IEOR E4602: Quantitative Risk Management

Extreme Value Theory

Martin Haugh

Department of Industrial Engineering and Operations Research
Columbia University
Email: martin.b.haugh@gmail.com

Motivation

- Suppose we wish to estimate VaR_α for a given portfolio.
- We could use the empirical α-quantile, q_α.
- But there are many potential problems with this approach
  - there may not be enough data
  - the empirical quantile will never exceed the maximum loss in the data-set
  - time series dependence is ignored, i.e., we will be working with the unconditional loss distribution.
- Extreme value theory helps overcome these problems.

Extreme Value Theory (EVT)

- Two principal parametric approaches to modeling the extremes of a probability distribution:
  1. The block maxima approach
  2. The threshold exceedances approach.
- Threshold exceedances approach is more modern and usually the preferred approach
  - makes better use of available data.
- The Hill Estimator approach is also commonly used
  - this is a non-parametric approach.
- EVT can be combined with time-series models to estimate conditional loss distributions
  - and therefore construct better estimates of VaR, ES, etc.

The GEV Distributions

Definition: The CDF of the generalized extreme value (GEV) distribution satisfies

$$H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{-1/\xi}}, & \xi \neq 0 \\ e^{-e^{-x}}, & \xi = 0. \end{cases}$$

where $1 + \xi x > 0$.

- A three-parameter family is given by $H_{\xi,\mu,\sigma}(x) := H_\xi((x-\mu)/\sigma)$
  - μ is the location parameter
  - σ is the scale parameter
  - ξ is the shape parameter.
- $H_\xi(\cdot)$ defines the type of the distribution
  - i.e. recall a type is a family of distributions specified up to location and scale.

The GEV Distributions

Definition: The right endpoint, x_F, of a distribution with CDF, $F(\cdot)$, is given by $x_F := \sup\{x \in \mathbb{R} \ : \ F(x) < 1\}$.

- When $\xi > 0$ obtain the Fréchet distribution
  - has an infinite right endpoint.
- When $\xi = 0$ obtain the Gumbel distribution
  - has an infinite right endpoint
  - tail decays much faster than tail of Fréchet distribution.
- When $\xi < 0$ obtain the Weibull distribution
  - a short-tailed distribution with finite right endpoint

Convergence of Maxima

- Role of GEV distribution in the theory of extremes is analogous to role of normal distribution in the Central Limit Theorem (CLT) for sums of random variables.

- Recall the CLT: if $X_1, X_2, \ldots$ are IID with a finite variance then

$$\frac{S_n - a_n}{b_n} \longrightarrow \mathsf{N}(0,1) \text{ in distribution where}$$

$$\begin{aligned} S_n &:= \sum_{i=1}^{n} X_i \\ a_n &:= n\,\mathsf{E}[X_1] \\ b_n &:= \sqrt{n\mathsf{Var}(X_1)} \end{aligned}$$

- Let $M_n := \max(X_1, \ldots X_n)$, i.e., the block maximum.

- The block maxima approach to EVT is concerned with the limiting distribution of M_n.

The Maximum Domain of Attraction

Definition: A CDF, F, is said to be in the maximum domain of attraction (MDA) of H if there exist sequences of constants, c_n and d_n with $c_n > 0$ for all n, such that

$$\lim_{n\to\infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = H(x) \tag{1}$$

for some non-degenerate CDF, H.

Note that (1) implies (why?)

$$\lim_{n\to\infty} F^n(c_n x + d_n) = H(x). \tag{2}$$

The Fisher-Tippett Theorem (1920's)

Theorem: If $F \in \text{MDA}(H)$ for some non-degenerate CDF, H, then H must be a distribution of type H_ξ, i.e., a GEV distribution.

- If convergence of normalized maxima takes place, then the type of the distribution is uniquely determined. The location, μ, and scaling, σ, depend on the normalizing sequences, c_n and d_n.
- Essentially all the commonly used distributions of statistics are in $\text{MDA}(H_\xi)$ for some ξ.

