SOA PREDICTIVE ANALYTICS SEMINAR
SEOUL, SOUTH KOREA
SEP. 8, 2017

Session 5

Gentle introduction to Extreme Value Theory for large insurance claims

Joseph (Hyun Tae) Kim, PhD, FSA, CERA

SOA Predictive Analytics Seminar

8, Sept. 2017

SOCIETY OF ACTUARIES

A gentle introduction to Extreme Value Theory for large insurance claims

JOSEPH H.T. KIM PHD, FSA, CERA

Associate Professor, Yonsei University

Sep 2017

SOCIETY OF ACTUARIES

Why Extreme Value Theory (EVT)?

- Traditional statistical methods focus on the centre of the sample; e.g, mean or median in estimations and tests
- In modern finance and insurance, jumbo or extreme losses are more important
  - Normal losses may hurt profitability, but can be absorbed by risk margins or reserves
  - Jumbo losses can make the firm bankrupt
  - To cover this unexpectedly large losses, we need the risk capital
- Extreme losses have a severe impact on business
  - Rare event (very low frequency)
  - Huge impact if it does occurs (very high severity)

Disciplines where extreme losses are relevant

- Structural engineering: Wind, rain, waves or flood damage to a structure
- Finance: Crash in stock prices and exchange rates shocks
- Earth sciences: Earthquakes, floods, hurricanes
- Traffic prediction, Internet traffics
- Insurance: Terror attacks, hurricanes, frauds

Examples

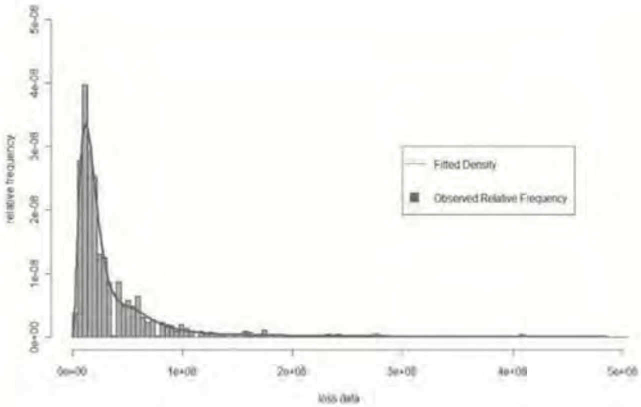

Figure: Catastrophic property losses in US, 1997–2005 (Lee & Lin, 2010)

Examples

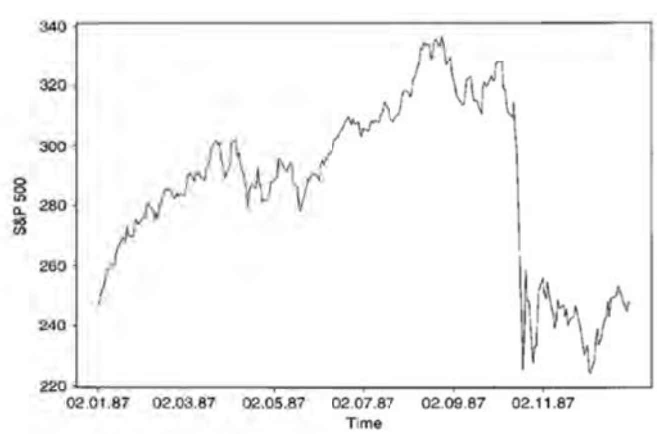

Figure: S&P index during the 1987 crash (Embrechts et al., 1999)

Examples

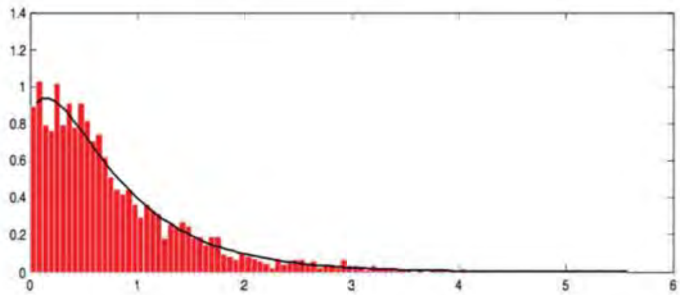

Figure: Danish fire insurance losses in 'log scale' (Park & Kim, 2016)

Challenges in modeling extreme losses

- Common theme is to model the tail area of the datasets (Right tail in insurance; left tail in banking)
- Important to understand and properly model extreme losses (observations) in the tail region of the dataset
- But modeling extreme losses is not easy:
  - Data are scarce in the tail area, so empirical values are highly volatile.
  - Shape of the tail can be different depending on the dataset
  - What kind of parametric distribution should be fitted for the tail area?

