Research Article

Research on Stock Market Portfolio Optimization Using Stochastic Matrix Theory and Genetic Algorithm

YanJin Zhang

Department of Economic Management Engineering, Inner Mongolia Technical College of Mechanics and Electrics, Hohhot, Inner Mongolia 010070, China

Correspondence should be addressed to YanJin Zhang; yjz@imufe.edu.cn

Received 14 June 2022; Revised 7 July 2022; Accepted 15 July 2022; Published 25 August 2022

1. Introduction

Finance is a subject full of vitality. Most people think that the origin of modern finance comes from Markowitz’s portfolio selection theory. Markowitz published a seminal article called “Portfolio Selection” in 1952 [1]. In this article, Markowitz describes how to use portfolios to diversify and maximize returns, and for the first time, under the guidance of classical probability theory, gives a probabilistic-based average—Variance method [2]. A large number of domestic and foreign experts have conducted in-depth discussions on securities investment management. This paper makes an in-depth discussion on the securities investment portfolio from three perspectives: seeking new methods of capital return and risk measurement under unstable conditions, focusing on building a portfolio selection model suitable for various investment conditions, and focusing on the optimization method of establishing the best combination decision-making mode and the corresponding practical application. Portfolio issues refer to investment decisions made under conditions of uncertainty. In recent years, many scholars at home and abroad have done a lot of work in this area [3]. Since the 1960s, many people have conducted in-depth discussions on the relevant theories of securities investment based on Markowitz’s mean-variance analysis [4]. According to Roy’s point of view on portfolios, Kataoka and Telser have successively introduced various types of “safety first” models [5]. In practice, investors often need to adjust their investment decisions to obtain the greatest benefits, which requires investors to regard investment decisions as a multilevel and dynamic issue. Many scholars have successively launched a series of multiperiod portfolio selection models [6]. The development of my country’s securities portfolio theory is relatively lagging behind, and there is still a certain distance from the mature foreign financial market theory [7]. But in recent years, it has shown a thriving trend, and there have been some results. Beyer et al. [8] proposed a specific solution to the optimal problem of securities portfolio and the definition of optimal decision-making, and from the perspective of the matrix, he obtained the risk size of the securities investment portfolio. Tang et al. also conducted a mathematical analysis of the effectiveness limit of portfolio securities under various constraints. The portfolio securities without negative constraints are analyzed. This
2.1. Portfolio Management. Securities investment portfolio refers to the allocation of various assets by investors according to their own needs and effective management of them during the investment process. Investors often weigh the risk of an investment against the expected return, and high return means high risk.

The MV model requires estimates of sampled income and variance. In addition, due to the errors caused by the data used, the accuracy of the variation-covariance matrix of stock returns has a greater effect on the effectiveness of the method. Many experiments have proved that the method contains many disturbances, which makes the optimal solution of this method produces a large offset, which results in an optimal solution to the model. The noise disturbance is eliminated in the portfolio correlation matrix to obtain an efficient security that matches the actual demand.

2.2. RMT-Based Portfolio Optimization. According to the principle of random matrix, in a portfolio, the variance-covariance (correlation matrix) composed of the covariance of securities and income, most of the information it contains is "useless," that is, it can be regarded as "noise" [15]. The RMT algorithm is used to decompose the noise in the correlation matrix and reorganize the corresponding asset portfolio to make it more efficient.

2.2.1. Establishment of Portfolio Correlation Matrix. Assuming that an investment project contains N resources, if the stock price of the t-th resource at t is $S_t(t)$, the rate of return of the asset t is as follows:

$$R_{t(t)} = \ln S_t(t + \Delta t) - \ln S_t(t).$$

(1)

Because different stocks have different volatility, the above returns need to be normalized.

$$r_i(t) = \frac{R_i(t) - \bar{R}_i(t)}{\sigma_i}$$

(2)

Here, $\sigma_i$ is the average return of $R_i$ in a certain period, representing the standard deviation of $R_i(t)$.

MV's portfolio model includes a variation-covariance matrix composed of various stocks, the covariance of each factor is proportional to each factor, and the correlation reflects the correlation of each asset. If each asset has a negative correlation, the correlation is weak and the covariance is small. Therefore, when constructing a portfolio, selecting a security that is positively correlated with it can effectively reduce the overall investment risk. If both bonds are negative, diversification completely removes this danger and is a very effective behavior. However, empirical surveys show that different stocks have a certain positive relationship. Therefore, in order to diversify investment risks, it is necessary to increase the asset size of securities.

