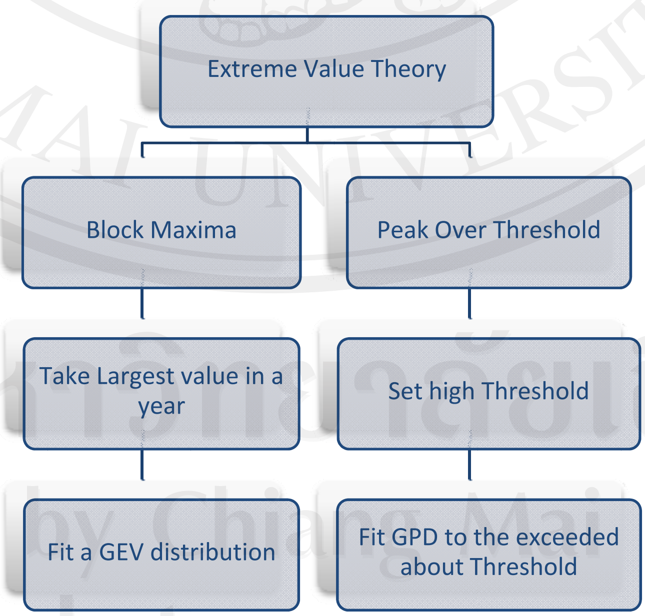
Chapter 3

Methodology

3.1 Research Design

This research investigates the evaluation and analysis of Value at Risk in daily gold price return by Extreme Value Theory Approach. Extreme Value Theory Approach modeled two methods for evaluation: 1) Block Maxima method (BM) modeled by the Generalized Extreme Value (GEV) distribution; 2) Peak Over Threshold (POT) models for large values over some high threshold, which can be simulated by the Generalized Pareto Distribution (GPD).

Figure 3.1 shows that how to calculate Value at risk of gold price return through Extreme Value Theory. First, Use of block maxima method is gathered the negative largest value of every year and then fit the Generalized Extreme Value (GEV). Second, Use of Peak Over Threshold is set a high threshold and then fit the Generalized Pareto Distribution to the exceedence about threshold by using R project econometrics package.



Source: Created by author

Figure 3.1 Concept Framework of Value at Risk Calculation

3.2 Data Sources

The data was gathered from secondary sources. This study uses daily gold prices in US dollar over the period spanning between January 1, 1985 and August 31, 2011. The source of gold prices is World Gold Council.

3.3 Research Methodology

3.3.1 Extreme Value Theory Approach

Extreme Value Theory models only the tail of the return distribution rather than the entire distribution with the extreme events. So, this approach can potentially perform better than other approaches in terms of forecasting unexpected extreme changes.

3.3.1.1 Block Maxima or Generalized Extreme Value Distribution (GEV)

This approach is the one the studies the limiting distributions of the sample extreme, which is presented under a single parameterization. In this case, extreme movements in the left tail of the distribution can be characterized by the negative numbers (Jiahn-Bang Jang, 2007)

Let X_i be the negative of the i th daily return of the gold prices between day i and day i-1. Define

$$X_i = -(\ln P_i - \ln P_{i-1})$$

where P_i and P_{i-1} are the gold prices of day i and day i-1 . Suppose that $X_1, X_2, \ldots, X_n$ be iid random variables with an unknown cumulative distribution function (CDF) F(x) = $\Pr(X_i \le x)$. Extreme values are defined as maxima (or minimum) of the n independently and identically distributed random variables $X_1, X_2, \ldots, X_n$.

Then, let X_n be the maximum negative side movements in the daily gold prices returns, that is

$$X_n = \max(X_1, X_2, \ldots, X_n)$$

Since the extreme movements are the focus of this study, the exact distribution of X_n can be written as

$$\Pr(X_n \le a) = \Pr(X_1 \le a, X_2 \le a, \ldots, X_n \le a)$$
$$= \prod_{i=1}^{n} F(a)$$
$$= F^n(a)$$

In practice the parent distribution F is usually unknown or not precisely known. The empirical estimation of the distribution F^n (a) is poor in this case. Fisher and Tippet (1928) derived the asymptotic distribution of F^n (a). Suppose μ_n and σ_n are sequences of real number location and scale measures of the maximum statistic X_n. Then the standardized maximum statistic

$$Z_n^* = \left(\frac{X_n - \mu_n}{\sigma_n}\right) \quad (1)$$

Converges to z = (x - μ) / σ which has one of three forms of non-degenerate distribution families such as

$$H(z) = \exp\{-\exp[-z]\}, -\infty < z < \infty$$
$$H(z) = \exp\{-Z^{-1/\xi}\}, z > 0$$
$$= 0, \text{else}$$
$$H(z) = \exp\{-[Z]^{-1/\xi}\}, z > 0$$
$$= 1, \text{else} \quad (2)$$

These forms go under the names of Gumbel, Frechet, and Weibull respectively. Here μ and σ are the mean return and volatility of the extreme values x and ξ is the shape parameter or called 1/ ξ the tail index of the extreme statistic distribution.

Embrechts et al. (1997) describe GEV distribution in detail, which fundamental types of extreme value distributions are defined by ξ:

1. If $\xi = 0$, the distribution is called the Gumbel distribution. In this case, the distribution spreads out along the entire real axis.

