that exceed \( \eta \).

Scatter plot of \( e_T(\eta) \) against \( \eta \) should show a linear function in \( \eta \) for \( \eta > \eta_0 \).

The command is `meplot` in `evir`. 
Two-dimensional Poisson process

For a given threshold $\eta$, consider $(t_i, r_{t_i})$ jointly. Treat it as a 2-dimensional Poisson process with intensity measure

$$\Lambda[(D_2, D_1) \times (r, \infty)] = \frac{D_2 - D_1}{D} S(r; \xi, \alpha, \beta),$$

where $D$ is the baseline time interval, e.g. $D = 252$, and

$$S(r; \xi, \alpha, \beta) = \left[1 + \frac{\xi(r - \beta)}{\alpha}\right]^{-1/\xi},$$

where it is understood that $1 + \xi(r - \beta)/\alpha > 0$. Intensity $\propto$ length of time interval $\times$ the survival function of GEV.
From the intensity measure,

\[
\frac{\Lambda[(0, D) \times (x + \eta, \infty)]}{\Lambda[(0, D) \times (\eta, \infty)\]} = \left[1 + \frac{\xi x}{\alpha + \xi(\eta - \beta)} \right]^{-1/\xi},
\]

which is precisely the survival function of the conditional distribution in Eq. (3).

**Intensity function:**

\[
\Lambda[(D_2, D_1) \times (r, \infty)] = \int_{D_1}^{D_2} \int_r^\infty \lambda(t, z; \xi, \alpha, \beta) dz dt,
\]

\[
\lambda(t, z; \xi, \alpha, \beta) = \frac{1}{D} g(z; \xi, \alpha, \beta) = \frac{1}{\alpha} \left[1 + \frac{\xi(z - \beta)}{\alpha} \right]^{-(1+\xi)/\xi}.
\]
Likelihood function

\[ L(\xi, \alpha, \beta) = \left( \prod_{i=1}^{N_{\eta}} \frac{1}{D} g(r_t; \xi, \alpha, \beta) \right) \times \exp \left[ -\frac{T}{D} S(\eta; \xi, \alpha, \beta) \right]. \]

For a given threshold \( \eta \), the parameters \( \xi, \alpha, \beta \) can be estimated by maximizing the log likelihood function.
Estimation results of IBM returns

<table>
<thead>
<tr>
<th>Thr.</th>
<th>Exc.</th>
<th>Shape Par. $\xi$</th>
<th>Log(Scale) $\ln(\alpha)$</th>
<th>Location $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Original log returns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0%</td>
<td>175</td>
<td>0.307(.090)</td>
<td>0.307(.124)</td>
<td>4.692(.191)</td>
</tr>
<tr>
<td>2.5%</td>
<td>310</td>
<td>0.264(.065)</td>
<td>0.315(.113)</td>
<td>4.741(.180)</td>
</tr>
<tr>
<td>2.0%</td>
<td>554</td>
<td>0.188(.044)</td>
<td>0.277(.099)</td>
<td>4.810(.172)</td>
</tr>
<tr>
<td>(b) Sample mean removed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0%</td>
<td>184</td>
<td>0.305(.088)</td>
<td>0.308(.124)</td>
<td>4.738(.192)</td>
</tr>
<tr>
<td>2.5%</td>
<td>334</td>
<td>0.282(.067)</td>
<td>0.320(.121)</td>
<td>4.768(.185)</td>
</tr>
<tr>
<td>2.0%</td>
<td>590</td>
<td>0.193(.044)</td>
<td>0.279(.099)</td>
<td>4.849(.173)</td>
</tr>
</tbody>
</table>

Ruey S. Tsay Booth School of Business University of Chicago
Extreme Values and Their Applications in Finance
Since GDP and GEV have the same parameters, we can make use of the results obtained before.

\[
\text{VaR} = \begin{cases} 
\beta - \frac{\alpha}{\xi} \left\{ 1 - \left[ -D \ln(1 - p) \right]^{-\xi} \right\} & \text{if } \xi \neq 0, \\
\beta - \alpha \ln \left[ -D \ln(1 - p) \right] & \text{if } \xi = 0,
\end{cases}
\]

