

Portfolio Model Based on Scenario Tree

Haiyan Xuan*, Cunliu Yao, Hongjian Li, Xiaoke Chang

School of Economics and Management, Lanzhou University of Technology, Lanzhou, China

Abstract—The uncertainty of return rate will affect the investment decision. In this paper, the ARMA-GARCH model is used to describe the data characteristics of stock returns, and the Monte Carlo method is used to construct a scenario tree containing the stock return rate and node probability. The decision rules are used to determine the nodes on the scene tree, and two mean-variance models are established based on the scene tree. Finally, four stock data are selected to optimize the portfolio of the constructed model, the results show that the scenario tree has good advantages in describing the uncertainty problem, and the constructed model is effective and feasible; the difference between the two models is analyzed and compared, which provides a reference for different investors.

1 Introduction

When to invest and how to make the return and risk reach a balance point satisfying investors are the main problems to be solved by portfolio. Markowitz[1] made a pioneering move, first using mean and variance to quantify the benefits and risks, and the proposed mean-variance model also laid a foundation for future research. Risk mainly comes from the uncertainty of yield, which gives decision-makers investment opportunities and makes decision-makers face gains or losses in investment. Based on the uncertainty of returns and risks, some scholars use fuzzy numbers to describe them, and give the corresponding portfolio model, which proves the practicability of fuzzy numbers in describing uncertainty[2]. However, fuzzy numbers are mainly based on past data, which can not provide a good guide for later investment decisions. In order to express the uncertainty better, Hoyland and Wallace[3] put forward the concept of scenario, which is the presentation of random variables at a certain moment. Yan et al.[4] used Copula method to give a scenario generation algorithm that can capture asymmetric tail dependence, and applied the generated scenario to multi-stage portfolio problems.

Through combing the above literature, it is found that few scholars have studied the portfolio directly from the perspective of scenario tree. In order to satisfy the portfolio with uncertain return rate and transform uncertainty into certainty, this paper combines scenario tree with the mean-variance model proposed by Markowitz[1], that is, studies the portfolio model based on scenario tree. The ARMA-GARCH model is used to predict the rate of return, Monte Carlo simulation is used to generate new scenarios, and the scenario tree is constructed by combining the sampling method. Finally, based on the scenario tree and combining the decision

rules, two different portfolio models are established to study the feasibility and difference of the two models.

2 Theory and Model

2.1 ARMA-GARCH Model

The ARMA-GARCH model not only satisfies the mean change process of time series, but also satisfies the random process of residual change, so the ARMA-GARCH model can describe and explain the mean and volatility change of time series data well. The ARMA-GARCH model is given below:

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + \sum_{j=1}^q \psi_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

$$\varepsilon_t = \sigma_t \eta_t \quad (2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (3)$$

Equations (1) is the general form of ARMA(p,q) model, Among them, ϕ_0 is constant, ϕ_i is the autoregressive coefficient of the model, and ψ_j is the moving average term coefficient of the model.

Equations (2) and (3) are general forms of GARCH(m,s) model, in which $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, m; \beta_j \geq 0, j = 1, \dots, s$, $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_j) < 1$. ε_t represents the disturbance term of the mobile phone and $\{\eta_t\}$ represents the sequence of independent distributed mobile phone variables with an average value of 0 and a tolerance of 1.

2.2 Scenario Tree

In the stock market, the basic idea of generating scenarios is to start from the historical returns of assets, explore the relationship between the statistical characteristics of historical data and returns, establish corresponding

*Corresponding author: haiyanxuan@msn.com

mathematical models, and then simulate and predict possible situations according to the established mathematical models, so as to generate scenario elements and build scenario trees. The scenario tree (Figure 1) contains a parent node and a child node. After the parent node, each child node may occur, and the probability sum of occurrence is 1. Each child node contains corresponding information.