Example: The Exponential Distribution

- Suppose the X_i's are IID Exp(λ) so that

$$F(x) \; = \; 1 \; - \; e^{-\lambda x}$$

  for $x \geq 0$ and $\lambda > 0$.

- Let $c_n := 1/\lambda$ and $d_n := \ln(n)/\lambda$.
- Can directly calculate the limiting distribution using (1).
- We obtain

$$F^n(c_n x + d_n) \quad = \quad \left(1 - \frac{1}{n} \, e^{-x}\right)^n, \qquad x \geq -\ln(n)$$

  so that

$$\lim_{n \to \infty} F^n(c_n x + d_n) \; = \; e^{-e^{-x}}.$$

- Therefore obtain $F \in$ MDA(H_0).

The Fréchet MDA

Definition:

(i) A positive function, L, on $(0, \infty)$ is slowly varying at ∞ if

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0.$$

(ii) A positive function, h, on $(0, \infty)$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$ if

$$\lim_{x \to \infty} \frac{h(tx)}{h(x)} = t^{\rho}, \quad t > 0.$$

e.g. The logarithmic function, $\log(x)$, is slowly varying.

The Fréchet MDA

Theorem: For $\xi > 0$,

$$F \in \mathrm{MDA}(H_\xi) \iff \bar{F}(x) \; = \; x^{-1/\xi} \, L(x)$$

for some function, L, that is slowly varying at ∞ and where $\bar{F}(x) := 1 - F(x)$.

- When $F \in \mathrm{MDA}(H_\xi)$, often refer to $\alpha \; := \; 1/\xi$ as the tail index of the distribution.
- e.g. Fréchet, t, F and Pareto are all in Fréchet MDA.
- Can be shown that if $F \in \mathrm{MDA}(H_\xi)$ for $\xi > 0$, then $\mathrm{E}[X^k] = \infty$ for $k > 1/\xi = \alpha$.

The Gumbel and Weibull MDA's

- The Gumbel and Weibull distributions aren't as interesting from a finance perspective
  - but their MDA's can still be characterized.
- e.g. exponential, normal and log-normal are in Gumbel MDA
  - $\mathsf{E}[X^k] < \infty$ for all $k > 0$ in this case.
- e.g. Beta distribution is in Weibull MDA.

The Non-IID Case

- So far have dealt with only the IID case.
- But in finance, data is rarely IID.
- Can be shown, however, that for most strictly stationary time series, our results continue to hold
  - e.g. our results hold for ARCH / GARCH models
  - if it exists, the extremal index, $\theta \in (0, 1]$, of the time series is key!
    - $n\theta$ can be interpreted as the number of independent clusters of observation in n observations.
  - see Section $7.1.3$ of *McNeil, Frey and Embrechts* for further details.

The Block Maxima Method

- Assume we have observation $X_1, \ldots, X_{nm}$
  - so that the data can be split into m blocks with $M_j := \max\{j^{th} \text{ block}\}$
  - each block contains n observations.
- Would like both n and m to be large but there are tradeoffs
  - would like n large so that convergence to the GEV has occurred
  - would like m large so that we have more observations and hence lower variances of MLE estimates.
- In practice, if we are working with daily data and we have sufficiently many observations, might take quarterly, semi-annual or annual block sizes.

The Block Maxima Method

- Let $h_{\xi,\mu,\sigma}$ be the log-density.
- Then log-likelihood for $\xi \neq 0$ given by

$$
\begin{aligned}
l(\xi, \mu, \sigma\ ;\ M_1, \ldots, M_m) &= \sum_{i=1}^{m} h_{\xi,\mu,\sigma}(M_i) \\
&= -m \ln(\sigma)\ -\ \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \ln\left(1 + \xi\ \frac{M_i - \mu}{\sigma}\right) \\
&\quad -\ \sum_{i=1}^{m} \left(1 + \xi\ \frac{M_i - \mu}{\sigma}\right)^{-1/\xi}.
\end{aligned}
$$

- We then maximize the log-likelihood over (ξ, μ, σ) subject to
  - $\sigma > 0$ and
  - $1 + \xi\ (M_i - \mu)/\sigma\ >\ 0$ for all $i = 1, \ldots, m$.