2. If $\xi > 0$, the distribution is called the Fréchet distribution. In this case, the distribution has a lower bound

3. If $\xi < 0$, the distribution is called the Weibull distribution. In this case, the distribution has an upper bound.

The Fisher and Tippett (1928) theorem suggests that the asymptotic distribution of the maxima belongs to one of the three distributions above, regardless of the original distribution of the observed data. Random variables fall into one of three tails shapes, fat, normal, and thin, depending on the various properties of the distribution. Thus, the tails of distributions are:

i. thin, that is, the tails are truncated,

ii. normal. In this case, the tails have an exponential shape,

iii. fat. The tails follow a power law.

Embrechts et al. (1997) proposed a generalized extreme value (GEV) distribution which included those three types and can be used for the case stationary GARCH processes. GEV distribution has the following form

$$H_{\xi}(X;\mu,\sigma) = \exp\{-\frac{\exp[-x-\mu)}{\sigma}]\}, -\infty \le x \le \infty;\ \xi = 0$$

$$= \exp\{-[1+\xi(x-\mu)/\sigma]^{-\frac{1}{\xi}}, 1+\frac{\xi(x-\mu)}{\sigma} > 0;\ \xi \neq 0 \quad (3)$$

Then, suppose that block maxima B_1, B_2,…., B_k are independent variables from a GEV distribution, the log-likelihood function for the GEV, under the case of $\xi \neq 0$, can be given as

$$\text{In L} = -\text{k In } \sigma - (1+\tfrac{1}{\xi})\sum_{i=1}^{k} \text{In}\{1+\ \xi\frac{(B_i-\mu)}{\sigma}\} - \sum_{i=1}^{k}\{1+\xi\frac{(B_i-\mu)}{\sigma}\}^{-1/\xi} \quad (4)$$

For the Gumbel type GEV form, the log-likelihood function can be written as

$$\text{In L} = -\text{k In } \sigma - \sum_{i=1}^{k}\frac{(B_i-\mu)}{\sigma} - \sum_{i=1}^{k}\exp\{\frac{(B_i-\mu)}{\sigma}\} \quad (5)$$

As Smith (1985) suggested that, for $\xi > 0.5$, the maximum likelihood estimators, for ξ, μ, and σ, satisfy the regular conditions and therefore having asymptotic and consistent properties. The number of blocks, k and the block size form a crucial tradeoff between variance and bias of parameters estimation.

3.3.1.2 Peak over threshold or Generalized Pareto Distribution (GPD)

Fitting models with more data is better than less, so Peaks over thresholds (POT) method utilizes data over a specified threshold. Then, Jiahn-Bang Jang (2007) defined the excess distribution as

$$F_h(x) = \Pr(X-h < x \mid X > h) = \frac{[F(x+h)-F(h)]}{1-F(h)} \quad (6)$$

where h is the threshold and F is an unknown distribution such that the CDF of the maxima will converge to a GEV type distribution. For large value of threshold h, there exists a function $\tau(h) > 0$ such that the excess distribution of equation (6) will approximated by the generalized Pareto distribution (GPD) with the following form

$$H_{\xi,\tau(h)}(x) = 1-\exp\left(-\frac{x}{\tau(h)}\right), \xi = 0$$

$$= 1 - (1 + \frac{\xi x}{\tau(h)})^{-1/\xi}, \xi \neq 0 \quad (7)$$

where x > 0 for the case of 0 , and $\xi \geq 0 \geq x$, and $0 \leq x \leq \tau(h) / \xi$ for the case of $\xi < 0$. Define $X_1, X_2,\ldots, X_k$ as the extreme values which are positive values after subtracting threshold h .

For large value of h, $X_1, X_2, \ldots, X_k$ is a random sample from a GPD, so the unknown parameters ξ and $\tau(h)$ can be estimated with maximum likelihood estimation on GPD log-likelihood function.

Based on equation (6) and GPD distribution, the unknown distribution F can be derived as

$$F(y) = \big(1 - F(h)\big)H_{\xi,\tau(h)}(x) + F(h) \quad (8)$$

where y = h + x . F(h) can be estimated with non-parametric empirical estimator

$$\hat{F}(h) = \frac{n-k}{n}$$

where k is the number of extreme values exceed the threshold h . Therefore the estimator of (8) is

$$\hat{F}(h) = \left(1 - \hat{F}(h)\right)\hat{H}\left(x; \hat{\xi}, \hat{\tau}(h)\right) + \hat{F}(h) \quad (9)$$

where $\hat{\xi}$ and $\hat{\tau}$(h) are mle of GPD log-likelihood. High quantile VaR and expected shortfall can be computed using (9). First, define $F(VaR_q) = q$ as the probability of distribution function up to q^{th} quantile VaR_q . Therefore

$$\widehat{VaR}_q = \hat{F}^{-1}(q) = h + \hat{\tau}(h)\left\{\left[\left(\frac{n}{k}\right)(1-q)\right]^{-\hat{\xi}} - 1\right\}/\hat{\xi} \quad (10)$$

Next, given that VaR_q is exceeded, define the expected loss size, expected shortfall (ES), as

$$ES_q = E\big(X \mid X > VaR_q\big) = VaR_q + E(X - VaR_q \mid X > VaR_q \quad (11)$$

From (10), $\widehat{ES}_q$ can be computed using $\widehat{VaR}_q$ and the estimated mean excess function of GPD distribution. Therefore,

$$\widehat{ES}_q = \frac{\widehat{VaR}_q}{1-\hat{\xi}} + (\hat{\tau}(h) - \hat{\xi}h)/(1-\hat{\xi})$$