Two approaches in EVT

1. Block maxima
   - Has long been used in hydrology
   - Collect the annual (or other period) maxima
   - Fit the maxima to the generalized extreme value (GEV) distribution (Fisher-Tippett theorem)
   - In practice the fit tends to be poor for GEV because of the limited number of relevant events, e.g., only 10 observations for 10 years $\Rightarrow$ Wasteful approach.
2. Peaks over threshold (POT)
   - Extract all observations above a given large threshold u, ignoring time element. These observations are called exceedances.
   - Fit exceedances to the generalized Pareto distribution (GPD) (Pickands-Balkema-de Haan theorem)
   - Less wasteful approach compared to the block maxima

Basic theory of block maxima

- Let $\{X_i\}$, $i = 1, 2, ...$, be a sequence of iid rv with common df $F_X(x)$
- Denote the right end point of $F_X(x)$ as $x_F = \sup\{x \in \Re : F(x) < 1\}$
- Define the sample maxima

$$M_n = \max\{X_1, \cdots, X_n\} \tag{1}$$

- The df of M_n is easily obtained as

$$\Pr\{M_n \le x\} = \Pr\{X_1 \le x, \cdots, X_n \le x\} = \{F_X(x)\}^n. \tag{2}$$

- However, as $n \to \infty$, for $x < x_F$,

$$\Pr\{M_n \le x\} = \{F_X(x)\}^n \to 0 \tag{3}$$

which implies that, for $x \ge x_F$,

$$\Pr\{M_n \le x\} = \{F(x)\}^n = 1 \tag{4}$$

$\Rightarrow$ Not very useful in practice.

- A more useful result can be obtained for standardized M_n.
- Fisher-Tippett theorem says that, if there exist constants sequences $c_n > 0$ and $d_n \in \Re$ such that

$$\frac{M_n - d_n}{c_n} \longrightarrow H, \quad as\ n \to \infty,$$

for some non-degenerate df H, then H must belong to one of the standard extreme value distribution class:

$$Frech\acute{e}t: \ \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases}$$

$$Gumbel: \ \Lambda(x) = \exp\{-e^{-x}\}, \ x \in \Re$$

$$Weibull: \ \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

- That is, if the standardized sample maximum has a proper distribution asymptotically, it should be one of these three.
- Most familiar distributions have this limit property, with some exceptions (e.g., Poisson maxima does not have this property)
- So it makes sense to use these three distributions to classify the tail heaviness for a given distribution.
- If a distribution's sample maxima has this limit property, we say that the distribution is in the maximum domain of attraction (MDA) of H. Some well-known examples are:
  - Pareto and Log-gamma distribution is MDA(Φ)
  - Exponential and Gamma distributions is MDA(Λ)
  - Uniform distribution is MDA(Ψ)
- We see that Frechét (Φ), Gumbel (Λ) and Weibull (Ψ) correspond to heavy, medium and short tailed distributions, respectively.

- These three distributions can be combined and represented as one expression using Jenkinson-von Mises representation
  $\Rightarrow$ called the generalized extreme value (GEV) distribution
- In GEV representation, we introduce a parameter ξ such that
  $\xi = \alpha^{-1} > 0$ corresponds to the Frechet distribution Φ_α,
  $\xi = 0$ corresponds to the Gumbel distribution Λ,
  $\xi = -\alpha^{-1} < 0$ corresponds to the Weibull distribution Ψ_α.
- The df of GEV is given by

$$H_\xi(x) = \begin{cases} \exp\left\{-(1+\xi x)^{-1/\xi}\right\} & \text{if } \xi \neq 0 \\ \exp\{-\exp(-x)\} & \text{if } \xi = 0, \end{cases} \tag{5}$$

  where $1 + \xi x > 0$.
- By adding a suitable location and scale parameter, one can verify that GEV has these 3 distributions as special cases.
- Also, we often use MDA(H_ξ) instead of MDA(Φ), MDA(Λ) or MDA(Ψ)

Example

If $X_1, X_2, \ldots$ is a sequence of independent standard exponential Exp(1) variables, $F(x) = 1 - e^{-x}$ for $x > 0$.
In this case, letting $c_n = 1$ and $d_n = \log n$,

$$\begin{aligned} Pr\{(M_n - d_n)/c_n \leq z\} &= F^n(z + \log n) \\ &= [1 - e^{-(z+\log n)}]^n \\ &= [1 - n^{-1}e^{-z}]^n \\ &\to exp(-e^{-z}) \end{aligned}$$

as $n \to \infty$, for each fixed $z \in \Re$. Hence, with the chosen c_n and d_n, the limit distribution of M_n as $n \to \infty$ is the standard Gumbel distribution, corresponding to $\xi = 0$ in the GEV family.
$\Rightarrow$ Exponential distribution is in MDA(H_0)

Example

If $X_1, X_2, \dots$ is a sequence of independent standard Frechét variables, $F(x) = exp(-1/x)$ for $x > 0$. Letting $c_n = n$ and $d_n = 0$,

$$\begin{aligned} Pr\{(M_n - d_n)/c_n \leq z\} &= F^n(nz) \\ &= [exp\{-1/(nz)\}]^n \\ &\to exp(-z^{-1}) \end{aligned}$$

as $n \to \infty$, for each fixed $z > 0$. Hence, the limit in this case is also the standard Frechét distribution : $\xi = 1$ in the GEV family.
$\Rightarrow$ Frechét distribution is in $MDA(H_1)$

Example

If $X_1, X_2, \dots$ are a sequence of independent uniform $U(0, 1)$ variables, $F(x) = x$ for $0 \leq x \leq 1$. For fixed $z < 0$, suppose $n > -z$ and let $c_n = 1/n$ and $d_n = 1$. Then,

$$\begin{aligned} Pr\{(M_n - d_n)/c_n \leq z\} &= F^n(n^{-1}z + 1) \\ &= (1 + \frac{z}{n})^n \\ &\to e^z \end{aligned}$$

as $n \to \infty$. Hence, the limit distribution is of Weibull type, with $\xi = -1$ in the GEV family.
$\Rightarrow$ Uniform distribution is in $MDA(H_{-1})$

