CHAPTER 2

LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1 Theoretical Background

2.1.1 Value at Risk

There are numerous characters of risk in financial markets. The three main categories of financial risk are credit risk, operational risk, and market risk. Value at Risk (VaR) is primarily concerned with market risk. The concept is also applicable to other forms of risk. Var is a single estimate of the amount by which an institution's perspective in a risk category could decline due to general market movement during a given holding period. Duffie and Pan (1997) and Jorion (1997) explained a general exposition of Var, the measure can be used by financial institutions to assess their risk or by regulatory committee to determine margin requirements. In either case, VaR is used to check the financial institution, which can still be business after catastrophic depression. From the outlook of a financial institution, VaR can be defined as the maximum loss of a financial view during the given time period for a given probability. In this outlook, one processes VaR as a measure of loss associated with a rare (or extraordinary) event under normal market conditions. Alternatively, from the perspective of a regulatory committee, VaR can be defined as minimum loss under extraordinary market events. The definition of both will result to same VaR measure, even though the concepts appear to be different.

In what follows, definition VaR under a probabilistic framework supposes that at the time index t. we are interested in the risk of financial position for the next ℓ periods. Let $\Delta V(\ell)$ be the change in value of the assets in the financial position for time t to $t+\ell$. The quantity is measured in dollars and is a random variable at the time index t. Denote the cumulative distribution function (CDF) of $\Delta V(\ell)$ by $F_\ell(x)$. We define the VaR of a long position over the time horizon ℓ with probability p as

$$p = \Pr\left[\Delta V(\ell) \le VaR\right] = F_\ell(VaR) \tag{2.1}$$

Since the holder of a long financial position sustain a loss when $\Delta V(\ell) < 0$. The VaR defined in (2.1) essentially assumes a negative value when p is small. The negative sign signifies a loss. From the definition, the probability that the holder would run into a loss greater than or equal to VaR over the time horizon ℓ is p. Alternatively, VaR can be explained as follows. With probability $(1-p)$, the potential loss run into by the holder position over the time horizon ℓ is less than or equal to VaR.

The holder of a short position sustains a loss when the value of the asset increases. The VaR is then defined as

$$p = \Pr\left[\Delta V(\ell) \ge VaR\right] = 1 - \Pr\left[\Delta V(\ell) \le VaR\right] = 1 - F_\ell(VaR) \tag{2.2}$$

for a small p, the VaR of a short position essentially assume a positive value. The positive sign signifies a loss.

The previous definition prove that VaR is concerned with tail behavior of the CDF $F_\ell(x)$. For the long position, the left tail of $F_\ell(x)$ is important. Thus far a short position focused on the right tail of $F_\ell(x)$. Notice that the definition of VaR in (2.1) proceeds to apply to a short position if one uses the distribution of $-\Delta V(\ell)$. Therefore, it suffices to discuss methods of VaR calculation using the long position.

For any univariate CDF $F_\ell(x)$ and probability p, such that $0 < p < 1$, the quantity

$$x_p = \inf\left\{x \mid F_\ell(x) \ge p\right\} \tag{2.3}$$

is called the p^{th} quantile of $F_\ell(x)$, where inf denotes the smallest real number satisfying $F_\ell(x) \ge p$. If the CDF $F_\ell(x)$ of (2.1) is known, the VaR is simply its p^{th} quantile (i.e., $VaR = x_p$). The CDF is unknown in practice, however. Learning of VaR are fundamentally concerned with estimation of the CDF and/or its quantile, particularly the tail behavior of the CDF.

In practical applications, calculation of VaR implies several factors:

1. The probability of interest (p), such as $p = 0.01$
2. The time horizon (ℓ), It might be set by a regulatory committee.
3. The frequency of the data, which might not be the same as the time horizon (ℓ). Daily observations are often used.
4. The CDF $F_\ell(x)$ or its quantile.
5. The amount of the financial position or the mark-to-market value of the portfolio.

2.1.2 Extreme Value

2.1.2.1 Classical Extreme Value Theory and Models

1. Asymptotic Model Formulation

In this section we want to study the distribution function of

$$M_n = \max\{X_1, ..., X_n\},$$

where $X_1, ..., X_n$, are independent random variables having a common distribution function *F*. So M_n represents the maximum of n observations.