The Return Level and Return Period Problems

- The fitted GEV model can be used to analyze stress losses. In particular we have the return level problem and the return period problem.

Definition: Let H denote the CDF of the true n-block maximum. Then the k n-block return level is

$$r_{n,k} \;:=\; q_{1-1/k}(H)$$

i.e., the $(1-1/k)$-quantile of H.

- The k n-block return level can be interpreted as the level that is exceeded once out of every k n-blocks on average.
- Using our fitted model, we obtain

$$\hat{r}_{n,k} \;=\; H^{-1}_{\hat{\xi},\hat{\mu},\hat{\sigma}}\left(1-\frac{1}{k}\right) \;=\; \hat{\mu} \;+\; \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(-\ln\left(1-\frac{1}{k}\right)\right)^{-\hat{\xi}} \;-\; 1\right)$$

  - estimates of $r_{n,k}$ should always (why?) be accompanied by confidence intervals.

The Return Level and Return Period Problems

Definition: Let H denote the CDF of the true n-block maximum. The return period of the event $\{M > u\}$ is given by

$$k_{n,u} \;:=\; 1/\bar{H}(u)$$

where $\bar{H}(u) = 1 - H(u)$.

- $k_{n,u}$ is the average number of blocks we must wait before we observe the event $\{M > u\}$.
- Again, an estimate of $k_{n,u}$ should always be accompanied by confidence intervals.

Threshold Exceedances

- The block maxima approach is inefficient as it ignores all but the maximum observation in each block.
- The threshold exceedance approach does not suffer from this approach
  - it uses all of the data above some threshold, u.
- The Generalized Pareto Distribution (GPD) plays the key role in the threshold exceedance approach.

The Generalized Pareto Distribution

Definition: The Generalized Pareto Distribution (GPD) is given by

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

where $\beta > 0$, and $x \geq 0$ when $\xi \geq 0$, and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

- ξ is the shape parameter
- β is the scale parameter.

- When $\xi > 0$ obtain the ordinary Pareto distribution.
- When $\xi = 0$ obtain the exponential distribution.
- When $\xi < 0$ obtain the short-tailed Pareto distribution.

Excess Distribution Over a Threshold

Definition: Let X be a random variable with CDF, F. Then the excess distribution over the threshold u has CDF

$$F_u(x) \;=\; P(X - u \leq x \mid X > u) \;=\; \frac{F(x+u) - F(u)}{1 - F(u)} \tag{3}$$

for $0 \leq x < X_f - u$ where $x_F \leq \infty$ is the right endpoint of F.

- In survival analysis F_u is known as the residual life CDF.

Definition: The mean excess function of a random variable, X, with finite mean is given by

$$e(u) \;:=\; \mathsf{E}[X - u \mid X > u].$$

Examples: Exponential and GPD Random Variables

- **e.g.** If $X \sim \text{Exp}(\lambda)$, then can show that $F_u(x) = F(x)$
  - reflects the memoryless property of exponential random variables.

- **e.g.** Suppose $X \sim G_{\xi,\beta}$. Then (3) implies

$$F_u(x) \;=\; G_{\xi,\beta(u)} \quad \text{where}$$

$$\beta(u) \;:=\; \beta + \xi u$$

$$0 \leq x < \infty \text{ if } \xi \geq 0 \;\text{ and }\; 0 \leq x \leq -\beta/\xi - u \text{ if } \xi < 0$$

- so the excess CDF remains a GPD with the same shape parameter but with a different scaling.
- can also show that the mean excess function satisfies

$$e(u) \;=\; \frac{\beta(u)}{1-\xi} \;=\; \frac{\beta + \xi u}{1-\xi}$$

where $0 \leq u < \infty$ if $0 \leq \xi < 1$ and $0 \leq u \leq -\beta/\xi$ if $\xi < 0$

- note that $e(u)$ is linear in u for the GPD, a useful property!