(4)

where \( D \) is the baseline time interval used in estimation. We use \( D = 252 \) for U.S. data.
Illustration: IBM data

Case I: Use the original daily log returns.
1. \( \eta = 3.0\%: \text{VaR}(5\%) = $228,239, \text{VaR}(1\%) = $359.303. \)
2. \( \eta = 2.5\%: \text{VaR}(5\%) = $219,106, \text{VaR}(1\%) = $361,119. \)
3. \( \eta = 2.0\%: \text{VaR}(5\%) = $212,981, \text{VaR}(1\%) = $368.552. \)

Case II: Sample mean removed.
1. \( \eta = 3.0\%: \text{VaR}(5\%) = $232,094, \text{VaR}(1\%) = $363,697. \)
2. \( \eta = 2.5\%: \text{VaR}(5\%) = $225,782, \text{VaR}(1\%) = $364,254. \)
3. \( \eta = 2.0\%: \text{VaR}(5\%) = $217,740, \text{VaR}(1\%) = $372,372. \)

The resulting VaR is more stable than that based on subgroup maxima.
In `evir`, the homogeneous POT approach can be estimated by the command `pot`.

Demonstration:
\[
m3 = \text{pot}(\text{nibm}, 2.5)
\]
\[
m3
\]
\[
\text{plot}(m3)
\]
\[
\text{riskmeasures}(m3, c(0.95, 0.99, 0.999))
\]
Alternative parameterization

Direct use of GPD with CDF

\[
G(x) = 1 - \left[1 + \frac{\xi x}{\beta}\right]^{-1/\xi}, \quad x > 0,
\]

where \( \beta = \psi(\eta) \) defined before. The new \( \beta \) is referred to as the scale parameter.

Thus, it uses two, instead of three, parameters. The pdf is

\[
f(x) = \frac{1}{\beta} \left[1 + \frac{\xi x}{\beta}\right]^{-(1/\xi+1)}.
\]

For a given threshold, one can estimate \( \xi \) and \( \beta \) directly via the maximum likelihood method of the above pdf.

This is the approach used in \texttt{evir} package. The command is \texttt{gpd}.
Let $y = x + \eta$ for a given threshold $\eta$. What we have is 
\[
P(r \leq y | r > \eta) \approx G(x).
\]
Thus, 
\[
\frac{F(y) - F(\eta)}{1 - F(\eta)} \approx G(x).
\]
If we estimate $F(\eta)$ by $\hat{F}(\eta) = \frac{T - N_\eta}{T}$, the sample fraction of less than the threshold. Then, 
\[
F(y) = F(\eta) + G(x)[1 - F(\eta)] = 1 - [1 - F(\eta)] + G(x)[1 - F(\eta)] \\
= 1 - [1 - F(\eta)][1 - G(x)] \approx 1 - \frac{N_\eta}{T} \left[1 + \frac{\xi(y - \eta)}{\beta}\right]^{-1/\xi}.
\]
For a upper tail probability $p$, letting $1 - p = F(y)$, we have 
\[
\text{VaR} = \eta - \frac{\beta}{\xi} \left[1 - \left(\frac{Tp}{N_\eta}\right)^{-\xi}\right].
\]
Expected shortfall

**Definition:** Expected loss given that the VaR is exceeded. That is, expected loss when the anticipated rare event occurred.

\[ \text{ES} = E(r|r > \text{VaR}) = \text{VaR} + E(r - \text{VaR}|r > \text{VaR}). \]

Using GPD, we have

\[ E(r - \text{VaR}|r > \text{VaR}) = \frac{\beta + \xi(V\text{aR} - \eta)}{1 - \xi}, \]

provided \(0 < \xi < 1\). Consequently,

\[ \text{ES} = \frac{\text{VaR}}{1 - \xi} + \frac{\beta - \xi\eta}{1 - \xi}. \]

In **evir**, VaR and ES can be calculated by the command `riskmeasures`. 
The GPD approach can be carried out in **evir** by the command `gpd` and `riskmeasures`.

