Financial Econometrics and Volatility Models
Extreme Value Theory

Eric Zivot

May 3, 2010

1 Lecture Outline

- Modeling Maxima and Worst Cases

- The Generalized Extreme Value Distribution

- Modeling Extremes Over a High Threshold
  - Generalized Pareto Distribution
  - Traditional Risk Measures

2 Reading

- MFTS, chapter 5
- FMUND, chapter 7

3 Modeling Maxima and Worst Cases

- *Example*: Analysis of daily negative returns on S&P 500 index

- What is the probability that next year's annual maximum negative return exceeds all previous negative returns? In other words, what is the probability that next year's maximum negative return is a new *record*?

- What is the 20-year *return level* of the negative returns? That is, what is the negative return which, on average, should only be exceeded in one year every twenty years?

To answer these questions, the distribution of extreme negative returns on the S&P 500 index is required.

4 The Generalized Extreme Value Distribution

Let $X_1, X_2, \ldots$ be iid random variables representing risks or losses with an unknown CDF $F(x) = \Pr\{X_i \leq x\}$. Define $M_n = \max\left(X_1, \ldots, X_n\right)$ as the worst-case loss in a sample of n losses. From the iid assumption, the CDF of M_n is

$$\begin{aligned}\Pr\{M_n &\leq x\} = \Pr\{X_1 \leq x, \ldots, X_n \leq x\} \\ &= \prod_{i=1}^{n} F(x) = F^n(x)\end{aligned}$$

Comments:

- The empirical distribution function is often a very poor estimator of $F^n(x)$
- $F^n(x) \to$ 0 or 1 as $n \to \infty$

4.1 Fisher-Tippet Theorem

Define the standardized maximum value

$$\begin{aligned} Z_n &= \frac{M_n - \mu_n}{\sigma_n} \\ \sigma_n &> 0 = \text{ scale parameter} \\ \mu_n &= \text{ location parameter} \end{aligned}$$

Fisher-Tippet Theorem: If the standardized maximum converges to some non-degenerate distribution function, it must be a *generalized extreme value* (GEV) distribution of the form

$$H_\xi(z) = \begin{cases} \exp\left\{-(1+\xi z)^{-1/\xi}\right\} & \xi \neq 0,\ 1+\xi z > 0 \\ \exp\{-\exp(-z)\} & \xi = 0,\ -\infty \leq z \leq \infty \end{cases}$$

Comments

- The CDF F of the underlying data is in the *domain of attraction* of H_ξ.
- The Fisher-Tippet Theorem is the analog of the *Central Limit Theorem* for extreme values.
- ξ is a *shape* parameter and determines the tail behavior of H_ξ.
- The parameter $\alpha = 1/\xi$ is called the *tail index* if $\xi > 0$.

4.1.1 GEV Types

Result: The tail behavior of the distribution F of the underlying data determines the shape parameter ξ of the GEV distribution

- If the tail of F declines exponentially, then H_ξ is of the *Gumbel* type and $\xi = 0$. Distributions in the domain of attraction of the Gumbel type are *thin tailed* distributions such as the normal, log-normal, exponential, and gamma. For these distributions, all moments usually exist

- If the tail of F declines by a power function, i.e.

$$1 - F(x) = c \cdot x^{-1/\xi} = x^{-\alpha}$$

for some constant c, then H_ξ is of the *Fréchet* type and $\xi > 0$. Distributions in the domain of attraction of the Fréchet type include *fat tailed* distributions like the Pareto, Cauchy, Student-t, alpha-stable with characteristic exponent in $(0, 2)$, as well as various mixture models. Not all moments are finite for these distributions: $E[X^k] = \infty$ for $k \geq \alpha = 1/\xi$. This type is most relevant for financial data.

- if the tail of F is finite then H_ξ is of the *Weibull* type and $\xi < 0$. Distributions in the domain of attraction of the Weibull type include distributions with bounded support such as the uniform and beta distributions. All moments exist for these distributions.