- Practical problem with this is that we never know (c_n, d_n) because the underlying loss df is unknown.
- However, this issue can be resolved as follows
- Note that ${c_n}^{-1}(M_n - d_n) \longrightarrow H_\xi$ is equivalent to

$$\lim_{n\to\infty} \Pr(M_n \le c_n x + d_n) = \lim_{n\to\infty} [F_X(c_n x + d_n)]^n = H_\xi(x), \;\; \xi \in \Re.$$

- If we let $y = c_n x + d_n$, for large n, we can rewrite the above as

$$\Pr(M_n \le y) \approx H_\xi\left(\frac{y - d_n}{c_n}\right), \;\; \xi \in \Re.$$

- Thus, if we have m block maxima values, $M_{1,n}, \ldots, M_{m,n}$, with large n, we can use these to estimate GEV parameters ξ and (c_n, d_n) based on, e.g., MLE, where c_n, d_n are the scale and location parameters, respectively.
- But this block maxima approach needs large m and n to work well, which in turn requires a huge sample size.

Peaks over threshold (POT) approach

- More popular than the block maxima method
- Using the block maxima only is not desirable as it ignores other relevant extremes (e.g., the 2nd largest loss for a given period).
- Pickands-Balkema-de Haan theorem states that a distribution function of the excess over a high threshold converges to generalized Pareto distribution (GPD) regardless of its original distribution, under some technical conditions.
- All observations exceeding a high threshold u can be used in modeling, so less wasteful compared to the block maxima method.

$\Rightarrow$ Our focus will be the POT.

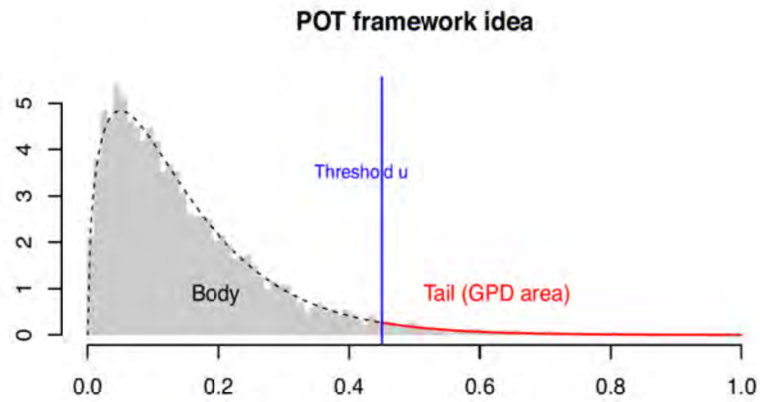

Figure: GPD is fitted for the tail which starts from threshold u

The Generalized Pareto Distribution (GPD)

- Like GEV, the GPD is a family of distributions
- The df of the GPD with shape, scale, location parameters $\xi \in \mathbb{R}$, $\sigma > 0$ and $\mu \in \mathbb{R}$ is defined as

$$G_{\mu,\sigma,\xi}(x) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}, & \xi \neq 0. \\ 1 - \exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right], & \xi = 0. \end{cases} \tag{6}$$

- Domain of x is $(0, \infty)$ for $\xi \geq 0$, and $(0, -\sigma/\xi)$ when $\xi < 0$.
- The case $\xi = 0$ is a shifted exponential distribution, which is obtained from taking limit $\xi \to 0$.
- In many EVT applications, we do not need the location parameter, so set $\mu = 0$ and write $G_{\sigma,\xi}(x)$