2.2.2. Denoising. According to the random matrix method, the correlation matrix C is divided into two parts: "information" and "noise." "Information" and "noise" are defined by the random matrix method. The theoretical maximum eigenvalue and the minimum eigenvalue, the eigenvalue of C is higher than the theoretical maximum eigenvalue and its corresponding eigenvector. The "information" of the eigenvector C and the eigenvalue and eigenvector between the eigenvalue and the theoretical maximum eigenvalue are "noise." On the basis of "noise," the correlation matrix C of the resource is removed, and the correlation matrix of the resource is reconstructed, that is, the new correlation matrix.
only contains some meaningful data. Therefore, compared with the original correlation matrix, a better combination can be obtained with an affinity matrix.

2.3. Stability of the Correlation Matrix. How is the stability of the correlation matrix measured? Krzanowski [16] noted that the degree of overlap of the eigenvectors in two successive cycles determines the similarity (convergence) of the eigenvectors, and that after the vectors are normalized, their number is multiplied by the cosine of the angle of the two vectors. This provides a measure of the degree of overlap of feature vectors. If the two are in the same direction, the remaining string force will become larger, otherwise, it will become very slim. The size of the correlation matrix is larger than the eigenvalue of \( \lambda_{\text{max}} \) and the corresponding eigenvector can obtain more information for the eigenvector. Therefore, to measure the stability of the correlation matrix, Krzanowski proposed to test the influence of small perturbations of eigenvalues on its corresponding eigenvectors \( v_k \), the angle \( \theta_k \) is the angle between the vectors \( v_k \) and \( v_k (p) \), and the cosine of the angle \( \theta_k \). It is determined by the following formula:

\[
\cos \theta_k = \begin{cases} 
\left(1 + \frac{\xi}{\lambda_k - \lambda_{k-1}}\right)^{-1/2}, \\
\left(1 + \frac{\xi}{\lambda_{k+1} - \lambda_k}\right)^{-1/2}.
\end{cases}
\]

3. Empirical Analysis and Discussion

3.1. Performance Comparison of Different Denoising Methods. Jolliffe argues that when the correlation matrix is stable, its characteristic vector and the economic implications of its principal components can be unambiguously explained. Therefore, after removing the noise, the stability of the correlation matrix is particularly critical.

In this study, 80 A-share stocks on the Shanghai Stock Exchange were taken as samples, and the real returns from 2016-06-02 to 2020-12-31 were taken as the research object. In order to study the stability of the covariance (correlation) matrix of asset portfolios using various noise reduction techniques in random matrices, the research was carried out through various noise reduction techniques.

3.1.1. The Stability of the Correlation Matrix. The rank-overlap relationship curve of the combination initial correlation matrix \( C \) is obtained by analyzing the sequential correlation matrix \( C \) of the portfolio returns. This method can better reflect the stability of the correlation matrix. Figure 1 is the rank-overlap relationship curve of the initial correlation matrix \( C \) of the portfolio. It can be seen from this figure that the higher the order of the correlation matrix, the worse its stability.

By eliminating some noise in the correlation matrix, denoising the combined correlation matrix, and then reconstructing a new combined correlation matrix, the rank-overlap relationship curves of the correlation matrices after various noise removals are obtained, as shown in Figures 2–6.

Figures 2–6 show the correlation matrix rank-overlap curve of the noise elimination method MI-MV. It can be seen from the figure that MIII has the best denoising performance, and the correlation matrix column value after denoising is the largest, while MII has the best denoising performance. The stability of the noise algorithm decreases sharply after the first few steps, while the MII increases sharply in the later stages. On this basis, this paper proposes a new algorithm based on MII, which adopts a new noise reduction algorithm.

MIV is based on MIII. It can be seen from this figure that the stability of the reconstructed correlation matrix obtained by the MIV denoising technique is similar to that of MIII and has good stability. Method MV uses the residual method to represent the characteristic quantity of noise. It can be seen from the image that only some orders have high stability. This algorithm ensures that the integrity of the data is maintained to the greatest extent possible, while avoiding the over-rendering of noise.

Figure 7 is the rank-overlap curve of the combined reorganization correlation matrix obtained by the two-point determination method (MV1). It can be seen from the figure that although the stability of the correlation matrix of the noise reduction algorithm is not as high as that of the MIII algorithm, but it also ensures the stability of the reconstructed correlation matrix.

Through the comparison of the above six methods, it is found that the MIII method has the best noise reduction effect, and the methods MIV, MV and MVI are MIV, MII and MVI follow respectively. This paper mainly compares the stability of asset allocation and compares it from the risk of asset allocation and the effective margin.
3.1.2. Minimum Variance Portfolio Risk. Table 1 shows the minimum variance of the asset portfolio, which is denoised by using a variety of denoising techniques. From the data shown in Table 1, it can be seen that compared with other corresponding risk values, the combined risk of MIII after noise reduction is significantly higher than other corresponding risks, and after using MIII noise reduction in Figure 4, the correlation matrix’s stability is the best, which shows that although this algorithm can obtain a high degree of stability, the risk prediction for the portfolio is relatively large, because when determining the alternative factors, the deviation of the slope selection is large, so posed a greater risk of forecasting. Because this denoising algorithm uses a minimum number of algebras as algebras, choosing the minimum number of algebras will have a certain impact on the final effect, and to achieve the best effect, the minimum number of algebras must be set. However, this requires a lot of complex calculations.

