Quantitative Analysis in Finance: Leveraging Statistical Methods for Improved Investment Decisions

Mingshen Liu[1,a,*]

[1]Arizona State University, Tempe AZ 85281, USA
a. mliu159@asu.edu
**corresponding author*

Abstract: Since modern times, with the vigorous development of the financial market, the explanation of some complex finan-cial phenomena can no longer be based on the qualitative analysis of the subjective judgment, but on the quantitative analysis of statistical models. This paper first analyzes and demonstrates the application and importance of statistics in the financial field from three perspectives of risk quantifi-cation, price prediction as well as portfolio optimization. Secondly, this paper points out that the traditional statisti-cal theory cannot meet the current practical needs in the above three aspects, and lists a series of relevant studies. Finally, this paper gives some possible solutions to the cur-rent problem based on existing research and personal un-derstanding and evaluates some existing solutions. The re-search conclusion of this paper is expected to provide a reference for the practical application of statistics in the fi-nancial field and provide some possible ideas for the re-search direction of related interdisciplinary.

Keywords: statistics, finance, risk quantification, price forecast, portfolio optimization

1. Introduction

Statistics and economic phenomena are the foundation of the field of finance. The financial market has become increasingly complex over the past few years as a result of ongoing financial instrument development and upgrading, as well as the increasing use of quantitative statistical research techniques in the financial sector. First, statistics plays an important role in the process of financial risk management. In the process of development, the financial sector will be affected by policies, markets and other factors, which will lead to certain risks. The occurrence of risks will often have a chain reaction, which will have an impact on the financial sector and various industries. In order to ensure the sound and orderly development of the financial market, it is necessary to strengthen its risk management and make corresponding risk prediction and risk response plans. The statistical method can analyze the data to obtain the appropriate investment management data, and can avoid the risks arising from investment to a certain extent. Now, with the development of network technology, there are a large number of statistical software, and the software can be combined with the financial field in the process of use, which has changed the huge amount of calculation and tedious calculation process in the past data analysis, and provided the basis for the prediction and prevention of financial risks. Secondly, statistical models are widely used in the fields of price forecasting analysis and portfolio optimization. In the financial field, the value of data is extremely important, and statistical analysis of data is a necessary means to realize the value of data. Through statistical

DOI: 10.54254/2754-1169/25/20230478

sampling and volatility analysis of financial data, it can significantly improve the accuracy of decision-making and achieve positive returns. With the continuous development of statistics, statistical methods applied in the fields of price forecasting and portfolio optimization emerge endlessly, such as the cointegration test, ECM model, non-stationary time econometric model and other statistical methods. At the same time, due to the rapid development of the financial market, the market competition and price fluctuations in the financial field have become increasingly fierce, and achieving accurate price judgment and rapid combination optimization has become the first step in the financial competition. For example, the stock futures industry in the financial field has realized the use of the CJR model statistical analysis method to analyze the leading financial situation from a micro perspective, which can effectively predict the future direction of the stock market. At present, the popular quantitative investment strategy in the financial market is to use statistical models for data analysis to capture price differences and obtain sustainable and stable returns. Finally, it is also worth noting that statistics has become the core required subject of the courses related to finance offered by many universities and institutions. All these indicate that statistics has gradually become one of the main methods of financial practice and theoretical research.

This paper focuses on using statistics in the financial industry to quantify risks, predict the future value of assets, and optimize investment portfolios. At the same time, this paper also points out a series of problems in the application of statistics in these three fields, and puts forward corresponding improvement and solution suggestions, in order to serve as a guide for the upcoming advancement of statistics in the financial industry.

