Portfolio Value-at-Risk Forecasting with GA-based Extreme Value Theory

P. C. Lin
Institute of Finance and Information[1]
P.C. Ko, P. S. Chiang
Department of Information Management[2]
[1,2]National Kaohsiung University of Applied Sciences

Abstract

Value-at-Risk (VaR) has become a popular risk measure since it was adopted by the Bank for International Settlements and US regulatory agencies in 1988. The VaR concept has also been further extended to the portfolio Value-at-Risk (PVaR) measure used for managing risks and returns under a multiple-asset portfolio. Precise prediction of PVaR provides better evaluation criteria in areas such as investment decision-making and risk management. The two issues concerned with portfolio risk are efficient set selection and volatility forecasting. Most of the statistical portfolio selection models are based on linear functions under specific assumptions. Due to the fat-tailed distribution in most real financial time-series data, extreme value theory (EVT) is powerful in determining the VaR of a portfolio by concentrating on estimating the shape of the fat-tailed probability distribution. However, using EVT to evaluate the portfolio's volatility is very difficult, because each individual within the portfolio has its own distinct peak threshold value. This study introduces an evolutionary portfolio volatility forecasting model to optimize portfolios under their maximum expected returns subject to a risk constraint. We use a genetic algorithm (GA) to extract the best portfolio set and most suitable peak threshold in order to estimate the portfolio's VaR by means of EVT.

Keywords: Portfolio Value-at-Risk (PVaR), Genetic Algorithms (GA), Extreme Value Theory (EVT), Peak Over Threshold (POT).

1. Introduction

Volatility forecasting has become an important input in the creation of many investments, options, and portfolios [10]. Portfolio managers and investors have certain risk threshold levels that they can bear. Portfolio volatility forecasting involves two key issues: efficient set selection and Value-at-Risk (VaR) measurement. Efficient set selection aims to allocate assets by maximizing the expected risk premium per unit of risk. The earlier portfolio selection methodologies were the mean-variance approach pioneered by Markowitz (1952) [1], quadratic programming [11], goal programming [6], and dynamic programming, [5] etc. However, most of them are based on linear functions under specific assumptions. Besides, VaR has become a popular risk measure since it was recommended and adopted by the Bank for International Settlements and US regulatory agencies in 1988. The VaR concept has been further extended to the portfolio Value-at-Risk (PVaR) measure used to evaluate the maximum potential loss of a portfolio with a given probability over a specified period [9]. For risk management, investment, and regulatory purposes, PVaR is also used to predict the probability of an extreme movement in the value of a portfolio [7]. The extreme movements are related to the tails of the distribution, which is especially important because of the fat-tailed characteristics in most financial time series data. Extreme value theory (EVT) focuses directly on the tails of distributions and therefore potentially provides better estimates of risk [2].

The genetic algorithm (GA), introduced by Holland in 1975, is a well-known efficient nonlinear search methodology in large spaces [3]. The efficient set selections within a portfolio have been efficiently solved by using GA [4][12][13]. These GA-based portfolio selection algorithms only focus on standard deviation as the appropriate measure for risk. This indicates that investors weigh the probabilities of negative returns equally against positive returns [7]. In this paper, risk evaluations concentrate on estimating the shapes of the tail distributions of financial return series.

To optimize the portfolios under their maximum expected returns subject to a risk constraint, we introduce a GA-based portfolio volatility forecasting model to extract the best portfolio set and dynamically estimate a suitable peak threshold for each asset in the portfolio, simultaneously. These peak thresholds are used to estimate the portfolio's VaR by using EVT. In addition, the Peak-Over-Threshold (POT) method is used to collect returns in the series that exceed a certain high threshold.