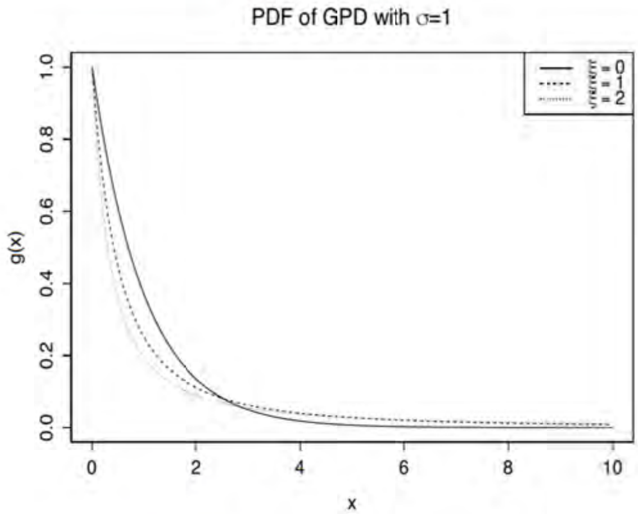

Figure: Density of GPD when $\xi \geq 0$

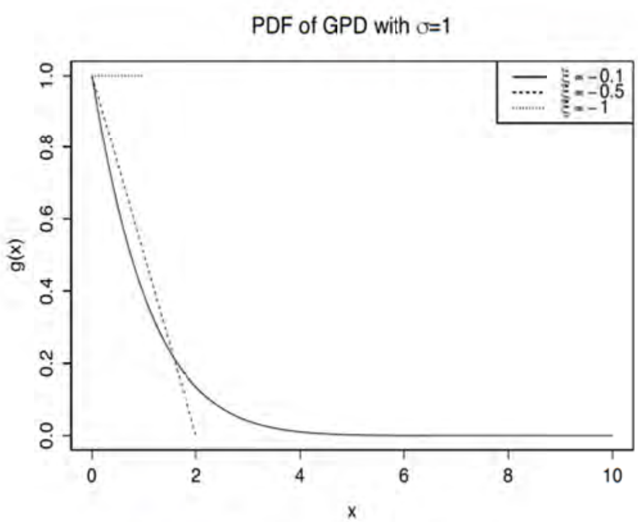

Figure: Density of GPD when $\xi < 0$

Theory behind POT

- Let us now connect GPD to the POT framework. For this, define the excess loss df of any loss variable X over threshold u, denoted $X - u|X > u$, by

$$F_u(x) = \Pr\{X - u \leq x|X > u\} = \frac{F(u + x) - F(u)}{1 - F(u)}.$$

- The famous Pickands-Balkema-de Haan theorem states that

$$\lim_{u \uparrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\sigma(u),\xi}(x)| = 0, \qquad (7)$$

  if and only if $F \in MDA(H_\xi)$, $\xi \in \mathbb{R}$. Here $\sigma(u) = \sigma + \xi u$.
- Being the most important result in the POT framework, this theorem says that, for large enough u, $F_u(y)$ tends to GPD.
- Or, equivalently, the df of the sample exceedances above u can be approximated by a GPD, as the threshold is raised.

POT: Formulation

- Important to know that GPD actually is a unified expression of three different distributions depending on the sign of ξ.
  1. $\xi > 0$ case means that the tail is heavy (power or polynomial tail)
  2. $\xi = 0$ case means that the tail thickness is medium (exponential tail)
  3. $\xi < 0$ case means that the tail is light (x has the finite upper limit)
- Also important to know there is a connection between Pickands-Balkema-de Haan theorem and Fisher-Tippet theorem: They share the same ξ (see next slide for examples)
- In most insurance and financial applications, we are interested in $\xi \geq 0$ and we will focus on this case.

Examples of which the exceedances produce GPD

- Pareto, log-gamma, and Cauchy are in MDA(H_ξ) with $\xi > 0$
  $\Rightarrow$ Hence their exceedances follow GPD with $\xi > 0$, the class of heavy-tailed distributions. (most important case in practice)
- Medium-tailed distributions, e.g., normal, exponential, gamma, and lognormal are in MDA(H_0)
  $\Rightarrow$ Their exceedances follow exponential distribution.
- Light-tailed distributions, e.g., uniform and beta distributions are in are in MDA(H_ξ) with $\xi < 0$
  $\Rightarrow$ Their exceedances follow Pareto II distribution (its support has a finite upper limit)

Challenges in applying GPD under POT approach

- Following issues are interrelated when the GPD is applied to the sample exceedances

(1) Selecting threshold u:
  - Pickands-Balkema-de Haan theorem holds as $u \to \infty$, but we must choose a finite u empirically
  - Higher u is better but this will leave less data to fit. i.e., large variance due to small sample size
  - Smaller u gives more data, but then the exceedances are not really extreme values, so it may violate the assumption of the Pickands-Balkema-de Haan theorem. i.e., large bias induced as we include more observations from the body of the dataset
  - Need to balance in this trade-off between bias and variance

(2) Estimate ξ, σ
  - Many estimators have been studied
  - The standard choice is MLE, but is known to be unstable for some parameter choices.
  - Estimation also affected by the choice of threshold u

Inspecting tail in practice

- The theory so far suggests that GPD can be used to fit the tail of the given dataset
- In practice, several graphical tools based on GPD properties are used to explore the dataset
  1. Double log quantile plot
  2. Mean excess plot
  3. Hill plot
  4. GPD index plot
- We consider two famous actual datasets
  - Danish Fire data: Insurance loss data collected at Copenhagen Reinsurance, comprising 2,167 fire losses between 1980 and 1990
  - SoA medical dataset: SOA Group Medical Insurance Large Claims Database, which records all the claim amounts exceeding $25,000 during 1991-1992
- `fExtremes` package in `R` is a good tool to use

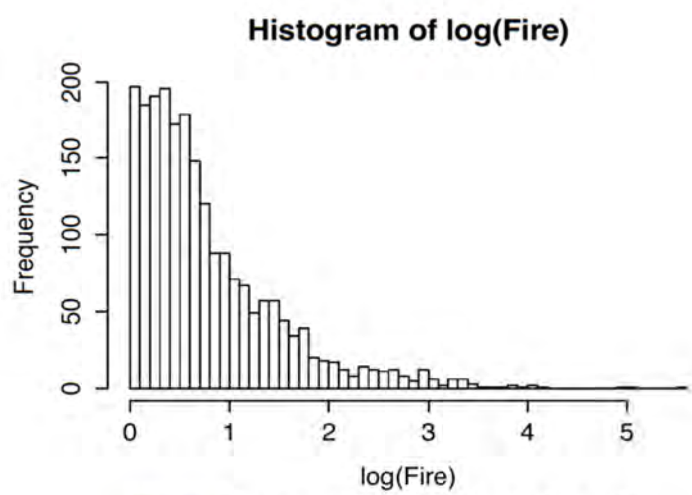

Figure: Histogram of Danish data in log scale

Double log quantile plot

- Log-log quantile plots

$$\Big(\ln x_i,\ \ln(1-F_n(x_i))\Big), \quad i=1,...,n. \qquad (8)$$

- Does not require any parameter estimation
- If the plot looks like a straight line, it indicates Pareto tail behaviour.
- The idea is from the Pareto distribution

$$F(x)=1-\left(\frac{x}{\sigma}\right)^{-\alpha}, \quad x\geq\sigma$$
$$\Rightarrow \ln(1-F(x))=-\alpha\ln x+\alpha\ln\sigma$$

- Hence this plot works best for pure Pareto data. But Pareto and GPD have similar tails, so it would reasonably work well for GPD tail as well.