2. Application Analysis

2.1. Risk Quantification

Financial systemic risk has the characteristics of wide impact, strong chain reaction, and great social harm. Statistics has the characteristics of strong data analysis and clear indicator quantification and can carry out data analysis and risk prediction on financial markets from various angles. Using statistical analysis to prevent financial systemic risk is a common risk prevention method. For instance, a mean-variance model is able to track the occurrence of systemic risk in the financial markets as well as offer early warning. Moreover, the network statistics and sequence distance of Bayesian network time series can also improve the accuracy of the prediction of extreme absolute returns, so as to evaluate the system risk [1]. Moreover, LSSVM is also a method to predict systemic financial risks. The suggested LSSVM can provide superior prediction performance and promotion compared to conventional approaches like SVM, BP neural network, and logical regression [2]. In addition, institutional and individual investors, after dealing with systemic risks through effective diversification, also need to consider the unique risks of each stock. Wang noted that the evaluation outcomes are frequently near to reality and the accuracy is relatively high when the GARCH model is applied to analyze company-specific financial investment risks [3]. Moreover, the sensitivity analysis model can also be used to quantify and evaluate the company-specific non-systematic risks. In general, as computer and mobile internet technology has advanced, a variety of intelligent and effective statistical software has appeared. This software allows for the tracking and statistical prediction of financial data in real-time, avoids data lag and complex calculations, and significantly enhances the timeliness and effectiveness of risk response. Statistics has become an important means to assess and manage financial risks. However, some studies have pointed out that people still tend to rely more on personal experience and judgment rather than statistical models when evaluating financial risks, which is more obvious in the financial markets and small investment markets in underdeveloped regions [4]. In addition, when using statistical models to evaluate financial risks, it is necessary to select a specific probability distribution, which will be arbitrary. That is to say, this

DOI: 10.54254/2754-1169/25/20230478

step is only based on the choice of statisticians or risk managers based on personal experience or preferences. It is difficult to prove its rationality through more objective East and West. This may lead to the result that the quantitative results obtained cannot reflect the real situation of risk very well. Finally, at present, scholars have studied many statistical models for assessing financial risk, but most existing models neither combine fuzzy sets with quantitative analysis nor take into account the historical data of the past few years.

2.2. Price Forecast

How to accurately predict the future price of assets is crucial for investors, but it is also a very complex problem, which requires comprehensive use of statistical methods and machine learning technology. Generally speaking, the current basic statistical forecasting methods are divided into the following two categories: first, the method based on time series analysis, that is, by analyzing historical data, building a time series model to predict future prices, volatility and other indicators. Secondly, the method based on Bayesian model is to use Bayesian theory to model historical data and market indicators with probability to predict the future price trend. More specifically, many scholars have proposed many specific advanced statistical methods for predicting the future price trend of assets. For instance, a statistical signal prediction model known as HMM has been extensively utilized to forecast stock values. Additionally, the feature fusion LSTM-CNN model, which integrates the traits discovered from several representations of the same data (such as stock time series and stock chart images), may also provide accurate stock price predictions. Additionally, the RVFL-GMDH statistical model can be utilized to accurately anticipate the turning point of the executive stock price after determining the turning point of the time series by the analysis and application of Chaos theory. In addition, Zuo and Kita compared the Bayesian method to other time series prediction algorithms and AR, MA, ARMA, and ARCH models, and used a Bayesian network to simulate the random dependence between past stock prices in order to predict future stock prices [5]. They came to the conclusion that the Bayesian algorithm has higher prediction accuracy than conventional time series prediction algorithms when comparing correlation coefficient and root mean square error. All of these examples demonstrate how statistical models have been extensively utilized to forecast stock prices in the future. Yet, the stock market has deviated from the conventional financial theory that holds that stock market price follows the fundamental law of "random walk" due to the emergence of several "abnormal phenomena" and numerous statistical prediction models. In various sample periods and markets, these price forecasting models are not always consistent and trustworthy. The reality is that many forecasting models can only serve as research tools for academics and cannot actually generate excess returns for investors. In addition, these statistical price forecasting models themselves have certain limitations. For example, time series analysis only uses time as the analysis factor, does not take into account the impact of other factors and only forecasts based on historical data, without taking into account the possibility of market changes. According to Hwang and Oh, using the conventional time series analysis model to predict stock prices is problematic due to the difficulty in simulating the uncertainty of the stock market [6]. The Bayesian model's flaws are more glaring. The Bayesian model makes the unworkable assumption that the qualities are independent of one another in real applications. In addition, the prior probability, which frequently depends on assumptions, must be known for the Bayesian model. Because of the presumptive prior model, the prediction effect may occasionally be subpar.