Demonstration:
```
mgpd = gpd(nibm,threshold=2.5)
mgpd
par(mfcol=c(2,2))
plot(mgpd)
plot(mgpd)
riskmeasures(mgpd,c(.95,.99,.999))
```
Non-homogeneous case

Use time-varying parameters, i.e. parameters depend on some explanatory variables.
For instance, consider the POT approach. Use

\[ \xi_t = \gamma_0 + \gamma_1 x_{1t} + \cdots + \gamma_v x_{vt} \equiv \gamma_0 + \gamma' x_t \]
\[ \ln(\alpha_t) = \delta_0 + \delta_1 x_{1t} + \cdots + \delta_v x_{vt} \equiv \delta_0 + \delta' x_t \]
\[ \beta_t = \theta_0 + \theta_1 x_{1t} + \cdots + \theta_v x_{vt} \equiv \theta_0 + \theta' x_t. \]

For IBM data, explanatory variables include past volatilities, etc. See Chapter 7 of Tsay (2005) [or handout] for more details and estimation results.
Poisson process → time durations between two consecutive events are independent and exponentially distributed.
Define $z_{t_i} = \frac{1}{D} \sum_{t=t_{i-1}+1}^{t_i} S(\eta; \xi_t, \alpha_t, \beta_t)$. Perform QQ-plot of $z_{t_i}$ against standard exponential dist.

Relation between GPD $X$ and standard exponential dist.

$$w_{t_i} = \frac{1}{\xi_{t_i}} \ln \left(1 + \xi_{t_i} \frac{r_{t_i} - \eta}{\psi_{t_i}} \right)_{+}.$$

For GPD, $\{w_{t_i}\}$ are iid exponential with mean 1.
Some explanatory variables:

- $x_{1t}$: indicator for October, November and December
- $x_{2t}$: indicator the the behavior of $r_{t-1}$, i.e. $x_{2t} = 1$ if and only if $r_{t-1} < 2.5\%$.
- $x_{3t}$: number of days between $t - 1$ and $t - 5$ such that $|r_{t-i}| \geq 2.5\%$.
- $x_{4t}$: Annual trend defined as $(\text{year of time } t - 1961)/38$.
- $x_{5t}$: volatility based on a Gaussian GARCH(1,1) model for $r_t$. 
### Illustration: estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>constant</th>
<th>Coef. of $x_{3t}$</th>
<th>Coef. of $x_{4t}$</th>
<th>Coef. of $x_{5t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Threshold 2.5% with 334 Exceedances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_t$ (Std.err)</td>
<td>0.3202 (0.3387)</td>
<td>1.4772 (0.3222)</td>
<td>2.1991 (0.2450)</td>
<td></td>
</tr>
<tr>
<td>$\ln(\alpha_t)$ (Std.err)</td>
<td>$-0.8119$ (0.1798)</td>
<td>0.3305 (0.0826)</td>
<td>1.0324 (0.2619)</td>
<td></td>
</tr>
<tr>
<td>$\xi_t$ (Std.err)</td>
<td>0.1805 (0.1290)</td>
<td>0.2118 (0.0580)</td>
<td>0.3551 (0.1503)</td>
<td>$-0.2602$ (0.0461)</td>
</tr>
<tr>
<td><strong>Threshold 3.0% with 184 Exceedances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_t$ (Std.err)</td>
<td>1.1569 (0.4082)</td>
<td></td>
<td>2.1918 (0.2909)</td>
<td></td>
</tr>
<tr>
<td>$\ln(\alpha_t)$ (Std.err)</td>
<td>$-0.0316$ (0.1201)</td>
<td>0.3336 (0.0861)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi_t$ (Std.err)</td>
<td>0.6008 (0.1454)</td>
<td>0.2480 (0.0731)</td>
<td>$-0.3175$ (0.0685)</td>
<td></td>
</tr>
</tbody>
</table>
Illustration: comparison 1