2. PVaR using the EVT model under the GA mechanism

The proposed GA-based PVaR forecasting approach that uses the EVT model depicted in Figure 1 applies the GA mechanism to learn and obtain the best population. In the GA mechanism, the target assets are selected from each of the chromosomes (individuals) to be the candidates to forecast the efficient portfolio set in the current population, whose encoding method is shown in 2.1. Then, this mechanism applies GA operations (e.g., selection, crossover, and mutation) to create the next-generation population according to the fitness of the chromosomes. The fitness function concentrates on the PVaR evaluation. Each individual will be evaluated by the $PVaR_{EVT}$ model illustrated in 2.2. In the $PVaR_{EVT}$ model, the POT method is used to calculate VaR which built the history returns in series that exceed a certain high threshold value and to model these returns separately from the rest of the distribution. Finally, "Backtesting" is used to estimate the success rate of each individual. In addition, the number of portfolio sets is also considered in order to arrive at a reasonable portfolio if the fitness function is being designed.

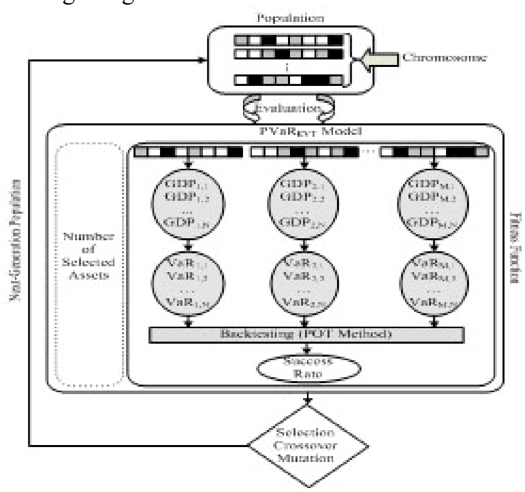

Figure 1. PVaR using the EVT model under the GA Mechanism

2.1 Encoding Method and Fitness Function Design

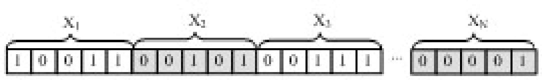

Figure 2. Encoding a portfolio and threshold into chromosomes

To represent a set of selected portfolios, we use a $5X_n$ bit string as a chromosome so that each genotype (X_i) in the chromosome is encoded into 5 bits as shown in Figure 2. The first bit represents the presence (1) or absence (0) of the corresponding assest in each genotype. The other 4 bits (2^4) represent the threshold value of each asset. The search space in the portfolio volatility forecasting for genetic algorithms is thus 2^{5n}.

In this paper, the higher success rate when less assets are considered in the portfolio is preferred. Thus, both the success rate criterion and the number of assets criterion are considered simultaneously. Let $F(V)$ denote the fitness function which is defined as follows, where SR_{VaR} denotes the success rate calculated by the $PVaR_{EVT}$ model, and S represents the number of selected assets:

$$F(V) = \frac{SR_{VaR}}{(|N| - |S|)/|N|} \tag{1}$$

2.2 The $PVaR_{EVT}$ model

The $PVaR_{EVT}$ model estimates the parameters (β_i, ξ_i) of the asymptotic distribution of selected minimal returns from the Generalized Pareto Distribution (GPD), and the probability (p_i) of a minimal return not exceeding a certain high threshold, u_i, in asset X_i. Assume that the collected returns in the series data $\{ r^i_{1,u_i}, r^i_{2,u_i}, \ldots, r^i_{Nu_i,u_i} \}$ are above the threshold u_i which follows the excess distrubution $F_{u_i}(x)$ shown in Equation (2).

$$\begin{aligned} F_{u_i}(x) &= P(r^i - u_i \le x \mid r^i > u_i) \\ &= \frac{F(u_i + x) - F(u_i)}{1 - F(u_i)} \\ &= \frac{F(r^i) - F(u_i)}{1 - F(u_i)} \qquad 0 \le x \le \infty, x = r^i - u_i \end{aligned} \tag{2}$$

$F_{u_i}(x)$ is approximated to GPD as shown in Equation (3), if u_i is high enough for a large class of distributions.

$$\lim_{u\uparrow} \sup_{0 \le x \le r^i - u_i} \left| F_{u_i}(x) - G_{\beta_i,\xi_i}(x) \right| = 0$$

$$G_{\beta_i,\xi_i}(x) = \begin{cases} 1 - \left(1 + \xi_i \dfrac{x}{\beta_i}\right)^{-\frac{1}{\xi_i}} & , \quad \xi_i \ne 0 \\ 1 - e^{-\frac{x}{\beta_i}} & , \quad \xi_i = 0 \end{cases} \tag{3}$$