The theoretical distribution of M_n can be calculated as follows:

$$\begin{aligned} \Pr\{M_n \le z\} &= \Pr\{X_1 \le z, ..., X_n \le z\} \\ &= \Pr\{X_1 \le z\} \times ... \times P\{X_n \le z\} \\ &= \{F(z)\}^n \end{aligned} \tag{2.4}$$

However this is not always very useful, as often the distribution function *F* is unknown. A possibility would be to estimate *F* with standard statistical techniques and to substitute into (2.4). But small discrepancies for *F* lead to bigger ones for F^n.

Therefore we use an approach to directly estimate F^n. We look at the behavior of F^n as $n \to \infty$. We remark that for any $z < z_+$, where z_+ is the smallest value of *z* such that $F(z) = 1$, $F^n(z) \to 0$ as $n \to \infty$, so that the distribution of M_n degenerates to a point mass on z_+. This can be avoided by making a linear renormalization of the variable M_n:

$$M_n^* = \frac{M_n - b_n}{a_n}$$

for constants $\{a_n > 0\}$ and $\{b_n\}$, which chosen appropriately stabilize the location and scale of M_n^* as n increases.

2. Extreme Types Theorem

If there is exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that[2]

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} \to G(z), \quad \text{as } n \to \infty$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$\text{I:} \qquad G(z) = G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty;$$

$$\text{II:} \qquad G(z) = \begin{cases} 0, & z \le b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b \end{cases}$$

$$\text{III:} \qquad G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)\right]^{\alpha}\right\}, & z < b \\ 1, & z \ge b \end{cases}$$

for parameters $a_n > 0$, b and, in the case of families II and III, $\alpha > 0$.

The distribution families in *Extreme Types Theorem* are called the extreme value distributions, with types I, II and III known as the Gumbel, Fréchet and Weibull families respectively.

This theorem implies that, if M_n can be stabilized, then the normalized variable M_n^* has a limiting distribution that is one of the three types of extreme value distribution.

[2] R. A. Fisher and L. H. C. Tippett, Limiting forms of the frequency distribution of the largest and smallest member of a sample, Proc. Cambridge Phil. Soc., 1928.

3. The Generalized Extreme Value Distribution

There is one problem with *Extreme Types Theorem*; the three types of limits have distinct forms of behavior. So there is a technique required to choose the most appropriate one of three families and subsequent inferences presume this choice to be correct.

The extreme types theorem can be reformulated, using only one family of distributions, which is called the generalized extreme value (GEV) family of distributions:

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} \to G(z), \quad \text{as } n \to \infty \tag{2.5}$$

for a non-degenerate distribution function G, then G is a member of the GEV family

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{2.6}$$

defined on $\{z : 1+\xi(z-\mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \xi < \infty$.

The apparent difficulty to know the normalizing constants can be resolved by assuming the hypothesis in (2.5),

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} \approx G(z)$$

for large enough *n*. Equivalently,

$$\Pr\{M_n \le z\} \approx \left\{\frac{M_n - b_n}{a_n}\right\}$$

$$= G^*(z)$$

where G^* is another member of the GEV family. So if we can approximate the distribution of M_n^* by a member of the GEV family for large n, the distribution of M_n can also be approximated by a different member of this family.

Quantiles of the GEV distribution can be obtained by inverting (2.6):

$$z_p = \begin{cases} \mu - \dfrac{\sigma}{\xi}\left[1-\{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\}, & \text{for } \xi = 0 \end{cases} \tag{2.7}$$

where $$G(z_p) = 1-p$$

The definition of the extreme quantile $z_p = G^{-1}(1-p)$ where G is the distribution function of M_n, is called the *return level* associated with the *return period*, $1/p$.

This terminology is due to the fact, that the level z_p is expected to be exceeded on average once every $1/p$ years, if our data consists of annual maxima.

4. Asymptotic Models for Minima

Sometimes we need models for extremely small instead of extremely large observations. Results for the GEV distribution for minima are analogues to those obtained above. So if $M_n = \min\{X_1,...,X_n\}$, we can reformulate (2.5) as follows:

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that[2]

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G(z), \quad \text{as } n \to \infty$$

for a non-degenerate distribution function G, then G is a member of the GEV family of distributions for minima:

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

defined on $\{z : 1+\xi(z-\mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \xi < \infty$.

Proof. Let $Y_i = -X_i$ for $i = 1,...,n,$ so if $M_n = \min\{X_1,...,X_n\}$ and then $M_n = -M_n$.