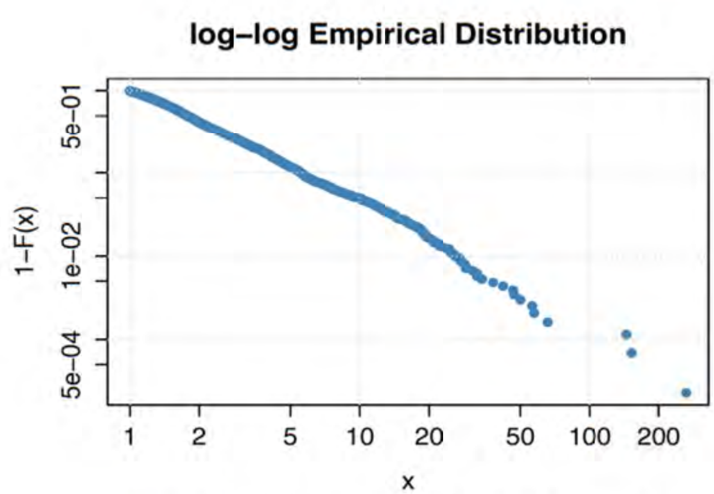

Figure: Double log plot (emdPlot in fExtremes)

Mean excess plot

- Mean excess function (MEF) is a well-known quantity in insurance pricing (e.g., reinsurance premium), defined as $e(u) = E(X - u|X > u)$.
- If X is GPD(σ, ξ), then $e(u)$ is a linear function of u:

$$e(u) = E(X - u|X > u) = \frac{\sigma}{1-\xi} + \frac{\xi}{1-\xi}u, \qquad (9)$$

  an important property of GPD
- Using the linearity, the plot

$$(u, e(u)) \qquad (10)$$

  should form a linear line on the plane for GPD
- The mean excess plot is the sample version of this with $e(u)$ replaced with $\hat{e}(u)$ which is the sample mean of the exceedances less u. So $\hat{e}(u)$ needs no parameter estimation

Mean excess plot

- Typically the plot looks like a (roughly) straight line from some large u value. Then this u may indicate the starting point of the GPD realm.
- The slope of the line indicates how heavy the tail is:
  1. If the slope is strictly positive, it indicates $\xi > 0$, a heavy tail
  2. If the slope is close to 0, it indicates $\xi = 0$, an exponential tail
  3. If the slope is negative, it indicates $\xi < 0$, a very short tail
- Mean excess plot has advantage over the log-log quantile plot because it uses a property of GPD directly.
- However the mean excess plot has some shortcomings:
  - As MEF is finite only if $E(X)$ is so, it can be defined only for $\xi < 1$. If $\xi \geq 1$, MEF is not useful.
  - The plot is generally sensitive to sampling error as it is a sample mean (not robust). In addition, it is highly volatile for large u, so it is not always clear how to read it

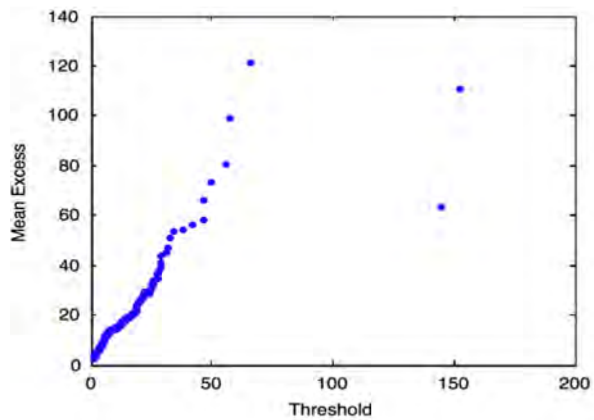

Figure: Sample MEF of Danish data

Hill (1975) plot

- An influential tool to find u and ξ from a given sample
- looks at the mean excess function of the log data
- For $X_{1,n} \geq ... \geq X_{n,n}$, the Hill estimator is

$$\hat{\alpha}_{k,n} = \left(\frac{1}{k} \sum_{j=1}^{k} \log X_{j,n} - \log X_{k,n} \right)^{-1}, \quad 2 \leq k \leq n, \qquad (11)$$

  and find the stable region in the Hill plot $\{(k, \hat{\alpha}_{k,n}^{-1})\}$, Here α is the tail index, corresponding to $1/\xi$ of GPD
- The Hill estimator is derived from asymptotic argument, but it turns out that it also is the (approx) MLE for the exponential distribution, which is the sample mean, when X is Pareto distributed.