Except for MIII, the R1 and R4 values of the other two algorithms are relatively similar. Through the analysis of the rank-overlap relationship between the correlation matrix and R4, it can be seen that both the MIV and MV algorithms can improve the correlation after denoising, while also reducing portfolio risk. Under the minimum variance, the MI and MII noise reduction algorithms are not much different from other noise reduction algorithms, but if the denoised correlation matrix is denoised, it cannot be ensured that the denoised correlation matrix remains stable. The new noise reduction algorithm MVI analyzes the stability of the correlation matrix and the risk of combination. Although it is
not as simple as the MI and MII algorithms, it does not need to set any parameters for noise reduction, so its processing process is much simpler. However, from the perspective of the risk analysis of the investment portfolio, the data in Table 1 is not ideal. Regardless of whether it has been processed by past noise, the prediction of future investment risks is relatively low. There is an important factor in this. The data used are weekly earnings data, and long-term data are used to estimate, which will inevitably reduce the expected risk of investment.

3.1.3. Portfolio Efficient Frontier. Figures 8–13 are the legal bounds for portfolios obtained by applying different denoising treatments (MI–MVI) to the correlation matrix (MI–MVI) of the portfolios. The bound is carried out by the initial correlation matrix. Figure 13 is the effective bound of the security estimated with the correlation matrix reconstructed by denoising. It is the security obtained by the initial correlation matrix Efficient bounds for combinations. Optimal bounds for combinations are found using the denoised correlation matrix. As can be seen from the graph, except for MI and MIV, the rest of the graphs are similar: curve (b) is above curve (a), that is, the risk of noise processing is lower than that without noise reduction.

The calculated value of MI is quite different from other calculated values, but MI is very close to MI in terms of stability, because their replacement unit spacing is 0, so the stability is poor. However, from the perspective of effective bounds, the difference between MI and other methods is the difference in the sum of the substitution factors. MI replaces the noise characteristic with 0, that is, the sum of the substitution factors is 0, and the sums of other substitution factors are similar to noise. MIII has good stability, but after the effective edge denoising, its effective edge (b) value is not lower than the noise-free value, indicating that the effect of the method is not lower than the noise-free condition under the same income. The main reason for this is that other elements in MIII are still determined by subjectivity, which can lead to errors and thus less satisfactory results. At the same time, MIV also needs to set some parameters in the noise reduction process, which affects the stability of noise reduction. In order to obtain better noise reduction performance, the parameters must be optimized, which increases the complexity of the operation. The MVI algorithm is a new noise reduction algorithm. According to the experimental results, it has a very good inhibitory effect on the stability of the reconstructed correlation matrix and the efficient frontier of the portfolio.

3.2. Empirical Analysis of Portfolio Optimization Based on RMT

3.2.1. Minimum Variance Portfolio. The optimal selection problems of portfolio investment include maximizing the benefit of the investment and minimizing the variance. Think of it as a simple way of putting it in terms of minimizing the risk of an investment so that you can get the least amount of risk. This simplified format makes it easy to analyze the effect of noise on securities.

Table 2 shows the risk of the minimum variance portfolio of 50 securities with $Q = T/N = 1, 2, 3, 4, and 5$. The following analysis can be obtained from the information in Tables 2 and 3:
When the value of \( Q(T) \) increases, the noise contained in the portfolio correlation matrix also decreases accordingly.

Regardless of noise exclusion or not, the forecast results for the first three data are higher than that of the “realized” hazard, i.e., the hazard estimates are often overestimated.

Under the conditions of \( Q = 1 \) to 3, the risk of the portfolio with or without noise is lower than that of the portfolio without noise, and through the noise reduction of RMT, a lower investment risk can be obtained.

From the effect point of view, all the predicted risks are smaller than that of the actual realization risks, which are despised because the increase of \( Q(T) \) will weaken the credibility of historical data, so that predictions cannot be made based on these data. In addition, when \( Q = 4 \) and \( Q = 5 \) predict an era of economic recession, the risk of investment is unavoidable, and the prediction of long-term historical data will inevitably lead to the underestimation of the risk of investment.

3.2.2. Portfolio Efficient Frontier. Figures 14–17 are the active bounds for a portfolio of 50 stocks with Qs of 2, 3, 4, and 5, respectively. It can be seen from the four subfigures...
that all curves (b) are located on the left side of curve (a), that is, under the same expected income level, the forecast risk after noise is smaller than that of the forecast without noise reduction.