Hence, for large n we have:

$$\begin{aligned}
\Pr\{M_n \le z\} &= \Pr\{-M_n \le z\} \\
&= \Pr\{M_n \ge -z\} \\
&\approx 1 - \exp\left\{-\left[1+\xi\left(\frac{-z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\
&\approx 1 - \exp\left\{-\left[1+\xi\left(\frac{-z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}
\end{aligned}$$

on $\left\{z : 1+\xi\left(z-\mu\right)/\sigma > 0\right\}$, where $\mu = -\mu$

5. Maximum Likelihood Estimation

Under the assumption that $Z_1,...,Z_m$ are independent variables having the GEV distribution, the log-likelihood for the GEV parameters when $\xi \ne 0$ is

$$\ell(\mu,\sigma,\xi) = -m\log(\sigma) - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{-1/\xi}$$

provided that $$1+\xi\left(\frac{z_i-\mu}{\sigma}\right) > 0, \text{ for } i = 1,...,m \qquad (2.8)$$

If the inequality above is violated, it corresponds to a configuration where at least one of the observed data falls beyond an end-point of the distribution, then the likelihood is zero and the log-likelihood equals $-\infty$.

The case $\xi = 0$ has to be treated separately using the Gumbel limit of the GEV distribution.

When $\xi = 0$ the log-likelihood becomes

$$\ell(\mu,\sigma,\xi) = -m\log(\sigma) - \sum_{i=1}^{m}\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\exp\left\{-\left(\frac{z_i-\mu}{\sigma}\right)\right\}. \qquad (2.9)$$

Maximization of (2.8) and (2.9) leads to the maximum likelihood estimate with respect to the entire GEV family. There is no analytical solution, but for any given dataset maximization can be easily done using standard numerical optimization algorithms. However some care is needed to ensure that such algorithms do not move to parameter combinations violating (2.8) and that numerical difficulties would not arise from $\xi = 0$. The second problem is easily solved using (2.9) instead of (2.8) for values ξ falling within a small window around zero.

6. Profile Likelihood

Numerical evaluation of the profile likelihood for any of the individual parameters μ, σ or ξ is quite easy. For example, to obtain the profile likelihood for ξ, we fix ξ_0 and maximize the log-likelihood with respect to the others parameters, μ and σ. We repeat this operation for several ξ_0. The corresponding maximized values of the log-likelihood constitute the profile log-likelihood for ξ.

This can be applied for any return level z_p as well. It requires a reparameterization of the GEV model, so that z_p is one of the model parameters. Then like before we obtain the profile log-likelihood by maximization with respect to the remaining parameters. The reparameterization is straightforward:

$$\mu = z_p + \frac{\sigma}{\xi}\left[1 - \left\{-\log\left(1-p\right)\right\}^{-\xi}\right]$$

with this reparameterization the GEV model is now expressed in terms of the parameters $\left(z_p, \sigma, \xi\right)$.

2.1.2.2 Threshold Model

1 Generalized Pareto Distribution

The models seen in previous chapters have some weaknesses. Modelings only block maxima as in part 2.1.2 that is a wasteful approach if other data on extremes are available. The *r* largest order statistic model is better but sometimes it is useful to avoid the procedure of blocking.

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, having marginal distribution function *F*. It is natural to regard as extreme events those of the X_i that exceed some high threshold u. The main result is contained in the following theorem.

Given $X_1, X_2, \ldots$ be a sequence of independent random variables with common distribution F, and let[3]

$$M_n = \max\{X_1, \ldots, X_n\} \tag{2.10}$$

Denote an arbitrary term in the X_i sequence by *X*, and suppose that *F* satisfies, so that for large n, $\Pr\{M_n \le z\} \approx G(z)$ (2.11)

where $G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$

for some $\mu, \sigma > 0$ and ξ. Then, for large enough u, the distribution function of $(X-u)$, conditional on $X > u$, is approximately

$$H(y) = 1-\left(1+\frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{2.12}$$

defined on $\{y : y > 0 \text{ and } (1+\xi y/\tilde{\sigma}) > 0\}$, where

$$\tilde{\sigma} = \sigma + \xi(u-\mu).$$

The family of distributions defined by Equation (2.12) is called the generalized Pareto family. The theorem above says that if block maxima follow a distribution G, then threshold excesses have a corresponding approximate distribution within the Pareto family. Moreover the parameters of the generalized Pareto distribution (GPD) are uniquely determined by those of the associated GEV distribution. We see that changing the block size n would affect the values of the GEV parameters, but not those of the corresponding GPD.

