Extreme Value Theory

MFM Practitioner Module: Quantitative Risk Management

John Dodson

November 13, 2019

Maxima

The *n*-block maxima[1] is a random variable defined as

$$M_n \triangleq \max\left(X_1, \ldots, X_n\right)$$

for i.i.d. random variables X_i with distribution function $F(\cdot)$. We are interested in $n \to \infty$. If there exists a sequence of normalizations of M_n (with $c_n > 0$) such that

$$\lim_{n\to\infty} F^n\left(c_n x + d_n\right)$$

converges to a non-degenerate distribution function, $H(x)$, then

$$H(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right) & \xi > 0 \text{ and } x \geq -1/\xi \\ \exp\left(-e^{-x}\right) & \xi = 0 \\ \exp\left(-(1+\xi x)^{-1/\xi}\right) & \xi < 0 \text{ and } x < -1/\xi \end{cases}$$

[1] Use $1 - F(-x)$ and $-\max\left(-X_1, \ldots, -X_n\right)$ for the minima.

Maxima

This remarkable result, the Fisher–Tippett–Gnedenko theorem (1927–28/1943), is analogous to the central limit theorem for an appropriately normalized $S_n \triangleq \sum_{i=1}^n X_i$:

$$\lim_{n\to\infty}\left(\frac{1}{\sqrt{n}}S_n - \sqrt{n}\mu\right) \sim \mathcal{N}\left(0,\sigma^2\right)$$

Generalized Extreme Value Distribution

$H(\cdot)$ from above is called the generalized extreme value distribution and it has a single parameter ξ.

Types

The GEV is continuous in ξ for each x, but its sign has can be used as a classifier

- $\xi > 0$ is the Fréchet with finite moments to order $1/\xi$.
- $\xi = 0$ is the Gumbel with finite moments of all orders.
- $\xi < 0$ is the Weibull with a finite right endpoint.

Maxima

The GEV result is about i.i.d. sequences of r.v.'s. We have seen that timeseries of innovations of financial invariants are independent, but not generally identically-distributed.

Stationary Time Series

If the normalized n-block maxima of the associated strict white noise for a stationary time series process has a limiting distribution $H(\cdot)$ in one of the GEV classes, then there exists $0 < \theta \leq 1$ such that the limit of the normalized n-block maxima of the innovations is $H^\theta(\cdot)$.

- In particular, the n-block maxima of the innovations can be re-normalized to yield the same GEV distribution as the associated white noise with the same ξ.
- Effectively $\tilde{n} = \theta n$ in the limit, which can be thought of as representing clustering in the extreme values of the innovations for $\theta < 1$.

Threshold Exceedances

Typically we are not as interested in the n-block maxima as we are in the relative frequency of a range of extreme outcomes. A result in this regard leads to the generalized Pareto distribution that we have already seen.

Generalized Pareto

The distribution function with scale parameter $\beta > 0$ is

$$G(x) = \begin{cases} 1 - (1 + \xi\, x/\beta)^{-1/\xi} & \xi > 0 \text{ and } x \geq 0 \\ 1 - e^{-x/\beta} & \xi = 0 \text{ and } x \geq 0 \\ 1 - (1 + \xi\, x/\beta)^{-1/\xi} & \xi < 0 \text{ and } 0 \leq x < -\beta/\xi \end{cases}$$

Excess Distribution

If r.v. X has distribution $F(\cdot)$, the excess distribution is

$$F_\eta(x) \triangleq \mathrm{P}\left[X - \eta \leq x | X > \eta\right] = \frac{F(x+\eta) - F(\eta)}{1 - F(\eta)}$$

Threshold Exceedances

Pickands–Balkema–de Haan theorem (1974–75)

Iff X is GEV with parameter ξ and right endpoint x_F, then there exists $\beta(\eta) > 0$ such that

$$\lim_{\eta \to x_F} \sup_{0 \leq x < x_F - \eta} \left| F_\eta(x) - G(x; \xi, \beta(\eta)) \right| = 0$$

That is, as the threshold level is raised, the excess distribution becomes arbitrarily close to a generalized Pareto distribution with the same shape parameter as the GEV.

Mean Excess

Note that in the limit $\eta \to x_F$, $\beta(\eta)$ becomes linear. Since $\mathrm{E}\left[X - \eta | X > \eta\right] = \beta(\eta)/(1 - \xi)$, the mean excess also becomes linear in the threshold for $\xi < 1$.

