Extreme Value Theory

The extreme value theory (EVT) is designed to model very large tails. This is known as black swans or rare events. The basic black swan or rare event story is that black swans or rare events are not really rare after all. In other words, tail probabilities (probabilities for very large losses) are not as small as people think.

There are fat tail distributions, such as t or Cauchy to better fit the real data. Yet a more common method is to use EVT with the Gaussian distribution. Thanks to the work by Gnedenko (1943)[1] who finds that the tails of many distributions share common properties.

Let $F(u)$ be the probability (or cumulative density function) of a random variable $x \le u$. Or formally,

$$F(u) = \Pr[x \le u] = \int_{-\infty}^{u} f(x)dx$$

where $f(x)$ is the density function. Then the probability of $x > u$ is $1 - F(u)$. Also, the probability of $u < x < u + \varepsilon$ is $F(u+\varepsilon) - F(u)$. Hence the conditional probability of $u < x < u + \varepsilon$ on $x > u$ is:

$$F_u(\varepsilon) = \frac{F(u+\varepsilon) - F(u)}{1 - F(u)}$$

Gnedenko argues that this (conditional) tail distribution has the same properties regardless of what the actual statistical distribution is (within a set of allowable distributions which includes the Gaussian distribution). In particular, the following result holds:

$$\lim_{u \to v} \sup_{0 \le \varepsilon \le v-u} \left| F_u(\varepsilon) - G_{\xi,\beta,u}(\varepsilon) \right| = 0$$

where

$$G_{\xi,\beta}(\varepsilon) = \begin{cases} 1 - \left(1 + \xi \dfrac{\varepsilon}{\beta}\right)^{-1/\xi} & \xi \ne 0 \\ 1 - \exp\left(-\dfrac{\varepsilon}{\beta}\right) & \xi = 0 \end{cases}$$

[1] See D.V. Gnedenko, 1943, "Sur la distribution limité du terme d'une serie aléatoire," Ann. Math. 44: 423-453.

is the generalized Pareto distribution. In words, for any given $v > u$, and the condition that $0 \le \varepsilon \le v - u$, the (conditional) tail probability $F_u(\varepsilon)$ will converge to $G_{\xi,\beta}(u+\varepsilon)$ regardless of the distribution.

Maximum Likelihood Estimation of ξ and β

To perform the maximum likelihood estimation, we need the density of the generalized Pareto distribution:

$$\begin{aligned} g_{\xi,\beta}(y) &= \frac{dG_{\xi,\beta}(y)}{dy} \\ &= \frac{1}{\beta}\left(1 + \frac{\xi y}{\beta}\right)^{-1/\xi - 1} \end{aligned}$$

To perform the MLE, we first note that, given this is a tail distribution, we can only take observations of the "tail". Hence to define a tail, we must first choose a value of u. It will become clear later that this u value is ideally the critical value of a normal (Gaussian) distribution. For example, the parametric VaR (i.e. normal distribution) at 5% is –6.8543% (see Section 4.2.2 where mean of FB is –0.205% and standard deviation is 0.0404) which is a good choice of $u = +6.8543\%$. Note that in VaR reports, all negative numbers are flipped to be positive (more intuitive that way).

Hence we need to use only those observations (returns) that are less than -6.8543%. To facilitate the estimation (and not to be confused about the signs), it is easier that we turn all negative returns into positive numbers and also u. This way, we deal with numbers that are larger than the positive u.

Assuming that there are there are n_u such observations, the likelihood function is (let $y = x - u$):

$$f(x_1, \cdots, x_n) = \prod_{i=1}^{n_u} \frac{1}{\beta}\left(1 + \frac{\xi(x_i - u)}{\beta}\right)^{-1/\xi - 1}$$

or

$$\ln f(x_1, \cdots, x_n) = \sum_{i=1}^{n} \ln\left[\frac{1}{\beta}\left(1 + \frac{\xi(x_i - u)}{\beta}\right)^{-1/\xi - 1}\right]$$

The ML estimators cannot be derived in closed form. Hence, we must solve numerically for ξ and β by maximizing the likelihood function.

VaR

Once the parameters are estimated, we can then compute EVT-adjusted VaR. The choice of the distribution for EVT is Gaussian (normal). This is because the Gaussian tail probability follows EVT. As a result, we can add probability to the tail of the Gaussian distribution.

Just now, we know that:

$$F_u(\varepsilon) = \frac{F(u+\varepsilon) - F(u)}{1 - F(u)}$$
$$\sim G_{\xi,\beta}(\varepsilon)$$

This says that conditional on $x > u$, the probability of x between $u + \varepsilon$ and u is $G_{\xi,\beta}(u+\varepsilon)$. Note that conditional on $x > u$, the probability of $u < x < u + \varepsilon$ is the same as the probability of $x < u + \varepsilon$, because the other condition ($u < x$) is redundant. Hence, the probability of $x > u + \varepsilon$ conditional on $x > u$ is simply $1 - G_{\xi,\beta}(\varepsilon)$.