The duality between the GEV and generalized Pareto families means that the shape parameter ξ is dominant in determining the qualitative behavior of the GPD, just as it is for the GEV distribution.

---

[3] A. C. Davison and R. L. Smith, Models for Exceedances over High Thresholds, Vol. 52, Journal of the Royal Statistical Society, Series B (Methodological), 1990.

If $\xi < 0$ the distribution of excesses has an upper bound of $u - \sigma/\xi$.

If $\xi > 0$ the distribution has no upper limit.

If $\xi = 0$ the distribution is also unbounded, which should again be interpreted by taking the limit $\xi \to 0$ in Equation (2.12), leading to

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), y > 0 \tag{2.13}$$

corresponding to an exponential distribution with parameter $1/\xi$.

2 Threshold Selection

The GPD approach contrasts with the block maxima approach tough the characterization of an observation as extreme if it exceeds a high threshold. The issue of threshold choice is analogous to the choice of block size in the bock maxima approach. The lower the threshold the easier the asymptotic model is violated, leading to bias. The higher the threshold is the fewer the excesses data available to estimate our model, leading to high variance. Usually we adopt the lower threshold as possible, subject to the limit model providing a reasonable approximation. Two methods are available to find the best threshold.

The first method is an exploratory technique carried out prior to model estimation, based on the mean of the GPD. When $\xi < 0$ the mean of the GPD is

$$E(Y) = \frac{\sigma}{1-\xi} \tag{2.14}$$

Then we find that starting from a threshold u_0 where the GPD is appropriate the mean is a linear function of u for every $u > u_0$. According to this, the estimates ξ and σ are expected to change linearly with *u*, at levels of *u* for which the GPD is appropriate. This leads to the following procedure. The locus of points

$$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}\left(x_{(i)} - u\right)\right) : u < x_{\max}\right\}$$

where $x_{(1)},\ldots,x_{(n_u)}$ consist of the n_u observations that exceed *u*, and $x_{\max}$ is the largest of the X_i that is the mean residual life plot. Above a threshold u_0 at which the generalized Pareto distribution provides a valid approximation to the excess distribution, the mean residual life plot should be approximately linear in *u*. Confidence intervals can be added to the plot based on the approximate normality of sample means.

The second procedure for threshold selection is to estimate the model at a range of thresholds. This method is described in detail in section "Threshold Choice Revisited".

3 Parameter Estimation

When the threshold is determined, the parameter of the GPD can be estimated by maximum likelihood. Suppose that the values $y_1,...y_k$ are the *k* excesses of a threshold *u*. For $\xi \neq 0$ the log-likelihood is derived from (2.12) as

$$\ell(\sigma,\xi) = -k\log\sigma - (1+1/\xi)\sum_{i=1}^{k}\log(1+\xi y_i/\sigma) \tag{2.15}$$

given $(1+\sigma^{-1}\xi y_i) > 0$ for $i=1,...,k$; otherwise, $\ell(\sigma,\xi) = -\infty$. If $\xi = 0$ the log-likelihood is obtained from (2.12) as

$$\ell(\sigma) = -k\log\sigma - \sigma^{-1}\sum_{i=1}^{k} y_i \tag{2.16}$$

To find these parameters numerical techniques are required as analytical maximization is not possible.

4 Return Levels

We have introduced the notion of return levels and return periods for the GEV distribution. The same can be done for the generalized Pareto distribution. Suppose that the generalized Pareto distribution with parameters σ and ξ is a suitable model for exceedances of a threshold, *u*, by a random variable *X*. That is, for *x* >*u* we have:

$$\Pr\{X > x | X > u\} = \left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}$$

Following that,

$$\Pr\{X > x\} = \varsigma_u\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi}$$

where $\varsigma_u = \Pr\{X > u\}$. So, the return level of x_m that is exceeded on average once every m observations is the solution of

$$\varsigma_u \left[1+\xi\left(\frac{x_m - u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m} \tag{2.17}$$

By rearranging of the terms we obtain:

$$x_m = u + \frac{\sigma}{\xi}\left[(m\varsigma_u)^{\xi} - 1\right] \tag{2.18}$$

if m is sufficiently large to be sure that $x_m > u$. For all the equations above we have assumed that $\xi \neq 0$. If $\xi = 0$ we obtain $x_m = u + \sigma \log(m\varsigma_u)$, sufficiently large.

