Predictive Risk Analysis of SSE Fund Index Based on POT Model

Haixia Lu

Nanjing Tech University Pujiang Institute, Nanjing, 211134, China

Keywords: Peaks over Threshold; Value－at－Risk; Expected Shortfall; Risk Analysis

Abstract: Accurate measurement of Value-at-Risk (VaR) and Expected Shortfall (ES) is a challenge for risk managers. Extreme value theory can accurately describe the quantile of the tail of the distribution. In this paper, the VaR and ES are calculated by Peaks over Threshold, and their error analysis is given. Then the Chinese SSE fund index is taken as an example to analyze and test the data, and the VaR and ES values and confidence intervals of the Shanghai Stock Index are given. Practice has proved that the extreme value method based on POT model can predict the risk of SSE fund index very well.

1. Introduction

Since the financial asset yield series has unique statistical characteristics such as peak, thick tail, skewers, agglomeration and leverage effect, it is a challenging task to find a reliable model to evaluate its risk value [1]. In financial research, risk assessment is a measure of uncertainty that comes from future price changes in assets [2]. While investors also want to get more than expected returns, they are more concerned about the biggest losses that can be suffered, and therefore require a risk assessment of the investment [3]. In the risk assessment process, the risk value model is an important technique for quantifying specific investment risks. VaR is the maximum expected loss due to changes in market factors over a certain period of time and at a certain level of significance [4]. Accurately estimating VaR is critical for investors and financial institutions in the risk assessment process. However, reliable risk value modeling and forecasting are based on certain assumptions [5]. For example, the basic parameter VaR model assumes that financial asset returns follow a normal distribution and the volatility is constant [6]. However, previous studies have shown that the distribution of financial assets returns is different from the normal distribution, with the characteristics of thick tail distribution and peak state distribution. The probability of extreme results is greater than that under normal distribution [7]. In addition, the fluctuations in financial asset returns are time-varying and are characterized by volatility agglomeration and leverage [8]. Academia and practice use a variety of methods to predict VaR [9]. However, since the distribution of a typical portfolio will change over time, no single method can give a prediction that is satisfactory to everyone [10].

2. The concept of VaR and ES:

VaR (Value-at-Risk) is a widely accepted risk measurement tool. In 2001, the Basel Committee designated the VaR model as a risk measurement tool for banks. It can be defined as the maximum loss of an asset or portfolio over a specified period of time at a certain confidence level p, or the quintile of the distribution function of the portfolio return loss. Assuming that X represents the return of a financial asset whose density function is f(x), then VaR can be expressed as:

$$VaR_p = -\inf\{x \mid f(X \le x) > (1-p)\} \tag{1}$$

When the density function f(x) is a continuous function it can also be written: $VaR_p = -F^{-1}(p)$,

Where F^{-1} is called a fractional function, which is defined as the inverse of the loss distribution f(x). The model is simple to calculate. When the portfolio loss X meets the normal distribution and the number of securities in the portfolio does not change, the risk of controlling the combination can

  DOI: 10.25236/ermbfe.2019.080

be relatively effective. However, the VaR model only cares about the frequency exceeding the VaR value, and does not care about the loss distribution exceeding the VaR value, and the performance is unstable when the processing loss conforms to the non-normal distribution (such as the tail phenomenon) and the portfolio changes. Appears as follows:

$$VaR_p(X+Y) \geq VaR_p(X) + VaR_p(Y) \tag{2}$$

This does not satisfy the sub-additive of the consistency risk metric model proposed by Artzner.

$ES_{(p)}$ (Expected shortfall)satisfies Artzner's sub-additive, homogeneity, monotonicity, and translation invariance conditions, and is a consistency risk measurement model. It is defined as follows: At a given confidence level p, let X be a random variable describing the loss of the portfolio, $F(x) = P[X \leq x]$ is its probability distribution function, make $F^{-1}(\alpha) = \inf\{x \mid F(x) \geq \alpha\}$, Then $ES_{(\alpha)}(X)$ can be expressed as:

$$ES_{(p)}(X) = -\frac{1}{p}\int_0^{1-p} F^{-1}(\alpha)d\alpha \tag{3}$$

When the density function of the loss X is continuous, $ES_{(p)}$ can be simply expressed as: $ES_p = -E\{x \mid F(x) \leq (1-p)\}$.

3. Extreme Value Theory

3.1. Peaks over threshold

Extreme value theory is a method of measuring risk loss under extreme market conditions. It has the ability to estimate beyond the sample data and can accurately describe the quantile of the tail of the distribution. It mainly includes two types of models: Block Maxima Method and Peaks over Threshold.