2.3. Portfolio Optimization

Another important application of statistics in the financial field is to optimize the portfolio. Securities investors should consider two basic factors when investing: one is income, and the other is risk. Most

DOI: 10.54254/2754-1169/25/20230478

investors want to maximize the return on investment and minimize the risk of investment. Because portfolio investment can reduce the risk of investment without reducing the income. As for portfolio investment, many scholars have established various optimization models. The conventional Markowitz mean-variance optimization theory no longer adequately describes the modern portfolio creation methods required by practice, and these newer methodologies call for the deployment of more robust and scalable numerical optimization techniques. For example, Ta et al used a variety of portfolio optimization models in their research, including mean variation optimization (MVO), Monte Carlo simulation (MCS), and equal-weighted modeling (EQ), to improve the performance of the portfolio [7]. The findings reveal that the optimized portfolio's return and Sharp ratio have significantly improved when compared to the S&P 500 index. Also, in their research, Tken et al. expanded the conventional Markowitz mean-variance portfolio optimization model by using the new two-step heuristic approach GRASP&SOLVER [8]. Moreover, Boyd et al. suggested that the convex optimization model can resolve the issue of multi-period portfolio optimization [9]. Earlier, Celikyurt and Oze-kici proposed that different models of the safety-first approach, coefficient of variation, and quadratic utility function can be dynamically programmed to solve the multi-period portfolio optimization problem, if the random evolution of the market is described by a perfectly observable Markov chain [10]. However, most existing statistical optimization methods do not take into account the constraints of the real world such as regulation or liquidity, which may lead to the concentration of portfolio weights on a small part of assets in the portfolio. When the process of portfolio optimization is affected by other constraints such as taxes, transaction costs and administrative expenses, the optimization results obtained by only using statistical models may deviate greatly from the actual situation. For example, the law may prohibit investors from holding certain assets, or sometimes the tax costs associated with holding certain assets are too high. In this case, appropriate constraints must be imposed on the optimization process. In addition, because the best portfolio will change over time, there is a motivation to often re-optimize, and changing the weight of the portfolio requires transaction costs. Too frequent transactions will lead to too frequent transaction costs. Therefore, the best strategy may also include finding the frequency of re-optimization and transactions.

3. Recommendation

3.1. Risk Quantification

First of all, as for how to popularize the use of financial risk statistical models in the financial markets and small individual investor groups in underdeveloped regions, an effective solution may be to accelerate the construction of a complete financial system by the local government. This includes not only strengthening the training and introduction of financial statistical technical personnel but also strengthening the publicity of the concept of quantitative investment risk by the financial functional departments of the government. Secondly, the previous article mentioned that when using statistical models to quantify financial risk, we must make assumptions about the probability distribution of the existing data in advance, which may lead to the deviation of the quantitative results from the actual situation to a large extent. To solve this problem, the machine learning method may be an effective solution. Now, the content of machine learning is quite rich. Many models do not require the same distribution of samples or the probability distribution of samples. Another possible solution is to measure the distribution of data samples in advance, rather than judge by experience. The specific operation method is to use BP neural network for distribution recognition after feature extraction of data samples. In addition, Zhang et al pointed out that the fuzzy mixed fractional Brownian motion model with jumps is an effective method to incorporate fuzzy sets into the statistical model [11]. In addition, when quantifying financial risk, using a random forest algorithm based on fuzzy

DOI: 10.54254/2754-1169/25/20230478

mathematics and constantly checking and testing the results is also an effective way to solve this problem [12].