To estimate return levels we substitute σ and ξ by their maximum likelihood estimates. For ς_u we remark that the number of exceedances of u follows the binomial Bin(n, ς_u) distribution, and therefore the maximum likelihood estimate of ς_u is $\hat{\varsigma}_u = \frac{k}{n}$ where k is the number of observations exceeding the threshold u.

From properties of the binomial distribution we deduce that $Var(\hat{\varsigma}_u) \approx \hat{\varsigma}_u(1-\hat{\varsigma}_u)/n$, so the variance-covariance matrix for $\hat{\varsigma}_u, \hat{\sigma}, \hat{\xi}$ is approximately

$$V = \begin{bmatrix} \hat{\varsigma}_u(1-\hat{\varsigma}_u)/n & 0 & 0 \\ 0 & v_{1,1} & v_{1,2} \\ 0 & v_{2,1} & v_{2,2} \end{bmatrix}$$

where $v_{i,j}$ denotes the (i, j) term of the variance-covariance matrix of $(\hat{\sigma}, \hat{\xi})$. Hence, using the delta method we obtain

$$Var(\hat{x}_m) \approx \nabla x_m^T V \, \nabla x_m \tag{2.19}$$

where

$$\begin{aligned} \nabla x_m^T &= \left[\frac{\partial x_m}{\partial \varsigma_u}, \frac{\partial x_m}{\partial \sigma} \frac{\partial x_m}{\partial \xi}\right] \\ &= [\sigma m^{\xi} \varsigma_u^{\xi-1}, \xi^{-1}\{(m\varsigma_u)^{\xi} - 1\}, \\ &\quad -\sigma\xi^{-2}\{(m\varsigma_u)^{\xi} - 1\} + \sigma\xi^{-1}(m\varsigma_u)^{\xi}\log(m\varsigma_u)] \end{aligned}$$

5 Threshold Choice Revisited

Sometimes the mean residual life plot can be difficult to interpret as a method of threshold selection. Therefore we introduce a complementary technique.

By (2.10), if a generalized Pareto distribution is a reasonable model for excesses of a threshold u_0, then excesses of a higher threshold u also follow a generalized Pareto distribution. The parameters $\mu, \xi,$ and σ coming from the GEV distribution G will not change. The only value that will change is the generalized Pareto scale parameter σ, which we denote by σ_u for a threshold of $u > u_0$. It follows that

$$\sigma_u = \sigma_{u0} + \xi(u - u_0) \tag{2.20}$$

By reparametrization of the scale parameter as

$$\sigma^* = \sigma_u - \xi u$$

we obtain that σ^* is constant with respect to u. So if u_0 is a valid threshold estimates of both σ^* and ξ should be constant above u_0.

This argument suggests plotting $\hat{\sigma}^*$ and $\hat{\xi}$ plotting against *u*, and selecting u_0 as the lowest value of u for which the estimates remain near-constant.

6 Model Checking

Probability plots, quantile plots, return level plots and density plots are all useful for assessing the quality of a fitted generalized Pareto model. Assuming a threshold *u*, threshold excesses $y_{(1)} \le \cdots \le y_{(k)}$ and an estimated model $\hat{H}$, the probability plot consists of the pairs

$$\left\{\left(i/(k+1), \hat{H}\left(y_{(i)}\right)\right); i = 1, \ldots, k\right\} \tag{2.21}$$

where

$$\hat{H}(y) = 1 - \left(1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right)^{-1/\xi}$$

by $\hat{\xi} \ne 0$. If $\hat{\xi} = 0$, using (2.13) constructs the plot in place of (2.12). Then assumes $\hat{\xi} \ne 0$ again, the quantile plot consists of the pairs

$$\{(\hat{H}^{-1}(i/(k+1)), y_{(i)}), i = 1, \ldots, k\} \tag{2.22}$$

where

$$\hat{H}^{-1}(y) = u + \frac{\hat{\sigma}}{\hat{\xi}}[y^{-\hat{\xi}} - 1]$$

If the generalized Pareto model is reasonable for modeling excesses of *u*, then both the probability and quantile plots should consist of points that are approximately linear.