Assuming that the distribution function of the sequence $\{z_t\}$ is F(x), and $F_u(y)$ is defined as a conditional distribution function in which the random variable Z exceeds the threshold value u, it can be expressed as:

$$F_u(y) = P(Z-u \leq y \mid Z > u) \qquad y \geq 0 \tag{4}$$

According to the conditional probability formula we can get:

$$\begin{aligned} F_u(y) &= \frac{F(u+y)-F(u)}{1-F(u)} = \frac{F(z)-F(u)}{1-F(u)} \qquad z \geq u \\ \Rightarrow F(z) &= F_u(y)(1-F(u)) + F(u) \end{aligned} \tag{5}$$

Theorem 2: For a large class of distribution F (almost all common distributions) conditional excess distribution function $F_u(y)$, there is a $G'_{\xi,\sigma}(y)$, such that:

$$F_u(y) \approx G'_{\xi,\sigma}(y) = \begin{cases} 1-(1+\dfrac{\xi}{\sigma}y)^{-1/\xi} & \xi \neq 0 \\ 1-e^{-y/\sigma} & \xi = 0 \end{cases} \qquad u \to \infty \tag{6}$$

When $\xi \geq 0$, $y \in [0,\infty)$; When $\xi < 0$, $y \in [0, -\frac{\sigma}{\xi}]$. The distribution function $G'_{\xi,\sigma}(y)$ is called a generalized Pareto distribution.

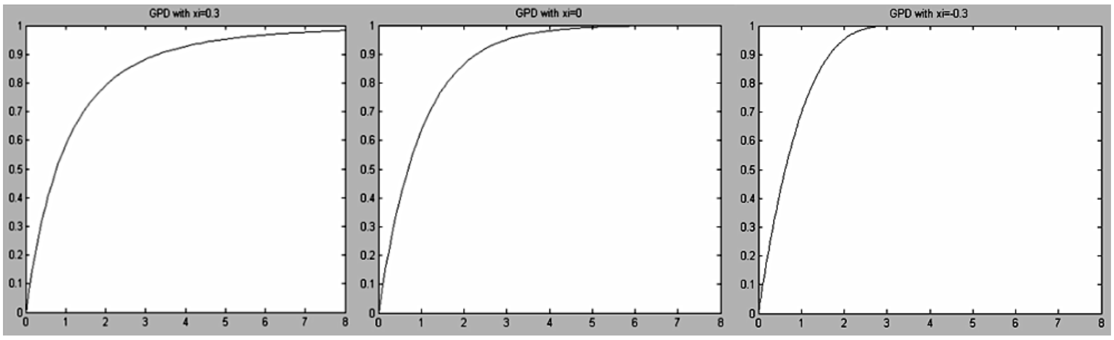

Fig.1. Generalized Pareto distribution when $\sigma = 1$, X takes a pattern of 0.3, 0, -0.3

From the graph, we can see that the different values of ξ determine the thickness of the tail. The larger the ξ is, the thicker the tail is. The smaller the ξ is, the thinner the tail is. From the $G'_{\xi,\sigma}(y)$ function, we can also get the maximum value of $\xi < 0$, The value is $-\frac{\sigma}{\xi}$ and has an upper bound.

When Lee and Saltoglu use EVT directly in the time series of financial asset returns, the determined ξ must be greater than zero due to the spikes in the sequence, but in our model, we perform extreme analysis on the residual sequence. So the ξ we get is not necessarily greater than zero.

We can get the probability density function $g'_{\xi,\sigma}(y)$ of the generalized Pareto distribution.

$$g'_{\xi,\sigma}(y) == \begin{cases} \frac{1}{\sigma}(1+\frac{\xi}{\sigma}y)^{-(1+\frac{1}{\xi})} & \xi \neq 0 \\ \frac{1}{\sigma}e^{-y/\sigma} & \xi = 0 \end{cases} \tag{7}$$

So for a given sample $\{z_1, \cdots, z_n\}$, Log likelihood function $L(\xi, \sigma \mid z)$ It can be expressed as:

$$L(\xi, \sigma \mid y) \begin{cases} -n\ln\sigma - (1+\frac{1}{\xi})\sum_{i=1}^{n}\ln(1+\frac{\xi}{\sigma}y_i) & \xi \neq 0 \\ -n\ln\sigma - \frac{1}{\sigma}\sum_{i=1}^{n}y_i & \xi = 0 \end{cases} \tag{8}$$

3.2. Estimation method of VaR and es confidence intervals of sequence z_t

In general, the estimation method of the parameter confidence interval can be obtained from the idea of the Likelihood Ratio Test in the case of a large sample. The likelihood ratio test is used to test the degree of fitting of two models of the same type. The likelihood ratios of two models of the same type are consistent with the χ^2 distribution, Its degree of freedom is equal to the number of newly added parameters in the complex model. Take the POT model as an example, To estimate the confidence intervals for the parameters ξ and σ at a given confidence level σ can be obtained by:

$$L(\xi, \sigma) > L(\hat{\xi}, \hat{\sigma}) - \frac{1}{2}\chi^2_{\alpha,2} \tag{9}$$