The GPD and MDA's

Theorem: We can find a positive function, $\beta(u)$, such that

$$\lim_{u \to x_F} \sup_{0 \leq x < x_F - u} \left| F_u(x) \; - \; G_{\xi,\beta(u)}(x) \right| \; = \; 0 \tag{4}$$

if and only if $F \in \text{MDA}(H_\xi)$, $\xi \in \mathbb{R}$.

- This theorem provides the link between the theories of block maxima and threshold exceedances.
- Since essentially all commonly used distributions are in $\text{MDA}(H_\xi)$ for some ξ, we see that the GPD distribution is the canonical distribution for excess distributions.
- Note that the shape parameter, ξ, does not depend on u.
- Can use (4) by taking u to be "large" and therefore assuming that $F_u(x) = G_{\xi,\beta}(x)$ for $0 \leq x < x_F - u$ and some ξ and $\beta > 0$.

Modeling Excess Losses

- Let $X_1, \ldots, X_n$ represent loss data from the distribution F.
- A random number N_u will exceed the threshold, u.
- Let $Y_1, \ldots, Y_{N_u}$ be the values of the N_u excess losses.
- We assume $F_u = G_{\xi,\beta}$ and estimate ξ and β using maximum likelihood.
- Obtain

$$\begin{aligned} l(\xi, \beta \; ; \; Y_1, \ldots, Y_{N_u}) &= \sum_{i=1}^{N_u} \ln g_{\xi,\beta}(Y_i) \\ &= -N_u \ln(\beta) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{N_u} \ln\left(1 + \xi \frac{Y_i}{\beta}\right) \end{aligned}$$

which we maximize subject to $\beta > 0$ and $1 + \xi Y_i/\beta > 0$ for all i.

When the Data is Not IID

- So far have assumed the data is IID
  - but of course we know financial return data is not IID!
- If the extremal index, θ, equals 1 then no evidence of extremal clustering
  - so fine to assume data is IID.
- If $\theta < 1$ then there is evidence of extremal clustering
  - situation not so satisfactory
  - but can still use the MLE method to estimate the parameters
  - technically this becomes quasi-MLE since the model is misspecified
  - point estimates of the parameters should still be fine
  - but standard errors might be too small in which case associated confidence intervals would also be too narrow.

Excesses Over Higher Thresholds

Lemma: Suppose $F_u(x) = G_{\xi,\ \beta}(x)$ for $0 \leq x < x_F - u$ for some ξ and $\beta > 0$. Then $F_v(x) = G_{\xi,\ \beta+\xi(v-u)}(x)$ for any higher threshold $v \geq u$.

- So excess distribution over higher thresholds remains a GPD with same shape parameter, ξ, but with a scaling parameter that grows linearly in v.
- If $\xi < 1$, the mean excess function satisfies

$$e(v) = \frac{\beta + \xi(v-u)}{1-\xi} = \frac{\xi v}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \qquad (5)$$

where $u \leq v < \infty$ if $0 \leq \xi < 1$ and $u \leq v \leq u - \beta/\xi$ if $\xi < 0$

- linearity of (5) in v can be used as a diagnostic for choosing the appropriate threshold, u
- this diagnostic tool is called the sample mean excess plot.

Sample Mean Excess Plot

Definition: Given loss data $X_1, \ldots, X_n$, the sample mean excess function is the empirical estimator of the mean excess function given by

$$e_n(v) \ := \ \frac{\sum_{i=1}^{n}(X_i - v)\, 1_{\{X_i > v\}}}{\sum_{i=1}^{n} 1_{\{X_i > v\}}}$$

- Now can construct the mean excess plot $\{X_{(i,n)},\ e_n(X_{(i,n)}) \ : \ 2 \leq i \leq n\}$ where $X_{(i,n)}$ is the i^{th} order statistic.
- If the data support a GPD model beyond a high threshold, then the plot should become linear for higher values of v
  - a positive slope indicates $\xi > 0$
  - a zero slope indicates $\xi \approx 0$
  - a negative slope indicates $\xi < 0$.
- Since final few values are based on very few data points they are often omitted from the plot.