Hill plot

- As the log variable of Pareto is exponentially distributed, the Hill estimator works best for Pareto datasets.
- However, once the underlying distribution departs from Pareto, its performance gets poorer and it becomes less clear what portion of the plot is most accurate.
- From the log-transformed data viewpoint, this means that the exponential distribution is inadequate to accommodate the shape of the log loss.

Hill plot

- There are alternative graphical methods to find ξ or α, and various variants of the Hill plot have been proposed
  - Pickand's estimator
  - Dekkers-Einmahl-de Haan estimator
  - Smoothed Hill plot, ...
- But there is no universally superior one, and the Hill plot seems most popular in practice
- Despite its wide use, the Hill plot (and others) still remains to be difficult to use due to substantial sampling variations in extreme sample quantiles.

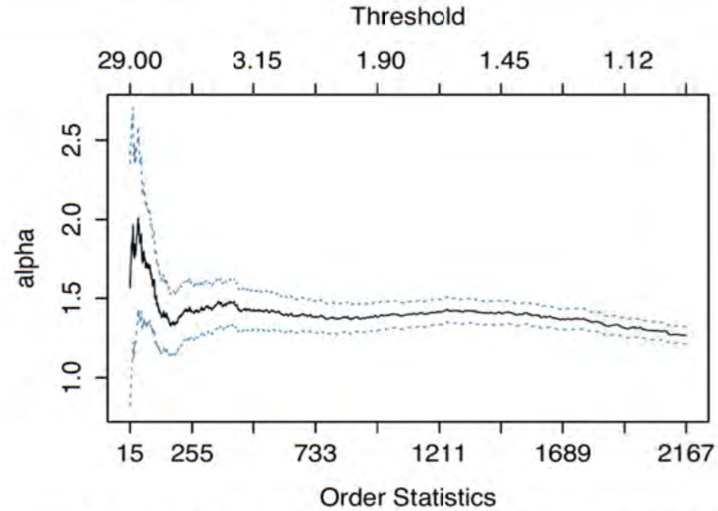

Figure: Hill plot of Danish data. Tail part is too volatile to see a pattern

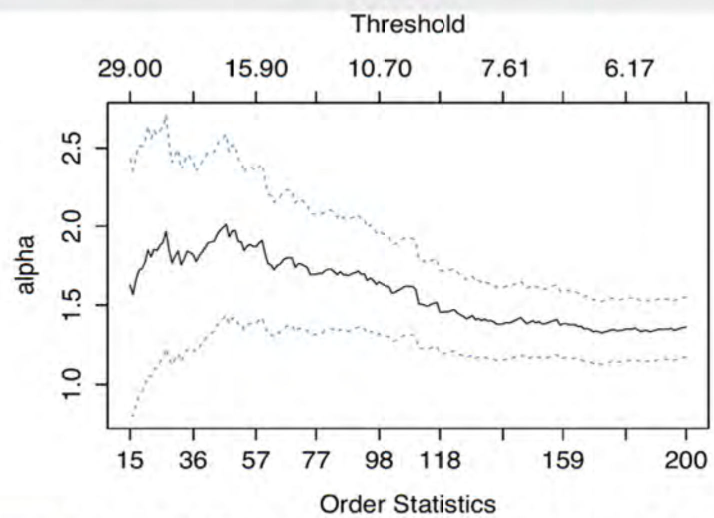

Figure: Hill plot of Danish data (Zoomed in). True α seems to be in [1.6, 2]

GPD index plot

- This plot uses the stability property of the GPD
- If X is $G_{\sigma,\xi}(x)$ distributed, its excess loss over d is again GPD distributed with parameter $(\sigma + \xi d, \xi)$. That is,

$$Pr(X - d < x | X > d) = 1 - \left(1 + \frac{\xi}{\sigma + \xi d} x\right)^{-1/\xi}. \qquad (12)$$

- This GPD stability resembles the memoryless property of the Exponential distn, a special case of the GPD with $\xi = 0$.
- The key is that ξ stays the same for all $d > 0$ values. We can use this property to find the threshold and the tail index.

GPD index plot

- Steps for the GPD index plot:
  - Assume that the GPD tail starts from u for a dataset
  - Estimate ξ using all exceedances larger than u using e.g., MLE
  - Increase threshold to $u' > u$. Then all exceedances over u' are again GPD distributed $\Rightarrow$ get another estimate for ξ
  - Repeat above precedure and obtain many ξ estimates
  - Draw plot of ξ estimates against various u values
- If u is indeed the true threshold, we will observe in the plot that estimated ξ would stabilize around u, because they should have the same ξ in theory.