Among them, $\hat{\xi}$ and $\hat{\sigma}$ is the estimated optimal value, $L(x, y)$ r σ epresents the likelihood function. So we get the joint confidence interval between ξ and. If we want to get an estimate of VaR_p, Then $\bar{L}(VaR_p) =_{\xi}^{\max L(\xi, VaR_p)}$, The confidence interval of VaR_p can be obtained by:

$$\bar{L}(VaR_p) > L(\hat{\xi}, \hat{\sigma}) - \frac{1}{2}\chi^2_{\alpha,1} \tag{10}$$

However, the amount of extreme data that exceeds the threshold is not much, making the gradual effect of this estimate may be poor. To this end, we introduce the Bootstrap method to obtain an estimate of the confidence interval. Since the sequences $\{z_t\}$ we obtained are independently and identically distributed, we can extract N points from each time to form a new sequence. Using this sequence to estimate VaR_p and ES_p, repeat this operation, you can get a series of VaR_p and ES_p estimates, find the empirical distribution of VaR_p and ES_p, and finally get the confidence intervals of VaR_p and ES_p according to the empirical distribution, and put VaR_p and ES_p The expected value of B is taken as an estimate of VaR_p and ES_p. Here we only give the method of finding the VaR_p confidence interval in the POT model. The confidence intervals of other parameters can be similarly obtained. This method is also a method to test the stability of the model while determining the confidence interval.

4. Empirical Analysis of POT Model

We use the daily income composite index P published by the Shanghai Stock Exchange as the raw data. The sample space is from March 12, 2000 to November 20, 2014. The sample size is 3391 (using Eviews and Matlab software).

Based on the POT model in extreme value theory, we need to use a sufficiently large threshold u to fit the GPD of the overrun distribution. When we give a threshold of 0.8, 0.9, we use the maximum likelihood estimation to get the parameters: The value of $VaR_{0.01}$, $ES_{0.01}$ and the 95% confidence interval (see Table 1), and the Q-Q and distribution maps under these parameters (Figures 2 and 3). From the graph we can se $u = 0.8$ e th $R_t = \ln P_t - \ln P_{t-1}$ at the extreme value distribution effectively fits our sample distribution, and only a few exceptions occur. And there is no significant difference in the fitting effect between $u = 0.8$ and $u = 0.9$. For this reason, we only give the graph when $u = 0.9$.

Table 1 Maximum likelihood estimation of parameters and 95% confidence interval

	$N_u = 369$				$u = 0.9 \quad N_u = 287$			
	$\hat{\xi}$	$\hat{\sigma}$	$VaR_{0.01}$	$ES_{0.01}$	$\hat{\xi}$	$\hat{\sigma}$	$VaR_{0.01}$	$ES_{0.01}$
Lower	0.15	0.33	1.82	2.46	0.16	0.32	1.81	2.46
estimated	0.229	0.367	1.967	2.791	0.254	0.373	1.958	2.818
Upper	0.34	0.42	2.15	3.39	0.38	0.43	2.14	3.50
Interval	0.19	0.09	0.33	0.93	0.16	0.11	0.23	1.04

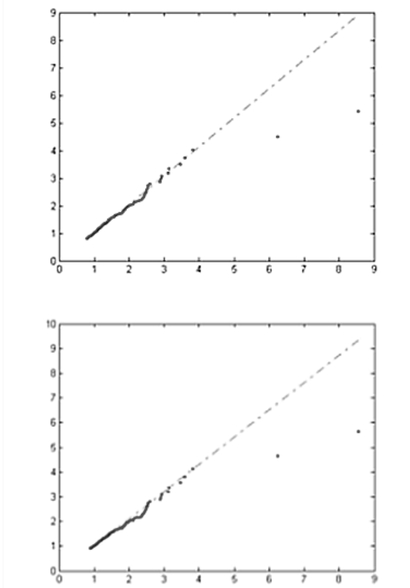

Fig.2. Q-Q diagram when $u = 0.8$ and $u = 0.9$

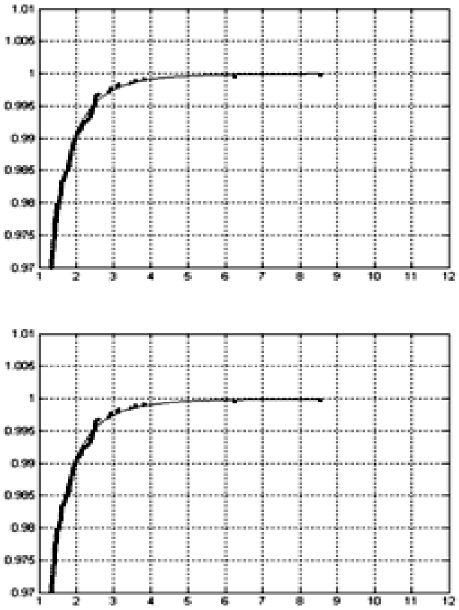

Fig.3. Comparison of extreme value distribution and empirical distribution when $u = 0.8$ and $u = 0.9$