3.2. Price Forecast

The efficient market theory is fundamentally at odds with the employment of statistical models to anticipate stock values, as was already mentioned. The short-term price forecasting model is stable even in an efficient market, but the stable forecasting model is unlikely to survive for a long time, according to Timmermann and Granger [13]. When majority investors find it, it will quickly self-destruct. In addition, the limitations of using the time series analysis model to predict the stock price have been described in the previous article. It should be noted that the future state of economic events cannot be just a simple repetition of the past, so the time series analysis is only applicable to short-term or medium-term forecasts. The turning point in future development and change is frequently difficult to anticipate, thus it is often required to integrate other approaches, particularly the qualitative prediction method, to build a thorough prediction that will have the desired outcome. Also, it is not possible to mechanically extend outward in accordance with the historical and current laws of market phenomena while analyzing and forecasting time series. The new qualities and new performance of market phenomena changes must be studied and analyzed, and the forecast value must fully take into account these new characteristics and new performance. It can only predict the market phenomena in this way with sufficient accuracy to account for both past changes in the phenomenon's behavior as well as current performance. Finally, there are four main ways to improve the naive Bayesian model at present: first, structure expansion, whose main direction is to explore the relationship between attributes; The second is attribute weighting (to solve the problem that each attribute has the same impact on the category decision and to assign a weight value to each feature attribute). Specific processing methods include the MCMC method, the information gain method, the mountain climbing method, and the combination of the MCMC method and the mountain climbing method. You can also start from the rough set theory, introduce the information entropy and conditional entropy of information theory, and determine the weight by the importance of attributes. Chi-square statistics can also be used to construct the correlation coefficient between attributes as the weight value; The third is local learning, that is, cutting the original data set into several subsets, and using some of them for the classification algorithm. The principle of local learning is that part of the data in the training set that does not meet the naive Bayesian hypothesis can meet the requirement of no dependency between attributes; The last is attribute selection (the method of removing redundant attributes). The specific operation method is to use a greedy strategy to select attributes to improve the performance of the classification algorithm. You can also use rough set theory to find a decision contribution index that can measure the contribution of a single attribute feature to a category, that is, association rules, and based on this eliminating the attribute feature with the contribution degree.

3.3. Portfolio Optimization

The previous article pointed out that when the existing statistical model is used to optimize the portfolio, the impact of transaction costs and other factors is basically not considered, which will lead to the theoretical optimization results may not be the actual optimal plan. The use of a stochastic differential equation and stochastic control theory to determine that the value function of optimal control is the smooth solution of its corresponding HJB equation based on the current optimization model, to demonstrate the existence of the optimal strategy, and to determine the weight composition of its optimal portfolio is one potential solution for incorporating the impact of transaction costs into the modeling process of portfolio optimization [14]. Another option is to utilize the following semi-absolute deviation to calculate the investment risk, create a nonlinear 0-1 fractional portfolio optimization model, and apply the chaotic differential evolution algorithm to solve the model to create the ideal portfolio. Also, when building the portfolio, certain constraining factors like legislation and liquidity must be taken into consideration. When modeling under these restrictive conditions, most of the existing optimization models are too sensitive to the changes in input variables, and slight changes may significantly affect the weight ratio of the optimal portfolio. A possible solution is to use the transfer coefficient to measure the impact of these restrictive conditions on portfolio performance. In this process, the transfer coefficient can be defined as the correlation coefficient between the risk-adjusted weight of the portfolio and the predicted return rate. When the optimization model takes into account the influence of some restrictive factors, there will no longer be a certain proportion relationship between the risk-adjusted weight and the alpha, so the transfer coefficient will become an important parameter to measure the matching degree of the portfolio weight and investment income. Of course, this method also has certain defects. For example, when there are multiple restrictive factors, the transfer coefficient can only reflect the overall impact of these factors in a general way, but cannot reflect the impact of each restrictive factor separately. In view of this, the shadow cost decomposition method may be an effective solution. This method can decompose the total impact of multiple restrictive factors into the impact of each factor on portfolio risk and return. Finally, it should be pointed out that when the input value of the optimization model itself has a large estimation error, the reliability of the output result of the model will be greatly reduced. At present, the actual value of input variables required by most portfolio optimization models is difficult to predict accurately. This problem can be solved by Bayesian technology, the Black-Litterman model, a robust optimization model, etc.