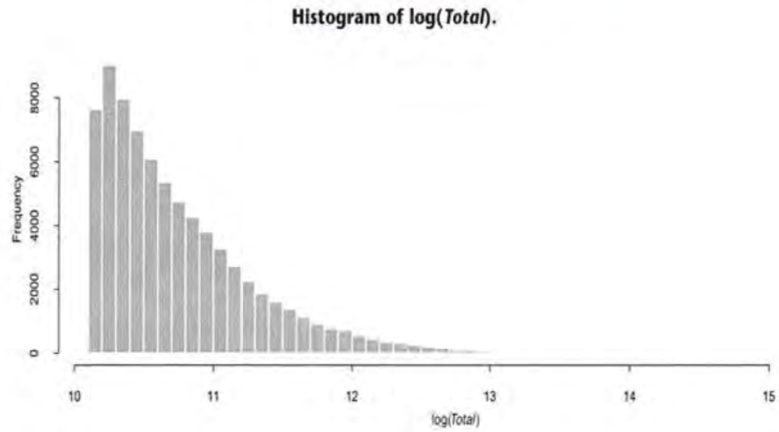

Figure: Histogram of SoA medical data in log scale

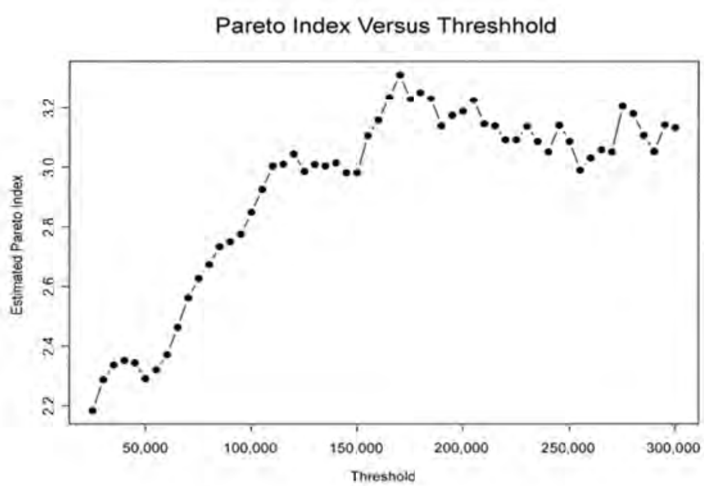

Figure: GPD index plot for SoA medical data (Cebrian et al., 2003)

Lesson

- These are basic tools to explore the dataset
- Finding the threshold is not easy. Its choice also affects the fitted GPD parameters.
- Some datasets are more volatile than others in the tail (extreme quintiles)
- Nonetheless, EVT has turned out to be successful in many applications (illustrated in the next example)
- Need experience and professional judgement

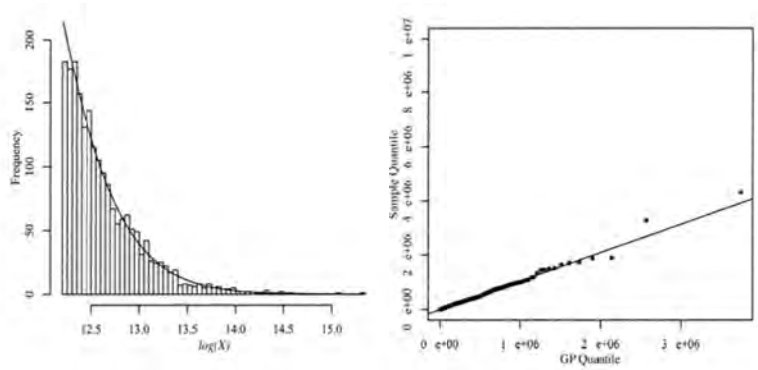

Figure: Tail fit with GPD for SoA medical data at $u = \$200,000$

GPD parameter estimation

- Once a certain u has been proposed as a threshold, we model the exceedances with GPD and estimate its two parameters
- As is true for many distributions, there are different estimators for GPD parameter in the literature
- MLE is the standard but has some known issues
- Others have other issues
- We will look at several estimators

Method of moment estimator (MME)

- Uses the first two moments of GPD
- The first two moments of GPD are

$$E(X) = \frac{\sigma}{(1-\xi)}, \quad \xi < 1, \tag{13}$$

and

$$Var(X) = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)}, \quad \xi < \frac{1}{2}. \tag{14}$$

- The MME are then obtained by

$$\hat{\sigma}_{MME} = \frac{1}{2}\bar{x}(\bar{x}^2/s^2+1), \quad \hat{\xi}_{MME} = \frac{1}{2}(1-\bar{x}^2/s^2), \tag{15}$$

- MME works only for $\xi < \frac{1}{2}$. It has an asymptotic normality only for $\xi < \frac{1}{4}$.

Probability weighted moments (PWM)

- Hosking and Wallis [1987] proposed to use

$$w_r := E[X(1+\xi X/\sigma)^r] = \frac{\sigma}{(1+r)(1+r-\xi)}, \quad r = 0, 1$$

- We can obtain then

$$\sigma = \frac{2w_0 w_1}{w_0 - 2w_1}; \qquad \xi = 2 - \frac{w_0}{w_0 - 2w_1}$$

- Replacing w_0 and w_1 with empirical moments, we get PWM estimator
- Works reasonably well for $0 \le \xi \le 1$, but not for other ranges

Maximum likelihood estimator (MLE)

- Suppose we have a sample $X_1, ..., X_n$ from GPD df $G_{\sigma,\xi}(x)$
- From its density the MLE of GPD can be obtained by maximizing the log-likelihood

$$\begin{aligned} l(\sigma, \xi) &= \sum_{i=1}^{n} \log g_{\sigma,\xi}(X_i) \\ &= -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log\left(1 + \xi \frac{X_i}{\sigma}\right), \end{aligned}$$

  subject to constraints $\sigma > 0$ and $1 + \xi X_i/\sigma > 0$ for all i.
- Partial derivative of $l(\sigma, \xi)$ yields the MLE of GPD
- The MLE is not analytically available; numerical methods needed.