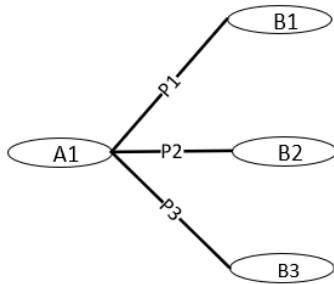


Figure 1. Scenario tree

The common methods of scenario tree construction include random sampling method, cluster analysis method and moment matching method. With the rapid development of computer technology, random sampling has become the most commonly used method. Random sampling method is mainly to generate new scenarios by Monte Carlo simulation of models describing data characteristics. Taking ARMA-GARCH model as an example, according to the known mean and covariance, a number of random numbers with normal distribution are randomly generated to represent the profit scenario at the next moment.

Generally speaking, the distribution of return on assets shows an approximate normal distribution of "high in the middle and bottom at both ends". In order to reduce the sampling area, we define the return of assets as the expected value at the midpoint and the width of 6 times the standard deviation, which can guarantee a large number of values. According to ARMA-GARCH model, the random data of asset returns are obtained, and various assets are defined in N intervals. Because there may be many values in a certain interval, we take the average value of these values as the scenario (yield) of this interval. By counting the frequency in each interval, and then calculating the frequency, the corresponding probability can be obtained, and the scenario tree can be constructed.

2.3 Mean-Variance Model Based on Scenario Tree

The symbols used in the model in this section are described as follows: I represents the number of all assets, and i represents the i -th asset; K represents the number of all child nodes in the scenario number; ω_i represents the allocation weight of asset i ; r_{ki} represents the rate of return of asset i in the k -th node; p^k represents the probability that node k appears.

On each sub-node of the scenario tree, there are the occurrence probability of each node and the return rate of each asset. After each node, there will be three sub-nodes (three cases, which respectively represent the increase, basically unchanged and decrease of the return rate). To determine the sub-nodes, a decision rule is given here, and the subsequent sub-nodes are determined according to this rule, namely:

$$\max(p^k r_{ki}) \tag{4}$$

Mean-variance model for maximizing revenue. Assuming that the investor is a rational person and pursues the maximization of income within the scope of the risk he can bear, then:

$$\begin{aligned} \max R &= \sum_{k=1}^K \sum_{i=1}^I \omega_i * \max(p^k r_{ki}) \\ \text{s. t. } &\begin{cases} \sum_{i=1}^I \sum_{j=1}^I \omega_i \omega_j \sigma_{ij} \leq V_0 \\ \sum_{i=1}^I \omega_i = 1 \\ \omega_i \geq 0, \end{cases} \end{aligned} \tag{5}$$

Where V_0 represents different levels of risk.

Mean-variance model for risk minimization. Assuming that the investor is a rational person who pursues risk minimization within an acceptable range of returns, then:

$$\begin{aligned} \min \sigma^2 &= \sum_{i=1}^I \sum_{j=1}^I \omega_i \omega_j \sigma_{ij} \\ \text{s. t. } &\begin{cases} \sum_{k=1}^K \sum_{i=1}^I \omega_{ki} * \max(p^k r_{ki}) \geq r_0 \\ \sum_{i=1}^I \omega_i = 1 \\ \omega_i \geq 0, \end{cases} \end{aligned} \tag{6}$$

Where r_0 represents different levels of revenue.

3 Empirical Research

3.1 Sample Analysis

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

In this paper, Zhejiang Pharmaceutical (600216), Yabao Pharmaceutical (600351), Tsingtao Beer (600600) and People's Network (603000) are selected as four stocks, the data date is from August 29, 2016 to August 14, 2019, excluding holidays, a total of 2884 days of data, the sample data source is official website, Netease Finance and Economics. $r_t = \log(P_t) - \log(P_{t-1})$ This text indicates the rate of return in terms of the number one factorial of the closing price of the underlying shares, i.e, the rate of return in terms of the closing price of the underlying shares at the time of the indication of t . Where P_t represents the closing price of the stock at time t and r_t represents the logarithmic rate of return. For convenience of description, the return rates of the four

TABLE 1. Descriptive Statistics of Sample Data

Sample	Average Value	Standard Deviation	Skewness	Kurtosis	J-B Statistics	P Value
RZJYY	-0.000346	0.024498	-0.072342	5.544416	194.8496	0.0000
RYBYY	0.000821	0.018923	-0.115425	7.992012	749.2042	0.0000
RQDPJ	0.000615	0.021186	0.719695	5.726061	285.0975	0.0000
RRMW	-0.000379	0.020885	0.311618	8.393019	895.2447	0.0000

stocks of Zhejiang Pharmaceutical, Yabao Pharmaceutical, Qingdao Beer and Renmin.com are expressed as RZJYY, RYBYY, RQDPJ and RRMW respectively.