4.1.2 Unstandardized Distributions

For location and scale parameters μ and $\sigma > 0$ we have

$$H_\xi(z) = H_\xi\left(\frac{x-\mu}{\sigma}\right) = H_{\xi,\mu,\sigma}(x)$$

For large enough n

$$\mathsf{Pr}\left\{Z_n < z\right\} = \mathsf{Pr}\left\{\frac{M_n - \mu_n}{\sigma_n} < z\right\} \approx H_\xi(z)$$

Letting $x = \sigma_n z + \mu_n$ then

$$\mathsf{Pr}\{M_n < x\} \approx H_\xi\left(\frac{x-\mu_n}{\sigma_n}\right) = H_{\xi,\mu_n,\sigma_n}(x)$$

This result is used in practice to make inferences about the maximum loss M_n.

4.2 Maximum Likelihood Estimation

Let $X_1, \ldots, X_T$ be iid losses with unknown CDF F and let M_T denote the sample maximum. Divide the sample into m non-overlapping blocks of essentially equal size $n = T/m$

$$
\begin{aligned}
& [X_1, \ldots, X_n | X_{n+1}, \ldots, X_{2n} | \ldots | X_{(m-1)n+1}, \ldots, X_{mn}] \\
M_n^{(j)} &= \text{maximum value of } X_i \text{ in block } j = 1, \ldots, m
\end{aligned}
$$

- Construct likelihood for ξ, σ_n and μ_n from

$$\{M_n^{(1)}, \ldots, M_n^{(m)}\}$$

- Assumed that the block size n is sufficiently large so that the Fisher-Tippet Theorem holds.

The log-likelihood function for $\xi \neq 0$ is

$$
\begin{aligned}
l(\mu, \sigma, \xi) \;=\; & -m \ln(\sigma) \\
& -(1 + 1/\xi) \sum_{i=1}^{m} \ln \left[1 + \xi \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) \right] \\
& - \sum_{i=1}^{m} \left[1 + \xi \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) \right]^{-1/\xi}
\end{aligned}
$$

and is maximized imposing the constraint

$$
1 + \xi \left(\frac{M_n^{(i)} - \mu}{\sigma} \right) > 0
$$

Remarks:

- For $\xi > -0.5$ the mles for μ, σ and ξ are consistent and asymptotically normally distributed with asymptotic variance given by the inverse of the observed information matrix

- The bias of the mle is reduced by increasing the block size n, and the variance of the mle is reduced by increasing the number of blocks m.

4.3 Example: S&P 500 negative returns

The maximum likelihood estimates of ξ, μ and σ based on annual block maxima are

$$\begin{aligned} \hat{\xi} &= 0.334,\ \widehat{SE}(\hat{\xi}) = 0.208 \\ \hat{\mu} &= 1.975,\ \widehat{SE}(\hat{\mu}) = 0.151 \\ \hat{\sigma} &= 0.672,\ \widehat{SE}(\hat{\sigma}) = 0.131 \end{aligned}$$

What is the probability that next year's annual maximum negative return exceeds all previous negative returns?

$$\begin{aligned} &\Pr\left(M_{260}^{(29)} > \max\left(M_{260}^{(1)}, \ldots, M_{260}^{(28)}\right)\right) = \Pr(M_{260}^{(29)} > 6.68) \\ = \ & 1 - H_{\hat{\xi},\hat{\mu},\hat{\sigma}}(6.68) = 0.0267 \end{aligned}$$

4.4 Return Level

The k n-block *return level*, $R_{n,k}$, is that level which is exceeded in one out of every k blocks of size n. That is, $R_{n,k}$ is the loss value such that

$$\Pr\{M_n > R_{n,k}\} = 1/k$$

$R_{n,k}$ is simply the $1 - 1/k$ quantile of limiting GEV distribution:

$$R_{n,k} \approx H^{-1}_{\xi,\mu,\sigma}(1 - 1/k) = \mu - \frac{\sigma}{\xi}\left(1 - (-\log(1 - 1/k))^{-\xi}\right)$$

5 Modeling Extremes Over High Thresholds

Idea: Modeling only block maxima data is inefficient if other data on extreme values are available. A more efficient alternative approach that utilizes more data is to model the behavior of extreme values above some high threshold. This method is often called *peaks over thresholds* (POT).

- An advantage of the POT approach is that common risk measures like *Value-at-Risk* (VaR) and *expected shortfall* (ES) may easily be computed

5.1 Risk Measures

Value-at-Risk (VaR). For $0.95 \leq q < 1$, say, VaR_q is the qth quantile of the loss distribution F

$$VaR_q = F^{-1}(q)$$

where F^{-1} is the inverse of F.