Maximum likelihood estimator (MLE)

- If we let $\hat{\theta} = (\hat{\sigma}, \hat{\xi})'$ be the MLE of GPD(σ, ξ), we have
$$\sqrt{n}(\hat{\theta} - \theta) \overset{d}{\sim} BVN(\mathbf{0}, [I_1(\theta)]^{-1}), \qquad (16)$$
where the covariance matrix
$$[I_1(\theta)]^{-1} = \begin{bmatrix} 2\sigma^2(1+\xi) & -\sigma(1+\xi) \\ -\sigma(1+\xi) & (1+\xi)^2 \end{bmatrix}, \quad \xi > -0.5. \qquad (17)$$
- For any given $\xi < -1$, we see that
$$\lim_{-\sigma/\xi \to y_{(n)}^{+}} l(\sigma, \xi) = \infty, \qquad (18)$$
implying that there is no finite MLE.
- In summary, the MLE exists only when $\xi > -1$, and its asymptotic normality further requires $\xi > -0.5$
- It is also known that MLE sometimes fails to numerically converge even when $\xi > -0.5$, which is troublesome

Other estimators for GPD

- Zhang's LME and its generalization (Zhang 2007, 2010)
- NLS by Song and Song (2012) and its weighted generalization by Park and Kim (2016)
- EPM of by Castillo and Hadi (1997)
- Phase-type approach by Kim and Kim (2015)
- And more ...

Risk measures under POT

- Modern financial risk management often requires accurate estimation of some tail risk measures of the given loss dataset for solvency and risk analyses purposes
- Most popular such risk measures are VaR and CTE
- The VaR of a rv X at the $100p\%$ level is the $100p$ quantile of the distribution of X, denoted by

$$Q_p(X) = {F_X}^{-1}(p). \tag{19}$$

- However, VaR is sometimes criticized for lacking the subadditivity property of the coherent risk measure axioms (Artzner et al.,1999)
- In this regard, the CTE has received much attention as an alternative, coherent tail risk measure.

- The CTE measures the average severity of the loss when the extreme loss does occurs, where the extreme loss is represented by the VaR. Formally, the CTE of X at $100p\%$ level is defined as

$$CTE_p(X) = E[X|X > Q_p(X)]. \tag{20}$$

- As the tail of an arbitrary dataset can be modelled via GPD, we may compute these risk measures under the POT approach
- For this one writes the underlying loss distribution as a combination of the body and tail parts:
    - Body part uses the def $F_n(x)$
    - Tail part modelled by the GPD $G_{\sigma,\xi}(x)$

- Formally, we let $F(x)$ be the distribution function of an arbitrary continuous distribution, and define the exceedance distribution of loss events over u by

$$F_u(y) = \Pr(X - u \le y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}, \qquad (21)$$

which is assumed to be GPD(ξ, σ).

- Then we can rewrite $F(x)$ as

$$\begin{aligned} F(x) = P(X \le x) &= (1 - F(u))F_u(x - u) + F(u) \\ &= (1 - F(u))G_{\xi,\sigma}(x - u) + F(u) \\ &= (1 - F(u))G_{\xi,u,\sigma}(x) + F(u), \qquad (22) \end{aligned}$$

using the three-parameter GPD, and its estimated version by

$$\hat{F}(x) = (1 - F_n(u))G_{\hat{\xi},u,\hat{\sigma}}(x) + F_n(u), \qquad (23)$$

where F_n is the EDF and $\hat{\xi}, \hat{\sigma}$ are the estimated GPD parameters from the sample.

- Assuming that the tail risk measures lie in the GPD realm, the VaR estimate is obtained from inverting (23) as

$$\widehat{Q}_p(X) = G^{-1}_{\hat{\xi},u,\hat{\sigma}}\left[1 - \frac{1-p}{1 - F_n(u)}\right] = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(\frac{n}{n_u}(1-p)\right)^{-\hat{\xi}} - 1\right],$$

where n_u is the number of exceedances over the threshold u

- The CTE is obtained as, subject to $\xi < 1$ for the mean to be finite, for $Q_p(X) > u$,

$$\widehat{CTE}_p(X) = \frac{\widehat{Q}_p(X) + \hat{\sigma} - \hat{\xi}u}{1 - \hat{\xi}} = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[\frac{1}{1-\hat{\xi}}\left(\frac{n}{n_u}(1-p)\right)^{-\hat{\xi}} - 1\right], \quad \xi < 1.$$

Thank you

- Topics not covered
  - Time series
  - Dependence
- References

  - Embrechts, P., Kluppelberg, C. and T. Mikosch. *Modelling extremal events*. Springer, New York, 1997.
  - Coles, S. *An introduction to statistical modeling of extreme values*. Springer, 2001.
  - Cebrian, A.C. and Denuit, M. and Lambert, P. *Generalized pareto fit to the society of actuaries' large claims database*. North American Actuarial Journal, 2003
  - Kim, J.H.T. and Kim, J. *A parametric alternative to the Hill estimator for heavy-tailed distributions*. Journal of Banking & Finance, 2015
  - M. H. Park and J.H.T Kim. Estimating extreme tail risk measures with generalized pareto distribution. Computational Statistics & Data Analysis, 2016.

- Software: R package `fExtremes`