Using Eviews8.0 software to analyze the statistical characteristics of the four stock returns, the results are shown in Table 1. In Table 1, it can be found that the skewness and kurtosis of the four stock returns do not meet the standard normal distribution (skewness is 0 and kurtosis is 3) assumption, and the J-B statistics of the four stock returns are not 0, and the corresponding P values are all equal to 0, which provides strong evidence for rejecting the original assumption that the return rates are normal distribution.

3.2 Parameter Estimation

Before establishing the ARMA model, it is generally necessary to test the stationarity of the data. After performing a section of difference on the data, the software Eviews is used for stationarity test (ADF test) and it is found that the data are all stationary. Then the autocorrelation of the four data is judged, and it is found that the four data are all correlated, and the initial order is determined. Finally, according to the autocorrelation test chart of the four stock return series, the significance of the coefficient corresponding to the maximum lag variable of each ARMA model and the minimum AIC, SC and HQ information criteria, affected by the length, the parameter results of the ARMA model corresponding to the four stock return series are directly given here, as shown in Table 2.

TABLE 2. ARMA Model of Sample Data

Sample	Parameter	Estimated Value	T-Test	P Value
RZJYY ARMA (2,2)	Intercept	-0.000361	-0.40712	0.6840
	AR1	-0.798184	-30.9753	0.0000***
	AR2	-0.967335	-38.0257	0.0000***
	MA1	0.767231	21.6713	0.0000***
	MA2	0.935301	26.5414	0.0000***
RYBYY ARMA (3,3)	Intercept	0.000832	1.16514	0.2443
	AR2	-0.799701	-67.0253	0.0000***
	AR3	-0.388623	-33.3277	0.0000***
	MA2	0.804278	182.421	0.0000***
	MA3	0.436040	94.0907	0.0000***
RQDPJ	Intercept	0.000615	0.77949	0.4359
RRMW ARMA (1,1)	Intercept	-0.000386	-0.18678	0.8519
	AR1	0.927664	23.0466	0.0000***
	MA1	-0.855988	-15.3815	0.0000***

a. ***, **, and * represent significant levels of 1%, 5% and 10%.

In the financial market, it is generally believed that the simplest GARCH(1,1) model can well depict volatility[5]. We use Eviews software to obtain the GARCH(1,1) model of four stock return series.

Then the GARCH(1,1) model of the RZJYY sequence is as follows:

$$R_{zjyy}(t) = -0.000536 - 0.80012R_{zjyy}(t-1) - 0.996709R_{zjyy}(t-2) + 0.790363\varepsilon_{t-1} + 0.997271\varepsilon_{t-2} + \varepsilon_t$$

$$\sigma_t^2 = 3.86 \times 10^{-6} + 0.034776\varepsilon_{t-1}^2 + 0.959915\sigma_{t-1}^2$$

Then the GARCH(1,1) model of the RYBYY sequence is as follows:

$$R_{ybyy}(t) = 0.000858 - 0.762727R_{ybyy}(t-2) - 0.393678R_{ybyy}(t-3) + 0.795363\varepsilon_{t-2} + 0.44023\varepsilon_{t-3} + \varepsilon_t$$

$$\sigma_t^2 = 1.04 \times 10^{-5} + 0.134519\varepsilon_{t-1}^2 + 0.84227\sigma_{t-1}^2$$

Then the GARCH(1,1) model of the RQDPJ sequence is as follows:

$$R_{qdpj}(t) = 0.000378 + \varepsilon_t$$

$$\sigma_t^2 = 1.97 \times 10^{-6} + 0.030939\varepsilon_{t-1}^2 + 0.966554\sigma_{t-1}^2$$

Then the GARCH(1,1) model of the RRMW sequence is as follows:

$$R_{rmw}(t) = -0.001735 - 0.650103R_{rmw}(t-1) + 0.641844\varepsilon_{t-1} + \varepsilon_t$$

$$\sigma_t^2 = 2.01 \times 10^{-5} + 0.079182\varepsilon_{t-1}^2 + 0.891686\sigma_{t-1}^2$$

In the fitting GARCH model, $\alpha + \beta$ are all less than 1, and the sum of both is close to 1, indicating that the impact received in the GARCH process is continuous.