Expected Shortfall (ES). ES_q is the expected loss size, given that VaR_q is exceeded:

$$ES_q = E[X|X > VaR_q]$$

Note: ES_q is related to VaR_q via

$$ES_q = VaR_q + E[X - VaR_q|X > VaR_q]$$

5.2 The Generalized Pareto Distribution

Let $X_1, X_2, \ldots$ be a sequence of iid random losses with an unknown CDF F and let $M_n = \max\{X_1, \ldots, X_n\}$. A natural measure of extreme events are values of the X_i that exceed a high threshold u. Define the excedences above the high threhold u:

$$Y = X - u, \; X > u$$

Then the *excess distribution* (aka tail distribution) above the threshold u is the conditional probability

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

for $y > 0$.

Result: For the class of distributions F such that the CDF of the standardized value of M_n converges to a GEV distribution, for large enough u there exists a positive function $\beta(u)$ such that $F_u(y)$ is well approximated by the *generalized Pareto distribution* (GPD)

$$G_{\xi,\beta(u)}(y) = \begin{cases} 1-(1+\xi y/\beta(u)) & \text{for } \xi \neq 0 \\ 1-\exp(-y/\beta(u)) & \text{for } \xi = 0 \end{cases}, \quad \beta(u) > 0$$

defined for $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\beta(u)/\xi$ when $\xi < 0$

Remarks:

- For a sufficiently high threshold u, $F_u(y) \approx G_{\xi,\beta(u)}(y)$ for a wide class of loss distributions F. To implement this result, the threshold value u must be specified and estimates of the unknown parameters ξ and $\beta(u)$ must be obtained.

- There is a close connection between the limiting GEV distribution for block maxima and the limiting GPD for threshold excesses. The shape parameter ξ of the GEV distribution is the same shape parameter ξ in the GPD and is independent of the threshold value u.

- Consider a limiting GPD with shape parameter ξ and scale parameter $\beta(u_0)$ for an excess distribution F_{u_0} with threshold u_0. For an arbitrary threshold $u > u_0$, the excess distribution F_u has a limiting GPD distribution with shape parameter ξ and scale parameter $\beta(u) = \beta(u_0) + \xi(u - u_0)$. Alternatively, for any $y > 0$ the excess distribution F_{u_0+y} has a limiting GPD distribution with shape parameter ξ and scale parameter $\beta(u_0) + \xi y$

5.3 Mean Excess Function

Suppose the threshold excess $X - u_0$ follows a GPD with parameters $\xi < 1$ and $\beta(u_0)$. Then the *mean excess* over the threshold u_0 is

$$E[X - u_0 | X > u_0] = \frac{\beta(u_0)}{1 - \xi}$$

For any $u > u_0$, define the *mean excess function* $e(u)$ as

$$e(u) = E[X - u | X > u] = \frac{\beta(u_0) + \xi(u - u_0)}{1 - \xi}$$

Alternatively, for any $y > 0$

$$e(u_0 + y) = E[X - (u_0 + y) | X > u_0 + y] = \frac{\beta(u_0) + \xi y}{1 - \xi}$$

- The mean excess function is a linear function of $y = u - u_0$.

5.3.1 Graphical Diagnostic for determining u_0

Define the *empirical mean excess function*

$$e_n(u) = \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u)$$

where $x_{(i)}$ $(i = 1, \ldots, n_u)$ are the values of x_i such that $x_i > u$.

- The *mean excess plot* is a plot of $e_n(u)$ against u and should be linear in u for $u > u_0$

5.4 Maximum Likelihood Estimation

Let $x_1, \ldots, x_n$ be iid sample of losses with unknown CDF F.

- For a given u, extreme values are those x_i values for which $x_i - u > 0$. Denote these values $x_{(1)}, \ldots, x_{(k)}$

- Define the threshold excesses as $y_i = x_{(i)} - u$ for $i = 1, \ldots, k$.