Tail Probabilities

- Again assuming that $F_u(x) = G_{\xi,\,\beta}(x)$ for $0 \leq x < x_F - u$ we obtain for $x > u$

$$
\begin{aligned}
1 - F(x) \; = \bar{F}(x) &= P(X > u)\; P(X > x \mid X > u) \\
&= \bar{F}(u)\; P(X - u > x - u \mid X > u) \\
&= \bar{F}(u)\; \bar{F}_u(x - u) \\
&= \bar{F}(u)\; \left(1 + \xi \frac{x - u}{\beta}\right)^{-1/\xi} \qquad (6)
\end{aligned}
$$

  - so if we know $\bar{F}(u)$ we have a formula for the tail probabilities
  - (6) can now be inverted to compute risk measures!

Risk Measures

- For $\alpha \geq F(u)$ obtain

$$\mathsf{VaR}_\alpha \;=\; q_\alpha(F) \;=\; u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\xi} \;-\; 1\right).$$

- If $\xi < 1$, then

$$\mathsf{ES}_\alpha \;=\; \frac{1}{1-\alpha}\int_\alpha^1 q_x(F)\;dx \;=\; \frac{\mathsf{VaR}_\alpha}{1-\xi} \;+\; \frac{\beta - \xi u}{1-\xi}.$$

- Also obtain

$$\lim_{\alpha\to 1}\frac{\mathsf{ES}_\alpha}{\mathsf{VaR}_\alpha} \;=\; \begin{cases} (1-\xi)^{-1}, & 1 > \xi \geq 0 \\ 1, & \xi < 0. \end{cases}$$

Estimation in Practice

- Can use the sample mean excess plot to choose an appropriate threshold, u.
- MLE methods then used to estimate ξ and β as well as their standard errors.
- We can use the empirical estimator, N_u/n, to estimate $\bar{F}(u)$.
- Then have

$$\hat{\bar{F}}(x) \quad = \quad \frac{N_u}{n} \left(1 + \hat{\xi}\, \frac{x - u}{\hat{\beta}}\right)^{-1/\hat{\xi}} \tag{7}$$

  as our tail probability estimator for $x \geq u$

  - should also compute confidence intervals for (7)
    - either using Monte-Carlo (how?) or by reparametrizing (how?).
- Should also study sensitivity of parameter estimates to the threshold, u
  - results are not reliable if estimates remain sensitive for large u.

Multivariate EVT

- Can also study extreme value theory for multivariate data
  - leads to multivariate EVT.
- The marginal distributions are as in the univariate case
  - e.g. GPD for the threshold exceedances method.
- So the main item of concern is the dependency structure
  - leads to extreme value copulas
  - e.g. the Gumbel copula is a 2-dimensional EV copula.
- Generally difficult to apply Multivariate EVT in high dimensions
  - too many parameters to estimate.
- A common solution is to simply collapse the problem to the univariate case by considering the entire portfolio value or return as a univariate random variable.

Example: Danish Fire Loss Data

- Dataset consists of $2,156$ fire insurance losses over 1m Danish Kroner from 1980 to 1990
  - representing combined loss for building and contents and sometimes, business earnings
  - losses are inflation adjusted to 1985 levels.
- Mean excess plot appears linear over entire range
  - so GPD with $\xi > 0$ could be fitted to entire dataset.
- We find $\hat{\xi} \approx .52$
  - so fitted model is very heavy-tailed with infinite variance. Why?
    - because $\mathsf{E}[X^k] = \infty$ for any GPD distribution with $k \geq 1/\xi$.

The Hill Estimator

- The Hill method assumes $F \in \text{MDA}(H_\xi)$ for $\xi > 0$, i.e., the Fréchet MDA
  - so $\bar{F}(x) \; = \; L(x) \, x^{-1/\xi}$ where L is slowly varying.
- The estimator satisfies

$$\hat{\xi}_k^{Hill} \;=\; \frac{1}{k} \sum_{i=1}^{k} \ln(X_{i,n}) \;-\; \ln(X_{k,n}), \quad 2 \leq k \leq n$$

where $X_{n,n} \leq \cdots \leq X_{1,n}$ are the order statistics.