3.3 Scenario Generation and Scenario Tree Construction

The quality of scene generation will affect the success of decision-making. In order to verify the ARMA-GARCH model of the stock returns, we use Python3.7 software to use Monte Carlo simulation method to compare and analyze the return rate data obtained by the ARMA-GARCH model with the historical data, and obtain the cumulative probability distribution diagram of the simulated data and the historical data.

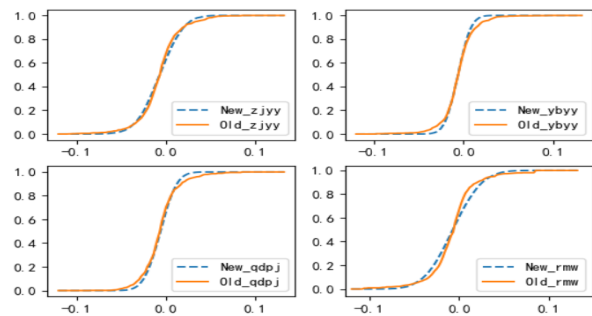


Figure 2. Comparison of cumulative probability distribution of four stocks

In figure2, the black dotted line represents the generated scene data, and the black solid line represents the history data. As can be seen from Figure 2, the cumulative probability curves of the scenario data of the four stock returns rates generated based on the simulation of ARMA-GARCH model can well approximate the cumulative probability curves of the historical data. Therefore, it is effective to use ARMA-GARCH model to generate scenarios.

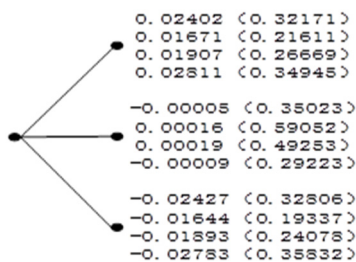


Figure 3. Scenario tree constructed by random sampling method

Based on the scenario elements generated by the ARMA-GARCH model, a scenario tree is constructed by using the sampling method. The tree contains a parent node and three child nodes. The three child nodes represent the increase (yield greater than 0.01), the basic unchanged (yield within [-0.01,0.01]) and the decrease (yield less than -0.01) of the yield respectively. According to the 10,379 scenarios simulated, the frequency of the corresponding scenario is selected, the frequency is taken as the probability of occurrence, and the average return rate in the corresponding scenario is taken as the return rate of occurrence. Among them, there are 8 values on each child node, representing RZJYY, RYBYY, RQDPJ and RRMW respectively from top to bottom. The values in brackets represent the probability of occurrence of the corresponding yield, as shown in Figure 3.

3.4 Model Solving

Take different variable V_0 for the mean-variance model (5) of revenue maximization and different variable r_0 for the mean-variance model (6) of risk minimization, and substitute all the data into Python3.7 software for solution. The results of Equations (5) and (6) are shown in Tables 3 and 4.

TABLE 3. Mean-variance model for revenue maximization

V_0	ω_{rzjyy}	ω_{rybyy}	ω_{rqdpj}	ω_{rrmw}	Revenue	Risk
0.0200	0	1	0	0	0.0130	0.00120
0.0175	0	0.9433	0.0567	0	0.0128	0.00111
0.0150	0	0.8328	0.1672	0	0.0124	0.00095
0.0125	0	0.6828	0.3172	0	0.0112	0.00079
0.0100	0.0778	0.5302	0.3279	0.0641	0.0093	0.00063
0.0088	0.0088	0.4793	0.2103	0.1302	0.0065	0.00058

TABLE 4. Mean-variance model for risk minimization

r_0	ω_{rzjyy}	ω_{rybyy}	ω_{rqdpj}	ω_{rrmw}	Revenue	Risk
0.225	0	1	0	0	0.0130	0.00120
0.200	0	0.8977	0.1023	0	0.01265	0.00104
0.175	0.0109	0.5635	0.4047	0.0210	0.01107	0.00071
0.150	0.0695	0.5344	0.3374	0.0588	0.00949	0.00064
0.125	0.1282	0.5052	0.2701	0.0966	0.00791	0.00060
0.100	0.1802	0.4793	0.2103	0.1302	0.00650	0.00058

