Extreme Value Theory in Risk Management

- See McNeil, *Extreme Value Theory for Risk Managers*
- Risk management is essentially about tail events.
- Probabilists have developed a theory specific to extreme, or tail, events.
- The generalized Pareto distribution (GPD) plays a central role.
Framework

- \(X_1, X_2, \ldots\) are identically distributed random variables with continuous distribution function

\[ F(x) = P(X_i \leq x). \]

- These \(X\)'s represent *losses*, so we shall focus on the *upper* tail of \(F(\cdot)\).

- For a level \(q\) such as 0.95 or 0.99,

\[ \text{VaR}_q(F) = F^{-1}(q) \]

and

\[ \text{ES}_q(F) = E[X | X > \text{VaR}_q(F)]. \]
Generalized Pareto Distribution

- The GPD has distribution function

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

where \( \beta > 0 \) is a scale parameter, \( \xi \) is a shape parameter, and:

- \( x \geq 0 \) when \( \xi \geq 0 \);
- \( 0 \leq x \leq -\beta/\xi \) when \( \xi < 0 \).

- The GPD is heavy-tailed when \( \xi > 0 \).
- With \( \xi > 0 \),

\[ E(X^k) < \infty \text{ only for } k < 1/\xi. \]
Excess Loss Distribution

- The conditional excess distribution function is
  \[ F_u(y) = P(X \leq u + y|X > u), \quad y \geq 0. \]

- That is,
  \[ F_u(y) = \frac{F(y + u) - F(u)}{1 - F(u)}. \]

- Key result: for “nice” \( F(\cdot) \) and for large \( u \),
  \[ F_u(y) \approx G_{\xi, \beta(u)}(y), \]
  where \( \xi \) and \( \beta(\cdot) \) depend on \( F(\cdot) \).
That is, for a high threshold $u$, losses over that threshold are approximately GPD.

Proposal: given data $X_1, X_2, \ldots, X_n$, choose a “sensible” $u$, and estimate $\xi$ and $\beta = \beta(u)$ by fitting the GPD (e.g. by Maximum Likelihood) to the $N_u$ observed excesses

$$\{X_i : X_i \geq u\}.$$ 

Note that

$$F(x) = [1 - F(u)]F_u(x - u) + F(u)$$

$$\approx [1 - F(u)]G_{\xi,\beta(u)}(x - u) + F(u)$$

and $F(u)$ can be estimated by $(n - N_u)/n$. 

$$5 / 13$$
So for \( x \geq u \), \( F(x) \) can be estimated by

\[
\hat{F}(x) = \left( \frac{N_u}{n} \right) G_{\hat{\xi}, \hat{\beta}}(x - u) + \left( \frac{n - N_u}{n} \right) \\
= 1 - \frac{N_u}{n} \left[ 1 + \hat{\xi} \left( \frac{x - u}{\hat{\beta}} \right) \right]^{-1/\hat{\xi}}
\]
Value at Risk

- The approximation \( \hat{F}(x) \) yields, for \( q > N_u/n \),

\[
\hat{\text{VaR}}_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left[ \left( \frac{n}{N_u} (1 - q) \right)^{-\hat{\xi}} - 1 \right].
\]
Expected Shortfall

For the approximating GPD

\[
\frac{ES_q}{VaR_q} = \frac{1}{1 - \xi} + \frac{\beta(u) - \xi u}{(1 - \xi)VaR_q}.
\]

So we can estimate \( ES_q \) by

\[
\hat{ES}_q = \frac{\hat{VaR}_q}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} u}{(1 - \hat{\xi})}
\]
EVT and GARCH

Above, losses have been modeled as the independent and identically distributed $X_1, X_2, \ldots$

This ignores volatility dynamics.

An EVT-GARCH approach models losses by $Y_1, Y_2, \ldots$, where

$$Y_t = \sigma_t X_t,$$

and $\sigma_t = \sigma(Y_{t-1}, Y_{t-2}, \ldots)$ satisfies a GARCH recursion.

Then for one step ahead, VaR or ES is calculated conditionally on $\sigma_t$:

$$\hat{\text{VaR}}_q(Y_t|\sigma_t) = \sigma_t \hat{\text{VaR}}_q(X)$$
For $k > 1$ steps ahead, it is necessary to simulate $Y_t$, $Y_{t+1}$, $Y_{t+k-1}$.

We have a model only for the upper tail of $X$.

Proposal: use the empirical distribution for the rest of the distribution (i.e., historical simulation, or *bootstrap*).
Extreme Value Distribution

- Classic extreme value theory deals with the distribution of

\[ M_n = \max(X_1, X_2, \ldots, X_n). \]

- We look for constants \( a_n > 0 \) and \( b_n \) such that

\[
\frac{M_n - b_n}{a_n}
\]

converges in distribution to some limit \( H(\cdot) \).

- Fisher-Tippett-Gnedenko Theorem: \( H(\cdot) \) must be the Generalized Extreme Value Distribution; for some \( \xi \),

\[
H(x) = \begin{cases} 
\exp[-(1 + \xi x)^{-1/\xi}] & \xi \neq 0 \\
\exp(-e^{-x}) & \xi = 0.
\end{cases}
\]
If constants $a_n$ and $b_n$ can be found, $F(\cdot)$ belongs to the (maximum) domain of attraction of $H(\cdot)$.

The three cases are:
- $\xi > 0$: the Fréchet distribution;
- $\xi = 0$: the Gumbel distribution;
- $\xi < 0$: the Weibull distribution.

Pickands-Balkema-de Haan Theorem: the distribution function $F(\cdot)$ is “nice” if and only if it belongs to the domain of attraction of $H(\cdot)$ for some $\xi$.

The parameter $\xi$ has the same value in the GEVD as in the GPD.
A necessary and sufficient condition for $F(\cdot)$ to belong to the domain of attraction of the Fréchet distribution (GEVD with $\xi > 0$) is

$$\frac{1 - F(tx)}{1 - F(t)} \rightarrow x^{-1/\xi} \text{ as } t \rightarrow \infty.$$ 

That is, $1 - F(x)$ is regularly varying as $x \rightarrow \infty$.

The same condition is therefore also necessary and sufficient for convergence of $F_u(y)$ to the GPD.