CHAPTER 3

METHODOLOGY

3.1 Scope of the Study and Data Collection

This study uses daily MSCI Emerging Market Index in Asia; China, India, Indonesia, Malaysia, Philippines, South Korea, Singapore, Taiwan, and Thailand over the period of January 1, 2001 through December 31, 2011. The paper is considered to estimate Value at Risk (VaR) using an Extreme Value Theory applied in modeling extremes : Block maxima models modeled by the generalized extreme value (GEV) distribution; threshold models realized large values over some high threshold, which can be simulated by the generalized pareto distribution (GPD).

3.2 Research Methodology / Data Analyzing Method

This paper is used extreme value theory and statistical approaches. Extreme value theory relates to the asymptotic behavior of extreme observations of a random variable. It provides the fundamentals for the statistical modeling of rare events, and is used to compute tail risk measures. Researchers have contributed abundant theoretical discussion on EVT such as Embrechts et al. (1997), Reiss and Thomas (1997), and Coles (2001). Modeling of extreme value theory, there are two ways if identified extremes in data.

This paper is considered a random variable which may represent daily losses or returns. The first approach considers the maximum (or minimum) the variable takes in periods. The second approach focuses on the largest value variable over some high threshold.

3.2.1 Block Maxima or Generalized Extreme Value Distribution (GEV)

An approach is the one of studying 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 MSCI Emerging Market Index price in Asia between day i and day $i-1$. Define as $X_i = -(\ln P_i - \ln P_{i-1})$

where, P_i and P_{i-1} are the daily MSCI Emerging Market Index price in Asia of day i and day $i-1$. Suppose that $X_{1,} X_2, ..., X_n$ be the random variables with an unknown cumulative distribution function CDF ; $F(x) = \Pr(X_i \le x)$. Extreme values are defined as maxima of the n independently and identically distributed random variable $X_1, X_2, ..., X_n$.

Then, let X_n be the maximum negative side movements in the MSCI Emerging Market Index price in Asia returns, that is, $X_n = \max(X_{1,} X_2, ..., X_n)$. Since the extreme movements are the focus of this study, the exact distribution of X_n can be written as

$$
\begin{aligned}
\Pr(X_n \le \alpha) &= \Pr(X_1 \le \alpha, X_2 \le \alpha, ..., X_n \le \alpha) \\
&= \prod_{i=1}^{n} F(\alpha) \\
&= F^n(a)
\end{aligned}
$$

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

$$Z_n^* = \left(\frac{X_n - \mu_n}{\sigma_n} \right) \tag{3.1}$$

Converges to $z = \left(\frac{x - \mu}{\sigma} \right)$ which has one of three forms of non-degenerate distribution families such as

$$
\begin{aligned}
H(z) &= \exp\{-\exp[-z]\}, && -\infty < z < \infty \\
H(z) &= \exp\left\{-z^{-\frac{1}{\xi}}\right\}, && z > 0 \\
&= 0, && else \\
H(z) &= \exp\left\{-[z]^{-\frac{1}{\xi}}\right\}, && z > 0 \\
&= 1, && else
\end{aligned}
\tag{3.2}
$$

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 $1/\xi$ the tail index of the extreme statistic distribution. With $\xi = 0, \xi > 0, \xi < 0$ represent Gumbel, Frechet, and Weibull types of tail behavior respectively. In fact Gumbel, Frechet, and Weibull types can be fit for exponential, long, and short tails respectively.

According to Embrechts and Mikosch (1997) suggested 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

$$
\begin{aligned}
H_\xi(X;\mu,\sigma) &= \exp\left(-\left[\frac{\exp(-x-\mu)}{\sigma}\right]\right), -\infty \le x \le \infty; \xi = 0 \\
H_\xi(X;\mu,\sigma) &= \exp\left(-\left[1+\frac{\xi(-x-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}\right), 1+\frac{\xi(-x-\mu)}{\sigma} > 0; \xi \ne 0
\end{aligned}
\tag{3.3}
$$

Then, suppose that block maxima $B_{1,} B_2, \ldots, B_n$ are independent variables from a GEV distribution, the log-likelihood function for the GEV, under the case of $\xi \ne 0$, can be given as

$$
\ln L = -k \ln \sigma - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{k} \ln\left\{1+\frac{B_i-\mu}{\sigma}\right\} - \sum_{i=1}^{k}\left\{1+\frac{B_i-\mu}{\sigma}\right\}^{-\frac{1}{\xi}} \tag{3.4}
$$

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

$$
\ln L = -k \ln \sigma - \sum_{i=1}^{k} \ln \frac{B_i-\mu}{\sigma} - \sum_{i=1}^{k} \exp\left\{\frac{B_i-\mu}{\sigma}\right\} \tag{3.5}
$$

As Smith (1985) declared that, for $\xi > 0.5$, the maximum likelihood estimators, for $\xi, \mu,$ 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.

3.2.2 Peak over threshold or Generalized Pareto Distribution (GPD)

Jiahn-Bang Jang, 2007 stated that Peaks over Thresholds (POT) method utilizes data over a specified threshold. Define the excess distribution as

$$F_h(x) = \Pr\left(X - h < x \mid X > h\right) = \frac{\left[F(x+h) - F(h)\right]}{1 - F(h)} \tag{3.6}$$

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 (2.21) will approximate by the generalized Pareto distribution (GPD) with the following form

$$\begin{aligned} H_{\xi,\tau(h)}(x) &= 1 - \exp\left(-\frac{x}{\tau(h)}\right), \xi = 0 \\ H_{\xi,\tau(h)}(x) &= 1 - \left(1 + \frac{\xi x}{\tau(h)}\right)^{-1/\xi}, \xi \neq 0 \end{aligned} \tag{3.7}$$

where $x > 0$ for the case of 0 , and $\xi \geq 0 \geq$ x, and $\dfrac{0 \leq x \leq \tau\left(h\right)}{\xi}$ 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, therefore the unknown parameters ξ and $\tau\left(h\right)$ can be estimated with maximum likelihood estimation on GPD log-likelihood function.

Based on equation (3.6) and GPD distribution, the unknown distribution F can be derived as

$$F(y) = \left(1 - F(h) H_{\xi,\tau(h)}(x) + F(h)\right) \tag{3.8}$$

Therefore the estimator of (3.8) is $F(h) = \left(1 - F(h)\right) H\left(x; \xi, \hat{\tau}(h)\right) + F(h)$ (3.9) where ξ and $\hat{\tau}(h)$ are parameters of GPD log-likelihood. High quantile VaR and expected shortfall can be computed using (3.8). First, define $F(VaR_q) = q$ as the probability of distribution function up to q^{th} quantile VaR_q.

Therefore,
$$VaR_q = F^{-1}(q) = h + \hat{\tau}(h)\left\{\left[\frac{n}{k}(1-q)^{-\xi} - 1\right]\right\}\Big/\xi \tag{3.10}$$

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

$$ES_q = E\left(X \mid X > VaR_q\right) = VaR_q + E\left(X - VaR_q \mid X > VaR_q\right) \tag{3.11}$$

From (3.10), (Jiahn-Bang Jang, 2007) ES_q can be computed using VaR_q and the estimated mean excess function of GPD distribution.

Therefore,
$$ES_q = \frac{VaR_q}{1-\xi} + \frac{\left(\hat{\tau}(h) - \xi h\right)}{\left(1-\xi\right)}. \tag{3.12}$$