4. Conclusion

In general, statistics has become an important method for risk quantification, price prediction, and portfolio optimization in the financial market. But at the same time, due to the high complexity and unpredictability of the financial market, the traditional statistical theory cannot fully reflect the actual financial market performance. That is to say, many conclusions obtained from statistical research on the financial market at present do not have practical significance and cannot be really applied to practical operations. Even in some cases, the conclusions drawn are completely contrary to the actual market performance, which will play a negative guiding role for investors. In view of this, this paper provides some operational possibilities and inspiration on how to solve these problems, starting from some of the more advanced statistical theories at present. In the current and future era of big data, the theories and methods of statistics will be introduced into the financial field. Moreover, the integration of financial and statistical applications will also be more complex. The analysis of the application of statistics in the financial field in this paper may be limited to several specific angles and will continue to be a more in-depth analysis and exploration in the future.

References

[1] *Chan, L. S. H., Chu, A. M. Y., & So, M. K. P.: A moving-window bayesian network model for assessing systemic risk in financial markets. PloS One, 18(1), e0279888 (2023).*

[2] *Zhao, D., Ding, J., & Chai, S.: Systemic financial risk prediction using least squares support vector machines. Modern Physics Letters. B, Condensed Matter Physics, Statistical Physics, Applied Physics, 32(17), 1850183 (2018).*

[3] *Wang, X.: Mathematical Model of Quantitative Evaluation of Financial Investment Risk Management System. Mathematical Problems in Engineering (2022).*

[4] *Masood, O., Aktan, B., & Chaudhary, S.: The investment decision-making process from a risk manager's perspective: a survey. Qualitative Research in Financial Markets, 1(2), 106–120 (2009).*

[5] *Zuo, Y., & Kita, E.: Stock price forecast using Bayesian network. Expert Systems with Applications, 39(8), 6729–6737 (2012).*

[6] *Hwang, H., & Oh, J.: Fuzzy models for predicting time series stock price index. International Journal of Control, Automation, and Systems, 8(3), 702–706 (2010).*
[7] *Ta, V., Liu, C., & Tadesse, D. A.: Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. Applied Sciences, 10(2), 437 (2020).*
[8] *Ötken, Ç. N., Organ, Z. B., Yildirim, E., et al.: An extension to the classical mean–variance portfolio optimization model. The Engineering Economist, 64(3), 310–321 (2019).*
[9] *Boyd, S., Busseti, E., Diamond, S. et al..: Multi-Period Trading via Convex Optimization (2017).*
[10] *Celikyurt, U., & Ozekici, S.: Multiperiod portfolio optimization models in stochastic markets using the mean–variance approach. European Journal of Operational Research, 179(1), 186–202 (2007).*
[11] *Zhang, W. G., Li, Z., Liu, Y. J., & Zhang, Y.: Pricing European Option Under Fuzzy Mixed Fractional Brownian Motion Model with Jumps. Computational Economics, 58(2), 483–515 (2021).*
[12] *Zhu, J., Geng, Y., Li, W. et al.: Fuzzy Decision-Making Analysis of Quantitative Stock Selection in VR Industry Based on Random Forest Model. Journal of Function Spaces, 1–12 (2022).*
[13] *Timmermann, A., & Granger, J.: Efficient market hypothesis and forecasting. International Journal of Forecasting, 20(1), 15–27 (2004).*
[14] *Cheng, X., Huang, R., & Dai, H.: Study on Optimal Investment and Consumption Model Considering Tax and Transaction Fee in Incomplete Markets. Journal of Liaoning University, 31(4), 320–323 (2004).*