If u is large enough, then $\{y_1, \ldots, y_k\}$ may be thought of as a random sample from a GPD with unknown parameters ξ and $\beta(u)$. For $\xi \neq 0$, the log-likelihood function based on the GPD is

$$l(\xi, \beta(u)) = -k \ln(\beta(u)) - (1 + 1/\xi) \sum_{i=1}^{k} \ln(1 + \xi y_i/\beta(u))$$

5.5 Estimating the Tails of the Loss Distribution

For a sufficiently high threshold u, $F_u(y) \approx G_{\xi,\beta(u)}(y)$. Seting $x = u + y$, an approximation to the tails of $F(x)$ for $x > u$ is given by

$$F(x) = (1 - F(u))G_{\xi,\beta(u)}(y) + F(u)$$

- Estimate $F(u)$ using the empirical CDF

$$\begin{aligned} \hat{F}(u) &= \frac{(n-k)}{n} \\ k &= \text{number of exceedences over } u \end{aligned}$$

- Combine $\hat{F}(u)$ with $G_{\hat{\xi},\hat{\beta}(u)}(y)$ to give

$$\hat{F}(x) = 1 - \frac{k}{n}\left(1 + \hat{\xi} \cdot \frac{x-u}{\hat{\beta}(u)}\right)$$

- This approximation is used for VaR computations based on the fitted GPD

5.6 Risk Measures Again

- VaR for GPD. Compute $\hat{F}^{-1}(q)$ using

$$\widehat{VaR}_q = u + \frac{\hat{\beta}(u)}{\hat{\xi}} \left(\left(\frac{n}{k}(1-q) \right)^{-\hat{\xi}} - 1 \right)$$

- ES for GPD

$$\widehat{ES}_q = \frac{\widehat{VaR}_q}{1-\hat{\xi}} + \frac{\hat{\beta}(u) - \hat{\xi}u}{1-\hat{\xi}}$$

5.7 Non-Parametric Tail Estimation

- The shape parameter ξ, or equivalently, the tail index $a = 1/\xi$, of the GEV and GPD distributions may be estimated non-parametrically in a number of ways.

- A popular method due to Hill (1975) applies to the case where $\xi > 0$ ($\alpha > 0$) so that the data is generated by some fat-tailed distribution in the domain of attraction of a Fréchet type GEV.

- Consider a sample of losses $X_1, \ldots, X_T$ and define the order statistics as

$$X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(T)}$$

For a positive integer k, the *Hill estimator* of ξ is defined as

$$\hat{\xi}^{Hill}(k) = \frac{1}{k} \sum_{j=1}^{k} \left(\log X_{(j)} - \log X_{(k)} \right)$$

and the Hill estimator of α is

$$\hat{\alpha}^{Hill}(k) = 1/\hat{\xi}^{Hill}(k)$$

- The Hill estimators of ξ and α depend on the integer k, which plays the same role as k in the analysis of the GPD.

- It can be shown that if F is in the domain of attraction of a GEV distribution, then $\hat{\xi}^{Hill}(k)$ converges in probability to ξ as $k \longrightarrow \infty$ and $\frac{k}{n} \longrightarrow 0$, and that $\hat{\xi}^{Hill}(k)$ is asymptotically normally distributed with asymptotic variance

$$\text{avar}(\hat{\xi}^{Hill}(k)) = \frac{\xi^2}{k}$$

By the delta method, $\hat{\alpha}^{Hill}(k)$ is asymptotically normally distributed with asymptotic variance

$$\text{avar}(\hat{\alpha}^{Hill}(k)) = \frac{\alpha^2}{k}$$

- In practice, the Hill estimators $\hat{\xi}^{Hill}(k)$ or $\hat{\alpha}^{Hill}(k)$ are often plotted against k to find the value of k such that the estimator appears stable.

5.7.1 Hill Tail and Quantile Estimation

Suppose that the loss distribution F is such that $1 - F(x) = x^{-\alpha}L(x)$ with $\alpha = 1/\xi > 0$, where $L(x)$ is a slowly varying function. Let $x > X_{(k+1)}$ where $X_{(k+1)}$ is a high order statistic. The Hill estimator of $F(x)$ is given by

$$\hat{F}^{Hill}(x) = 1 - \frac{k}{n}\left(\frac{x}{X_{(k+1)}}\right)^{-\hat{\alpha}^{Hill}(k)}, \; x > X_{(k+1)}$$

Inverting the Hill tail estimator gives the Hill quantile estimator

$$\hat{x}_{q,k}^{Hill} = X_{(k+1)} - X_{(k+1)}\left(\left(\frac{n}{k}(1-q)\right)^{-\hat{\xi}^{Hill}(k)} - 1\right)$$

where $q > 1 - k/n$.