(a) Homogeneous: $z$

(b) Homogeneous: $w$

(c) Inhomogeneous: $z$

(d) Inhomogeneous: $w$
Illustration: comparison 2

(a) Homogeneous: z

(b) Homogeneous: w

(c) Inhomogeneous: z

(d) Inhomogeneous: w

Ruey S. Tsay
Booth School of Business
University of Chicago

Extreme Values and Their Applications in Finance
For December 31, 1998, we have $x_{3,9190} = 0$, $x_{4,9190} = 0.9737$ and $x_{5,9190} = 1.9766$. The parameters become:

\[ \xi_{9190} = 0.01195, \quad \ln(\alpha_{9190}) = 0.19331, \quad \beta_{9190} = 6.105. \]

If $p = 0.05$, then quantile = 3.03756\% and

\[ \text{VaR} = 10,000,000 \times 0.0303756 = 303,756. \]

If $p = 0.01$, then Var is $497,425$. 
For December 30, 1998, we have $x_{3,9189} = 1$, $x_{4,9189} = 0.9737$ and $x_{5,9189} = 1.8757$ and

$$\xi_{9189} = 0.2500, \quad \ln(\alpha_{9189}) = 0.52385, \quad \beta_{9189} = 5.8834.$$ 

The 5% VaR becomes

$$\text{VaR} = \$10,000,000 \times 0.0269139 = \$269,139.$$ 

If $p = 0.01$, then VaR becomes $\$448,323$. 

Illustration continued.
Extremes of a stationary time series

**What we have:** \( \{\tilde{x}_i\}_{i=1}^n \) a sequence of iid random variables. Let \( \tilde{x}(n) = \max\{x_i\} \). Then,

\[
\frac{\tilde{x}(n) - \beta_n}{\alpha_n} \to \tilde{F}_*(x),
\]

where \( \tilde{F}_*(x) \) is the GEV.

**What we need:** In practice, it is more appropriate to assume that \( x_i \) forms a stationary time series.

Serial dependence can introduce cluster in extremes.
A heuristic argument

Consider a strictly stationary time series $x_t$. Suppose the data are $\{x_t\}_{t=1}^n$. Let $x_{(n)}$ be the sample maximum. We seek the limiting distribution of $(x_{(n)} - \beta_n)/\alpha_n$ as $n \to \infty$.

IF the serial dependence of $x_t$ decays quickly such that $x_i$ and $x_{i+\ell}$ are essentially independent for a sufficiently large $\ell$.

Divide the sample into disjoint blocks of size $k$. Let $g = \lfloor n/k \rfloor$ be the integer part of $n/k$.

The $i$th block is $\{x_j | j = (i-1)k + 1, \ldots, ik\}$ for $i = 1, \ldots, g + 1$, where it is understood that the last block may have fewer data points.
Denote the $i$th block maximum as $x_{k,i}$. Then it is easily seen that

$$x(n) = \max_{1 \leq i \leq g+1} x_{k,i}.$$ 

Sample maximum is also the maximum of the block maxima.

Suppose that $k$ is sufficiently large and $x_{k,i}$ does not occur near the end of the $i$th block.

In this case, $\{x_{k,i}\}_{i=1}^{g+1}$ can be regarded as an iid random sample. The limiting distribution of its maximum, i.e. $x(n)$, should be GEV.
The limiting distribution is GEV under certain conditions.

But the GEV will differ from that of the iid case, because the sample size is reduced to $g + 1$, not $n$.

A sufficient condition is derived by Leadbetter (1974, 1983) called the $D(u_n)$ condition.

1. $\{u_n\}$ is a sequence of increasing thresholds such that
   \[
   \limsup_n n[1 - F(u_n)] < \infty,
   \]
   where $F(x)$ is the marginal CDF of $x_i$.