β_i is a positive scaling parameter, and ξ_i is the tail index for the fat-tailed distributions found in finance. We estimate them using Maximum Likelihood Estimation (MLE) as shown below:

$$g_{\beta_i,\xi_i}(x_j) = \frac{1}{\beta_i}\left(1 + \xi_i \frac{x_j}{\beta_i}\right)^{-\frac{1}{\xi_i}-1} \tag{4}$$

After estimating the GPD distribution and its parameters, the VaR_i of the underlying return distribution $F_{u_i}(x)$ is calculated as in Equation (4) for asset X_i.

$$VaR_i = u_i + \frac{\beta_i}{\xi_i}\left\{ \left[\frac{T}{N_{u_i}}(1 - p_i) \right]^{-\xi_i} - 1 \right\} \tag{5}$$

3. Experimental results

The weekly time series data on stock returns for 34 companies are arbitrarily selected to be our testing targets. The stocks of these companies have been listed and traded on the Taiwan Stock Exchange (TSE). Based on the recommendations of the Basle Capital Accord (BCA) in 1996, we use the "backtesting" method to evaluate the reliability of our model. Let T denote the observed period in

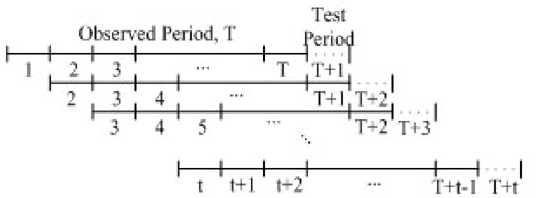

Figure 3. Each window consists of T observed values and one test value. The overall simulation process uses the sliding window methodology, which gradually shifts one time slot until the end of our experiment. By substituting these T observed values into Equations (3)-(5), the VaR_i is calculated. By comparing VaR_i with actual return data based on test values, success is achieved if VaR_i is larger, otherwise, failure results. After shifting the sliding window t times, the success rate is defined as follows:

$$Success\ Rate = \frac{Number\ of\ Success}{t} \tag{6}$$

In our experiment, T contains 764 observations in each window, the sliding window shifts 100 times (t=100), and the overall period of the experiment extends from 1987 to 2004.

The parameters used in GA are set as follows: the population size is 100, the chromosome length is 170 (5*34), the number of generations is 500, the crossover rate is 0.6 and the mutation rate is 0.001. Besides, the selection method consists of a roulette wheel and the crossover method is the one-point crossover.

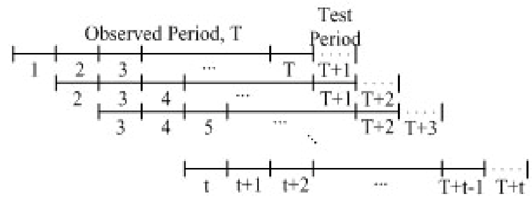

Figure 3. Sliding window simulation process

Three experiments, A, B, and C, with respective confidence levels of 90%, 95%, and 99% significance, are listed in Table 1, where higher confidence levels result in the selection of less assets and vice versa. For example: 10, 4 and 2 assets are selected if the 99%, 95% and 90% confidence levels are considered, respectively. The PVaR in Table 1 are 6.58, 7.09, and 10.46. This shows that higher confidence levels produce a larger PVaR. Furthermore, a higher confidence level avoids the situation where the risk for the portfolio as a whole is undervalued, such as in experiments B and C, where the success rates are higher than the confidence levels. Experiment B is listed in detail in Table 2. We find the GPD approxiamtion to be preferable, because it can deal with asymmetries in the fat-tailed distribution with $\xi>0$ and $\beta>0$.