A return level plot consists of the locus of points $\{(m, x_m)\}$for large values of *m*, where $\hat{x}_m$ is the estimated *m*-observation return level:

$$\hat{x}_m = u\frac{\hat{\sigma}}{\hat{\xi}}\left[(m\hat{\varsigma}_u)^{\hat{\xi}} - 1\right] \tag{2.23}$$

if $\hat{\xi} = 0$, as with the GEV return level plot, it is usual to plot the return level curve on a logarithmic scale to emphasize the effect of extrapolation, and also to add confidence bounds and empirical estimates of the return levels.

Finally, the density function of the fitted generalized Pareto model can be compared to a histogram of the threshold excesses (Stuart Coles 2001, p.84-86).

2.2 Literature Review

2.2.1 Value at Risk

In term of evaluation in Value at Risk, Jaroslav Baran and Jiří Witzany (2010) applied extreme value theory in estimating low quantiles of profit and losses distribution and the results are compared to common VaR methodologies. The result confirms that EVTGARCH is superior to other methods. Gençay and Selçuk (2004), they investigate the extreme value theory to generate Value at Risk to estimate and study the tail forecasts of daily returns for stress testing. Then, Bali (2003) studies how to estimate volatility and Value at Risk by an Extreme Value Approach and determines the type of asymptotic distribution for the extreme changes in U.S. Treasury yields. In this paper, the thin-tailed Gumbel and exponential distribution are worse than the fat-tailed Fréchet and Pareto distributions.

In the analysis of Stelios Bekiros and Dimitris Georgoutsos (2003) conduct a comparative evaluation of the predictive performance of various Value-at-Risk (VaR) models. Both estimation techniques are based on limit results for the excess distribution over high thresholds and block maxima respectively. The results we report reinforce previous ones according to which some “traditional” methods might yield similar results at conventional confidence levels but at very high ones the extreme value theory methodology produces the most accurate forecasts of extreme losses. Moreover, Yasuhiro Yamai and Toshinao Yoshiba (2002) investigate the comparison of value-at-risk (VaR) and expected shortfall under market stress. The paper found that First, VaR and expected shortfall may underestimate the risk of securities with fat-tailed properties and a high potential for large losses. Second, VaR and expected shortfall may both disregard the tail dependence of asset returns. Third, expected shortfall has less of a problem in disregarding the fat tails and the tail dependence than VaR does.

2.2.2 Extreme Value Theory Approach

Extreme value theory is most use in evaluation of Value at Risk in Financial market. Martin Odening and Jan Hinrichs[4] (2010), who investigate on Using Extreme Value Theory to Estimate Value-at-Risk, stated that this article examines problems that may occur when conventional Value-at-Risk (VaR) estimators are used to quantify market risks in an agricultural context. For example, standard Value at risk methods, such as variance-covariance method or historical simulation, can fail when the return distribution is fat tailed. This problem is aggravated when long-term Value at risk forecasts is desired. Extreme Value Theory is proposed to overcome these problems. For a stock market study, Vladimir Djakovic, Goran Andjelic, and Jelena Borocki (2010**)** investigate the performance of extreme value theory (EVT) with the daily stock index returns of four different emerging markets. Research results according to estimated Generalized Pareto Distribution (GPD) parameters indicate the necessity of applying market risk estimation methods and it is clear that emerging markets such as those of selected emerging markets have unique characteristics.

---

[4] Professor of Farm Management and PhD candidate, respectively. Department of Agricultural Economics, Humboldt University Berlin, Germany.

In the analysis of gold price return, the study of Jiahn-Bang Jang (2007) has been examined to illustrate the main idea of extreme value theory and discuss the tail behavior. The results show that GPD model with threshold is a better choice. Also, Blake LeBaron and Ritirupa Samanta (2004) investigate that how to apply Extreme Value Theory (EVT) to construct statistical tests. The result shows that EVT elegantly frames the problem of extreme events in the context of the limiting distributions of sample maxima and minima. In financial market study, Neftci (2000) found that the extreme distribution theory fit well for the extreme events in financial markets. Moreover, Alexander J. McNeil (1999) investigates about Extreme Value Theory for Risk Managers. In this paper, the tail of a loss distribution is of interest, whether for market, credit, operational or insurance risks, the POT method provides a simple tool for estimating measures of tail risk.