2. For any positive integers $p$ and $q$, suppose that $A_i = \{i_v\}_{v=1}^p$ and $A_2 = \{j_v\}_{v=1}^q$ are two sets of arbitrary integers satisfying
   \[
   1 \leq i_1 < i_2 < \cdots < i_p < j_1 < \cdots < j_q \leq n,
   \]
   where $j_1 - i_p \geq \ell_n$ such that $\ell_n \to \infty$ and $\ell_n/n \to 0$ as $n \to \infty$. 
The condition $D(u_n)$ is satisfied if

$$|P(\max_{i \in A_1 \cup A_2} x_i \leq u_n) - P(\max_{i \in A_1} x_i \leq u_n)P(\max_{i \in A_2} x_i \leq u_n)| \leq \delta_n, \ell_n,$$

where $\delta_n, \ell_n \to 0$ as $n \to \infty$.

The condition essentially says that any two events of the form $\{\max_{i \in A_1} x_i \leq u_n\}$ and $\{\max_{i \in A_2} x_i \leq u_n\}$ are asymptotically independent when $A_1$ and $A_2$ are separated by $\ell_n$ with $\ell_n/n \to 0$ and $n \to \infty$. 
If the $D(u_n)$ condition holds with $u_n = \alpha_n x + \beta_n$ for each $x$ such that $\tilde{F}_*(x) > 0$ and if $P[(x(n) - \beta_n)/\alpha_n \leq x]$ converges for some $x$, then

$$P \left( \frac{x(n) - \beta_n}{\alpha_n} \leq x \right) \rightarrow F_*(x) = \tilde{F}_*^\theta(x),$$

for some $\theta \in (0, 1]$. 

$\theta$ is called the extremal index.
VaR for strictly stationary series

\[ \text{VaR} = \beta_n - \frac{\alpha_n}{\xi_n} \left\{ 1 - \left[ -n \theta \ln(1 - p) \right]^{-\xi} \right\}, \quad \xi \neq 0, \]

where \( n \) is the length of the subperiod.

How to estimate \( \theta \)?
Not easy, but several methods are available.

- Blocks method:
- Runs method:
Blocks method

From $P(x(n) \leq u_n) \approx P^\theta(\tilde{x}(n) \leq u_n) = [F(u_n)]^{n\theta}$ and $n[1 - F(u_n)] \to \tau$, we have

$$\lim_{n \to \infty} \frac{\ln P(x(n) \leq u_n)}{n \ln(F(u_n))} = \theta.$$ 

Also,

$$\hat{F}(u_n) = 1 - \frac{1}{n} \sum_{i=1}^{n} I(x_i > u_n) = 1 - \frac{N(u_n)}{n},$$

where $N(u_n)$ is the number of exceedances over the threshold $u_n$, and

$$P(x(n) \leq u_n) = P(\max_{1 \leq i \leq g} x_{k,i} \leq u_n) \approx [P(x_{k,i} \leq u_n)]^g,$$

and $\hat{P}(x_{k,i} \leq u_n) = 1 - \frac{G(u_n)}{g}$, where $G(u_n)$ is the number of exceedances over the threshold $u_n$ of the block maxima $x_{k,i}$. 
Therefore,

$$\hat{\theta}^{(1)} = \frac{1}{k} \ln\left(1 - \frac{G(u_n)}{g}\right) \frac{\ln(1 - N(u_n)/n)}{\ln(1 - N(u_n)/n)}.$$

For the IBM negative log returns, using $u_n = 2.5\%$ and block size $k = 10$, we have $\hat{\theta}^{(1)} = 0.823$.

Consider the 1% VaR obtained before by GEV with $n = 63$. The value is 3.0497 for the iid case. If we use $\hat{\theta}^{(1)}$, then the VaR becomes 3.2714, which, as expected, is higher.

For the runs method, see the handout.
Summary

- Provided an introduction to applying EVT in VaR calculation.
- Demonstrated the concepts using daily IBM stock returns.
- Care must be exercised in using VaR, regardless of which method is used in the calculation.
- Other risk measure, such as expected shortfall, should also be considered.