In Table 3, as the level of risk acceptable to investors decreases, so does the portfolio return from the investment and the corresponding risk. Moreover, the maximum risk that the investor can bear is 0.0012 and the maximum return is 0.01298. When the risk level exceeds

0.0012, the return from the investment will not increase. In Table 4, as the value of the return expected by the investors decreases, the portfolio return obtained from the investment also decreases, and the corresponding risk also decreases. And the maximum expected return rate of investors is 0.01298, which corresponds to the maximum risk of 0.0012.

In order to better understand the relationship between benefits and risks and to understand the differences between model (5) and model (6), this paper compares and analyzes the effective boundaries of the two models based on the data in tables 3 and 4, as shown in figure 4. In figure 4, the dotted line represents the revenue maximization mean-variance module. The solid line represents the effective boundary of the risk minimization mean-variance model. Due to the limitations of the data, the difference between the two curves is not obvious. As can also be seen from figure 4, the beginning and the end of the two curves are basically coincident, and there is no significant difference, indicating that there is no difference between considering the maximum return when investors can bear greater risks and considering the minimum risk when investors expect the maximum return, which is consistent with the conclusion that large returns are always accompanied by large risks. In the middle part of the two curves, the two curves show certain differences. When investors can bear certain risks, considering the revenue maximization model can obtain higher returns; when the investor's risk tolerance is weak, the risk minimization model can obtain higher returns.

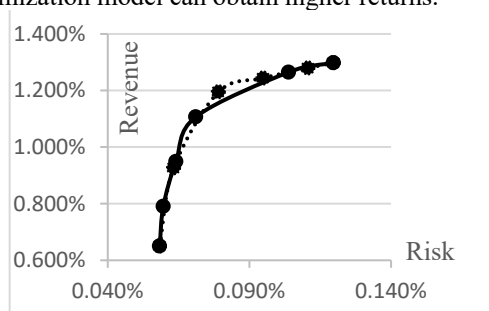


Figure 4. Comparison of effective boundaries between model (5) and model (6). a. The dotted line represents revenue maximization and the solid line represents risk minimization.

4 Conclusions

Based on the scenario tree, this paper improves the traditional mean-variance model, establishes two new models, and verifies the feasibility of the models through empirical research, and analyzes the differences between the models. The scenario tree is constructed by using sampling method to represent the possible future situation, and a decision rule is given to determine the future nodes. Based on this, the mean-variance model for revenue maximization and the mean-variance model for risk minimization are established. By solving them, it is found that both models conform to the general rule that there is a positive correlation between return and risk in investment. In the comparative analysis, it is found that investors with different preferences can choose the corresponding model to achieve their expected investment

benefits. The essence of the investment problem is the choice problem. Using the scenario tree to express the future return rate of the risk assets can be used as a reference for decision-making, and can provide intuitive and clear guidance for decision-makers.

Acknowledgments

Supported projects: projects supported by the National Natural Science Foundation of China (11261031); supported by Gansu Philosophy and Social Science Fund (GL 065).

References

1. Markowitz, H, (1952) Portfolio selection. *J. Financ.*, 7: 77-91.
2. Sadjadi, S. J., Seyedhosseini, S. M., Hassanlou, K. (2011) Fuzzy multi period portfolio selection with different rates for borrowing and lending. *Appl. Soft. Comput.*, 11: 3821-3826.
3. Hoyland, K., Wallace, S. W. (2011) Generating scenario trees for multistage decision problems. *Manage. Sci.*, 47: 295-307.
4. Yan, Z., Chen, Z. P., Consigli, G., et al. (2020) A copula-based scenario tree generation algorithm for multiperiod portfolio selection problems. *Ann. Oper. Res.*, 292: 849-881.
5. Kim, C. B., Yo, K. C., Park, S. A. (2017) The effect of Korea-India AEO-MRA on trade balance and volume using GARCH(1,1)-M model. *The e-Business Studies*, 18: 259-271.