For the estimation of $\hat{\xi}$, Embrechets believes that the financial sequence $\alpha = 1/\xi$ ranges from 3 to 4, and the $[3,4]$ we calculate here hardly falls within the area of $\hat{\xi}$. This is mainly because we have filtered the financial sequence with the ARMA-(Asymmetric) GARCH model, and the resulting sequence z_t has eliminated the spikes and tails to some extent, making the value estimated by $\hat{\xi}$ smaller, which is in line with the conclusion of Embrechets. Not contradictory.

In addition, in the Q-Q diagram, we can see that the fitting effect is very good before the fraction of 0.99, and there are individual outliers in the back, which will not affect our estimation of $VaR_{0.01}$. Because $VaR_{0.01}$ only cares about the distribution before the 0.99 is divided into numbers, and is not affected by the distribution after the 0.99 is divided into numbers. However, the estimation of $VaR_{0.01}$ is affected by the distribution after the 0.99 is divided into numbers, so this will cause a certain error in the estimation of $VaR_{0.01}$. This is one of the reasons why we see in Table 1 that the 95% estimation interval of $VaR_{0.01}$ is significantly wider than the 95% estimation interval of $VaR_{0.01}$.

We find that the estimated $VaR_{0.01}$ and $ES_{0.01}$ of the POT model are significantly smaller than the unadjusted estimates. This is because China's stock market was in its infancy between 2002 and 2004. The supervision was not strong enough, and the market volatility was relatively large. The unadjusted over-return income mainly occurred before 2005 and did not take into account in 2006. After the securities market is further standardized, the stock market over-return fluctuations are reduced, and $VaR_{0.01}$ and $ES_{0.01}$ should be adjusted accordingly. The data-adjusted extreme value forecast can effectively consider the impact of this factor, making the future It is estimated that more consideration is given to the current market risks, and it is more accurate to estimate the current risks of the market.

5. Conclusion

In risk management, a reasonable assumption of the distribution of returns is a prerequisite for the correct measurement of risk. However, the existing distribution, especially the widely used normal distribution, has a large gap with the actual financial income distribution. The POT model of extremum theory only considers the tail of the distribution, rather than modeling the whole distribution, which avoids the problem of distribution hypothesis; and the extreme value theory can accurately describe the quantile of the tail of the distribution, which is more helpful for processing. Our test results also show that the extreme value theory can accurately measure VaR and ES. The empirical test shows that the POT model can effectively improve the accuracy of extreme risk prediction, and the model is effective.

References

[1] Solari S, Egüen M, Polo M J. Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p - value. Water Resources Research, 2017, 53.

[2] Kang S, Song J. Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework ☆. Journal of the Korean Statistical Society, 2017, 46(4):S1226319217300200.

[3] Zhou X Y, Schmidt F, Toutlemonde F. A mixture peaks over threshold approach for predicting extreme bridge traffic load effects. Probabilistic Engineering Mechanics, 2016, 43:121-131.

[4] Silva A T, Portela M M, Naghettini M. A Bayesian peaks-over-threshold analysis of floods in the Itajaí-açu River under stationarity and nonstationarity. Stochastic Environmental Research and Risk Assessment, 2017, 31(1):185-204.

[5] Kumar N, Kalyani S. Modeling the Behavior of Peaks of OFDM Signal Using ‘Peaks Over Threshold’ Approach. IEEE Transactions on Wireless Communications, 2017, 16(6):3590-3600.

[6] Tong S T W, Li W K. Extreme values identification in regression using a peaks-over-threshold approach. Journal of Applied Statistics, 2015, 42(3):566-576.

[7] Villarini G, Smith J A, Vecchi G A. Changing Frequency of Heavy Rainfall over the Central United States. Journal of Climate, 2013, 26(1):351-357.

[8] Ferreira A, Haan L D. The generalized Pareto process; with a view towards application and simulation. Bernoulli, 2014, 20(4):1717-1737.

[9] Beirlant J, Alves I F, Reynkens T. Fitting tails affected by truncation. Electronic Journal of Statistics, 2017, 11(1):2026-2065.

[10] Baker V R. Stream-channel response to floods, with examples from central Texas. Geological Society of America Bulletin, 2015, 88(8):1057.