Table 1. The optimal portfolio's VaR for empirical distribution

Experiment	Num. of stocks	Portfolio success rate	Confidence level (1-P)	Risk estimation	PVaR
A	10	85.23%	90%	Undervalued	6.58
B	4	98.75%	95%	Overvalued	7.09
C	2	100.00%	99%	Overvalued	10.46

Table 2. The detailed individual VaR in portfolios with a 95% confidence level

Selected Asset	Success rate	Threshold u_i	VaR_i	p_i	β_i	ξ_i
X_1	98%	-6.34	10.04	0.14	0.23	0.010
X_2	97%	-5.43	4.13	0.15	0.22	0.001
X_3	100%	-5.22	5.26	0.09	0.29	0.003
X_4	100%	-6.88	8.92	0.06	0.22	0.025
Average	98.75%	-5.97	7.09	0.12	0.24	0.010

4. Conclusion

Forecasting the volatility of a portfolio is important in investment decisions. Conventionally, forecasting PVaR and selecting an efficient set of portfolios should have serious limitations if statistical linear models are used based on specific assumptions. Traditional risk management models might fail to forecast VaR precisely because the tail distributions are estimated incorrectly. Most real financial time series data comprise fat-tailed distributions. The EVT model has been proved to be a powerful approach to risk forecasting in that it focuses directly on the distribution of extreme returns instead of the whole distribution of returns. However, forecasting the PVaR is very difficult because EVT provides less maturity when analyzing the multivariate model. In this article, a GA-based PVaR forecasting mechanism using the EVT model is introduced, where the GA mechanism extracts an efficient set of portfolios, and simultaneously estimates more suitable peak threshold values for all assets in order to forecast PVaR more precisely. Using 34 companies listed and traded on the TSE, we try to demonstrate the stability and robustness of our mechanism. The preliminary results show that the higher confidence level produces a larger PVaR. Furthermore, the results also show that a higher confidence level effectively results in the avoidance of risk undervaluation for the whole portfolio, because undervaluing risk will lead to possible financial distress or bankruptcy.

REFERENCES

[1] H. Markowitz, "Portfolio selection," Journal of Finance, Vol. 7, pp. 77-91, 1952.

[2] H. N. E. Bystrom, "Managing extreme risks in tranquil and volatile markets using conditional extreme value theory," International Review of Financial Analysis, Vol. 13, pp. 133-152, 2004.

[3] J. Holland: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. (University of Michigan Press, 1975).

[4] J. Shoaf and J. A. Foster, "The efficient set GA for stock portfolios," In Proceedings of the IEEE International Conference on Computational Intelligence, pp. 354 – 359, 1998.

[5] K. F. C. Yiu, "Optimal portfolios under a value-at-risk constraint," Journal of Economic Dynamics & Control, Vol. 28, pp. 1317-1334, 2004.

[6] P. M. Arenas, T. A. Bilbao, and U. M. V. Rodríguez, "A fuzzy goal programming approach to portfolio selection," European Journal of Operational Research, Vol. 133, pp. 287-297, 2001.

[7] R. Campbell, R. Huisman and K. Koedijk, "Optimal portfolio selection in a value at risk framework," Journal of Banking & Finance, Vol. 25, pp. 1789-1804, 2001.

[8] R. Gencay, F. Selcuk and A. Ulugulyagci, "High volatility, thick tails and extreme value theory in value-at-risk estimation," Insurance: Mathematics and Economics, Vol. 33, pp. 337-356, 2003.

[9] S. Ganganelli and R. F. Engle, "Value at risk models in finance," Working paper No. 75, European Central Bank, 2001.

[10] S. H. Poon and C. Granger, "Forecasting volatility in financial markets: a review," Vol. 41, pp. 478-539, 2003.

[11] Y. Crama and M. Schyns, "Simulated annealing for complex portfolio selection problems," European Journal of Operational Research, Vol. 150, pp. 546-571, 2003.

[12] Y. Xia, B. Liu, S. Wang, and K. K. Lai, "A model for portfolio selection with order of expected returns," Computers & Operations Research, 27, pp. 409-422, 2000.

[13] Y. Xia, S. Wang and X. Deng, "A compromise solution to mutual funds portfolio selection with transaction costs," European Journal of Operational Research, Vol. 34, pp. 564-581, 2001.