  - often a very good estimator of ξ when the tail probability is well approximated by a power function
- It is common to plot the Hill estimator for different values of k
  - obtain the Hill plot
  - and to then choose a value of k from a region where the estimator is relatively stable.

Where Does the Hill Estimator Come From?

Consider the mean excess for function, $e(\cdot)$, for $\ln(X)$. We obtain:

$$
\begin{aligned}
e(\ln(u)) &= \mathsf{E}\left[\ln(X) - \ln(u) \mid \ln(X) > \ln(u)\right] \\
&= \frac{1}{\bar{F}(u)} \int_u^\infty (\ln(x) - \ln(u))\, dF(x) \\
&= \frac{1}{\bar{F}(u)} \int_u^\infty \frac{\bar{F}(x)}{x}\, dx \qquad \text{(using integration by parts)} \\
&= \frac{1}{\bar{F}(u)} \int_u^\infty L(x) x^{-(1+1/\xi)}\, dx \\
&\approx \frac{L(u)}{\bar{F}(u)} \int_u^\infty x^{-(1+1/\xi)}\, dx \qquad \text{(for } u \text{ sufficiently large)} \\
&= \frac{L(u) u^{-1/\xi} \xi}{\bar{F}(u)} \\
&= \xi. \qquad (8)
\end{aligned}
$$

Conditional or Dynamic EVT for Financial Time Series

- So far, our applications of EVT lead to estimates of the unconditional loss distribution.
- But we are usually (much) more interested in the conditional loss distribution
  - at least in the case of financial applications
  - generally not true in the case of insurance applications. Why?
- Can apply EVT to obtain estimate of the conditional loss distribution using time series models
  - in particular, ARCH / GARCH models.

Conditional or Dynamic EVT for Financial Time Series

- Suppose the negative log-returns (from date $t-1$ to date t) are generated by a strictly stationary time series

$$L_t = \mu_t + \sigma_t Z_t$$

  - μ_t and σ_t are known at time $t-1$
  - and the Z_t's are IID innovations with unknown CDF, $G(\cdot)$.

- The risk measures VaR_α^t and ES_α^t (at date $t-1$) satisfy

$$\begin{aligned} \text{VaR}_\alpha^t &= \mu_t + \sigma_t\, q_\alpha(Z) \\ \text{ES}_\alpha^t &= \mu_t + \sigma_t\, \text{ES}_\alpha(Z) \end{aligned}$$

  where q_α is the α-quantile of Z.

- We can estimate VaR_α^t and ES_α^t by first fitting a GARCH model to the L_t's
  - but we don't know the distribution, $G(\cdot)$, of Z
  - so we need to use quasi-maximum likelihood estimation (QMLE) instead of the usual MLE.

Conditional or Dynamic EVT for Financial Time Series

- The fitted GARCH model can be used to estimate μ_t and σ_t.
- We want to apply EVT to the innovations, Z, but we don't observe the Z's.
- Instead we take the GARCH residuals as our data for EVT.
- We fit the GPD to the tails of the residuals and estimate the corresponding risk measures to obtain

$$
\begin{aligned}
\widehat{\text{VaR}}_\alpha^t &= \hat{\mu}_t + \hat{\sigma}_t \, \hat{q}_\alpha(Z) \\
\widehat{\text{ES}}_\alpha^t &= \hat{\mu}_t + \hat{\sigma}_t \, \widehat{\text{ES}}_\alpha(Z)
\end{aligned}
$$

- See Section 3 of *"Extreme Value Theory for Risk Managers"* by McNeil or Sections 7.2.6 and 2.3.6 of MFE
  - note how well the dynamic EVT VaR method back-tests!
- See also *Risk Management and Time Series* lecture notes.