Let $u^* > u$ (or equivalently ε does not need to be small). Then it is clear that the probability of $x > u^*$ conditional on $x > u$ is $1 - G_{\xi,\beta}(u^* - u)$. That is,

$$\Pr[x > u^* \mid x > u] = 1 - \Pr[x < u^* \mid x > u]$$
$$= 1 - G_{\xi,\beta}(u^* - u)$$
$$= \left(1 + \xi\frac{(u^* - u)}{\beta}\right)^{-1/\xi}$$

Then the unconditional probability is:

$$\Pr[x > u^*] = [1 - F(u)]\left(1 + \xi\frac{(u^* - u)}{\beta}\right)^{-1/\xi}$$

Now, we approximate the right-tail probability $1 - F(u)$ by the frequency count of the sample: n_u / n where n is the total number of observations and n_u is (as defined earlier) the number of the observations in the tail (e.g. any return that is more negative than –6.8543%).

Hence,

$$\Pr[x > u^*] = \frac{n_u}{n}\left(1 + \xi\frac{(u^* - u)}{\beta}\right)^{-1/\xi}$$

Now, we flip the right tail to the left tail. For an arbitrary (usually fat tailed) distribution, we would like to find the VaR number by inverting the distribution function:

$$\text{VaR}_\alpha = F^{-1}(\alpha)$$

For example, if $\alpha = 5\%$ and $F(u)$ is normal then $\mathrm{VaR}_{5\%} = +1.645 \times \sigma$. Now with EVT, we have:

$$\alpha = \frac{n_u}{n}\left(1 + \xi \frac{(u^* - u)}{\beta}\right)^{-1/\xi}$$

Solving for u^*, we obtain:

$$u^* = u + \frac{\beta}{\xi}\left[\left(\frac{n}{n_u}\alpha\right)^{-\xi} - 1\right]$$

Recall that we select u from the normal distribution (i.e. $u = +1.645 \times \sigma$). Hence u^* reflects the same risk (i.e. α) under EVT. In other words, EVT adds to the Gaussian VaR value. Note that u and u^* are positive.

The intuition of the equation is very nice. To gauge a fail tail, u^* must be great than u. Hence,

$$\frac{\beta}{\xi}\left[\left(\frac{n}{n_u}\alpha\right)^{-\xi} - 1\right] > 0$$

Given that both β and ξ are greater than 0, it must be the case that:

$$\left(\frac{n}{n_u}\alpha\right)^{-\xi} > 1 \quad \text{or} \quad \left(\frac{n_u}{n}\frac{1}{\alpha}\right)^{\xi} > 1$$

Then

$$\frac{n_u}{n} > \alpha$$

in order to have a fat tail.

From Hull, $\beta = 32.532$ and $\xi = 0.436$. Now we can compute the EVT-adjusted VaR:

Hull (14.8)

$$\Pr[v > x] = \frac{n_u}{n}\left(1 + \xi \frac{x - u}{b}\right)^{-1/\xi}$$

$$\Pr[v > 300] = \frac{22}{500}\left(1 + 0.436 \times \frac{300 - 160}{32.532}\right)^{-1/0.436}$$

$$= 0.0039$$

as when $u = 160$, $n_u = 22$.

Hull (14.9)

$$
\begin{aligned}
u^* &= u + \frac{\beta}{\xi}\left[\left(\frac{n}{n_u}\alpha\right)^{-\xi} - 1\right] \\
&= 160 + \frac{32.532}{0.436}\left[\left(\frac{500}{22}\times 0.01\right)^{-0.436} - 1\right] \\
&= 227.8
\end{aligned}
$$

Expected Shortfall

The expected shortfall (ES) is regarded as a better measure of tail risk than VaR. It is defined as the expected loss conditional on a tail probability. Formally,

$$
\begin{aligned}
\mathrm{ES}_\alpha &= \mathbb{E}[x \mid x > u] \\
&= \int_{-\infty}^{\infty} x f(x \mid x > u) dx
\end{aligned}
$$

Under a log-normal distribution for x (i.e. Black-Scholes) where $\mathbb{E}[\ln x] = \mu$ and $\mathbb{V}[\ln x] = \sigma^2$, we have:

$$
\begin{aligned}
\mathrm{ES}_\alpha &= \int_{-\infty}^{\infty} x f(x \mid x > u) dx \\
&= \frac{1}{1 - N(d_-)} \int_u^{\infty} x f(x) dx \\
&= \frac{1}{1 - N(d_-)} \mathbb{E}[x] N(d_+) \\
&= \frac{N(d_+)}{1 - N(d_-)} \mathbb{E}[x]
\end{aligned}
$$

where (note that $N(d_\pm) = \Pr^{\pm}[x > u]$ which is ITM probability)

$$
d_\pm = \frac{\ln(x / u) + (\mu \pm ½\sigma^2)}{\sigma}
$$

Under EVT, the ES is:

$$
\mathrm{ES}_\alpha = \frac{u^* + \beta - \xi u}{1 - \xi}
$$