Comparing the 4 figures, it is found that in the case of $Q = 2$, the best result is obtained. When $Q = 2$, when the investment return reach the same level, the investment risk is the lowest. When $Q = 2$, the optimal security portfolio is in the range of less risk. When the $Q$ value increases, its risk increases and the expected return does not increase. An increase in $Q$ indicates an increase in the duration of historical data. According to RMT theory, when $Q$ increases, the noise contained in the correlation matrix will decrease. However, increasing $Q$ does not make it reach the optimal investment allocation because of its time effect. At the same time, since the model is based on weekly income and has
Table 3: The results of the two groups of parameter evaluation indicators.

<table>
<thead>
<tr>
<th></th>
<th>Elapsed time</th>
<th>SP</th>
<th>Number of assets</th>
<th>Elapsed time</th>
<th>SP</th>
<th>Number of assets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>184</td>
<td>0.8961</td>
<td>59.22</td>
<td>2027</td>
<td>15.4361</td>
<td>48.19</td>
</tr>
<tr>
<td>2</td>
<td>181</td>
<td>0.5907</td>
<td>59.71</td>
<td>1672</td>
<td>14.2136</td>
<td>48.74</td>
</tr>
<tr>
<td>3</td>
<td>182</td>
<td>0.5070</td>
<td>59.77</td>
<td>1690</td>
<td>14.7405</td>
<td>48.06</td>
</tr>
<tr>
<td>4</td>
<td>192</td>
<td>0.3165</td>
<td>59.91</td>
<td>1563</td>
<td>15.4361</td>
<td>48.19</td>
</tr>
<tr>
<td>5</td>
<td>189</td>
<td>0.4063</td>
<td>59.79</td>
<td>1706</td>
<td>14.2136</td>
<td>48.74</td>
</tr>
<tr>
<td>The mean</td>
<td>185.6</td>
<td>0.5433</td>
<td>59.68</td>
<td>1731.6</td>
<td>14.8080</td>
<td>48.38</td>
</tr>
</tbody>
</table>
large time intervals, $Q$ increases, that is, from long-term historical data, the risk of the security is predicted. When the $Q$ value is 3, 4, and 5, with the increase of $Q$, the optimal side length curve of the noise gradually tends to decrease. So the choice of $Q$ depends on the sampling frequency, and RMT is used in RMT to analyze the parameters of high and low frequencies.

4. Empirical Analysis of NSGA-II Genetic Algorithm Parameter Optimization

Table 3 is an empirical study of the evaluation index, and the optimal level is combined as a parameter of its algorithm. As can be seen from the table, the empirical analysis results of A1B5C1 have shorter running time and smaller SP, but the resources in its portfolio have not decreased, while A5B1C5, as the calculation method of calculation parameters, although the calculation period is longer, the SP is larger, but the total amount of resources is reduced. Therefore, under the same evolutionary algebra, comparing SP should select ABC as the parameter of the optimization algorithm from the optimal mode.

Figures 18 and 19 are experimental data for the combination of two optimal levels of A1B5C1 and A5B1C5 as an optimized computational method. The empirical analysis of the optimal combination of A5B1C5 shows that the optimal combination of A5B1C5 is much better than that of A1B5C1. Therefore, when optimizing the parameters of NSGA-II, it is necessary to evaluate it not only from its own characteristics, but also from a specific problem. The conventional test method is not only ineffective, but a kind of blind obedience. Through example analysis, it is proved that it is a feasible way to use orthogonal experiment to determine the reasonable combination of factors.

From the optimal solution, it can be known that when the rate of return and the degree of risk of stocks reach a certain level, this method can effectively select stocks and reduce the number of assets invested in stocks. Before investing, investors should choose different types of investment targets based on various factors such as the macro situation, stock market conditions, and their own investment experience, and then choose the most suitable investment targets according to their own preferences and use the above methods, providing reference for investors to make investment decisions, and can effectively guide the investment type and investment scale of the fund, reducing the investment risk of investors to the fund. At the same time, investors can also make periodic adjustments to their assets through this model.

5. Conclusion

Today, when the capital market is becoming more mature and standardized, quantitative investment has gradually formed a brand-new asset allocation and combination. Quantitative research is also an important feature and development direction of contemporary capital. Based on the actual needs of investors, using NSGA-III random matrix analysis technology and NSGA-III genetic algorithm, the technical process and technical system of securities data analysis, processing, and model solving are established, and the model is tested through comprehensive empirical analysis. The correctness and feasibility provide theoretical and technical support for the management of securities investment portfolio. This paper conducts an empirical study on investment portfolios containing multiple securities and finds that diversification can effectively reduce the risk of investment, but when the investment scale reaches a certain level, this effect becomes negligible. Using this method, we can reduce the funds in securities investment and reduce the management of securities investment, thus verifying the correctness and effectiveness of the model.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.
Conflicts of Interest

The author declares no conflicts of interest regarding this work.

References