Estimators

Generalized Extreme Value

It is possible to estimate the shape of the limiting GEV from a sample of n-block maxima of i.i.d. data using maximum likelihood, but this is very inefficient, since you effectively reduce all but a fraction $1/n$ of your data to partial ranks.

Generalized Pareto

In the GP setting you can use a higher fraction of the data depending on how you choose the threshold; and we can make use of results such as the limiting linearity of the mean excess above to help set it.

Exceedance Point Process

In a strict white noise setting, the interval between threshold exceedances is an exponential r.v. in the threshold limit, so we can use even more of the data taking that into account.

GP Maximum Likelihood

We have already worked with the MLE for GP. In this setting, we form the log-likelihood of the GP approximation of the threshold excess

$$\log L(\xi, \beta) = \\ -N_\eta \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{N_\eta} \log\left(1 + \xi \frac{x_{(N-N_\eta+i)} - \eta}{\beta}\right)$$

where N_η is the number of observations that exceed η and the N observations are indexed by their ranks. The estimator is of course

$$\left(\hat{\xi}, \hat{\beta}\right) = \arg\max_{(\xi,\beta)} \log L(\xi, \beta)$$

Hill Estimator

In the Fréchet class, for $\xi > 0$, the distribution function has a tail of the form $\bar{F}(x) = x^{-1/\xi} L(x)$ for some slowly varying function $L(\cdot)$.

Hill Estimator

The Hill estimator is based on the observation about the mean excess of the logarithm of X.

$$\mathrm{E}\left[\log X - \log \eta | X > \eta\right] \approx \frac{L(\eta)\eta^{-1/\xi}\xi}{\bar{F}(\eta)} \approx \xi$$

in the limit $\eta \to \infty$ from Karamata's theorem. The estimator is hence

$$\hat{\xi}_k^{(H)} = \frac{1}{k} \sum_{i=1}^{k} \log x_{(n-k+i)} - \log x_{(n-k)}$$

in terms of the observations indexed by their ranks.

Points over Thresholds

In this model, the order of the data matter. Let $t = i/n$ for $i = 1, \ldots, n$. Define a state space $\mathcal{X}$ for (t, x). The marked point process defined by X exceeding some high threshold η before t has intensity rate

$$\tau(x) \triangleq H_{\xi,\mu,\sigma}(x)$$

where σ, μ represent c_n, d_n, and excess magnitude

$$\bar{F}_\eta(x) = \bar{G}_{\xi,\beta}(x)$$

for scale $\beta = \sigma + \xi(\eta - \mu)$.

Estimator

The likelihood function for this model is given by

$$\log L(\xi, \sigma, \mu) = \log L_{GP}(\xi, \sigma; x - \eta) - \tau + N_\eta \log \tau$$

hence $\hat{\tau} = N_\eta$; and from the MLE of the GP, we can infer $\hat{\mu}, \hat{\sigma}$.

Generalized Pareto Loss Distributions

The convergence of excess loss to a GP random variable can be used to calculate VaR and ES. In particular, we can set the (right) tail mass θ to some sufficiently small value, then

$$F_L(x) \approx \begin{cases} ? & x < \eta \\ 1 - \theta \left(1 + \xi \frac{x-\eta}{\beta}\right)^{-1/\xi} & (\xi > 0 \text{ and } x \geq \eta) \text{ or} \\ & (\xi < 0 \text{ and } \eta \leq x < \eta - \frac{\beta}{\xi}) \\ 1 & \xi < 0 \text{ and } x \geq \eta - \frac{\beta}{\xi} \end{cases}$$

and we can invert this to get

$$\text{VaR}_\alpha \approx \eta + \frac{\beta}{\xi}\left(\left(\frac{\theta}{1-\alpha}\right)^{\xi} - 1\right) \quad \text{for } 1 - \theta \leq \alpha < 1$$

Generalized Pareto Loss Distributions

Furthermore, we can integrate the quantile function to get expected shortfall (for $\xi < 1$),

$$\mathrm{ES}_\alpha \approx \eta + \frac{\beta}{\xi}\left(\frac{1}{1-\xi}\left(\frac{\theta}{1-\alpha}\right)^{\xi} - 1\right) \quad \text{for } 1-\theta \leq \alpha < 1$$

Recalling our previous comparison of VaR and ES in terms of the Cornish-Fisher moment expansion, we again see that there is a fundamental equivalence to the extent that $\xi \approx 0$.

$$\lim_{\alpha \to 1} \frac{\mathrm{ES}_\alpha}{\mathrm{VaR}_\alpha} = \frac{1}{1-\xi}$$