Prediction Research and Application of Financial Time Series Based on Big Data

Rui Wang*
School of Economics, Shanghai University, Shanghai, China

*Corresponding author e-mail: wangr@shu.edu.cn

Abstract. Financial risk is conducive in the market. In this paper, correlation function is used to model the relevant structure of the financial market. The j-B test statistic value of four stocks is far greater than the critical value of 6.7325, and the critical value of LJung-Box-Pierce test is 32.1342. At the same time, it is proved that the selection of edge distribution function is independent of the dependent structure, but the risk value estimated by semi-parametric POT model is higher than that estimated by kernel density function.

Keywords: Time Series, Financial Forecast, Financial Risk, Venture Capital

1. Introduction
The price fluctuation of products in the financial market is a complex and changeable nonlinear process, which is closely related to the various economic and political activities of human beings in the financial market and shows the characteristics of dynamic changes. Investors' judgment on the price fluctuation pattern of financial products will have an important influence on their own investment behavior [1].

The applied research of time series has been widely concerned by scholars in various fields, and the prediction of time series is also a hot issue. How to establish a prediction model with high accuracy and low computational complexity is the key to the prediction of time series. For the financial field, the pursuit of high-precision numerical results of financial time series prediction models often consumes a lot of energy and material resources, but the results may not be satisfactory. However, the research on the prediction of price fluctuations of financial time series often gets twice the result with half the effort. Because, for financial investors, they often prefer to obtain the short-term price fluctuation trend of financial time series rather than precise value, which is more conducive to investors' decision-making. Therefore, it is particularly important to study the volatility of financial time series. Its serial data can be generated by various financial markets such as stocks, exchange rates, funds, etc. Volatility is one of the inherent characteristics of financial time series, which is influenced by economy, politics and even human psychology. Since the volatility analysis of financial time series can effectively obtain or predict the state of the financial market, it is of great significance for investors to make correct investment strategies and reduce investment risks. Therefore, the study of financial volatility has also attracted more and more researchers' attention. For example, in 2010, Brandt et al. studied market indexes of the three major U.S. stock markets, analyzing the effect of...
market and company size on stock price volatility. Then, it makes a further empirical analysis of the volatility of retail stock prices in emerging markets. In 2013, Waild et al. used futures price data to establish GARCH model family to analyze the volatility of crude oil stock price. In 2016, the volatility of futures market is discussed by using ARFIMA-Garch model. In financial markets, where there is uncertainty about the rise and fall of share prices, what investors want most is a forecast for the future of share prices. Research shows that rational use of stock historical prices can help people analyze the future volatility of stock prices, which can provide valuable information for investors. In this process, Markov model provides an important mathematical analysis model. The state division method in Markov chain is used to analyze the future state of stock prices, which simplifies the workload and facilitates the analysis and research of investors on the stock market [2-4].

In this paper, time series are divided, fitted and clustered to construct fuzzy information granule. Furthermore, the fuzzy information granule is used to construct the granule complex network to present the fluctuation pattern of financial time series. Among them, as nodes of the network, information grains represent wave patterns with similar characteristics, and directional edges represent the transmission process of wave patterns. A new oriented weighted network community detection algorithm is proposed to divide the community into complex granular networks. By using the distance between information nodes in complex networks, this algorithm can adjust the intimacy degree of nodes adaptively, and then achieve the purpose of community division.

2. Elaboration of Relevant Concepts

2.1. Basic Knowledge of Complex Networks

A Complex Network is a group of networks with certain characteristics. Complex networks are not man-made, but come from the real world. In order to better study of the network in real world, people tend to convert them into collection easier to understand by the edge and vertex set figure, figure set of vertices (or nodes) on behalf of all sorts of different types of entities in the real world the collection of information, and the edge collection represents the complex relationship between each entity, can turn in the real world networks after more intuitive to reflect the relationship between the various entities information. At the same time, with the popularization of computers and the increasing number of Internet users, the scale and number of the network have become increasingly large, and various connections in the complex network have become more and more complex, and the exploration of complex network has also been favored by many scholars and experts. Although the network models obtained from different fields (such as physics, biology, medicine and computer, etc.) may seem different, they have many similarities in network characteristics [5].

Small world is one of the important characteristics of complex network. The small-world feature of complex network means that although there are a large number of nodes and edges in the network, any node can be associated with any node in the network with a small number of connections. In real social networks, we summarize this phenomenon as the theory of six degrees of space, that is, in real social circles, any person wants to contact with any stranger in the world, there are at most six people between them. In a particular application scenario, the small-world nature of a complex network can be exploited to dramatically change the performance of the network by simply changing a few connections in the network [6].

Complex networks usually have community structures. As we say, people cluster by type and things cluster by group, which is a kind of clustering characteristic of nodes in complex network. In a complex network, the nodes in the network can be divided into different "groups" or "clusters". The nodes in these "groups" have more frequent contact or more similar properties than the nodes between "groups". These "groups" or "clusters" are called community structures. Within each community, the number of connections between each node is large and the information transmission is relatively frequent. Among different communities, there are fewer connections and less information transmission between nodes in different communities [7].
2.2. VaR and ES Measures of Financial Market Risk
Risk is one of the fundamental characteristics of the financial system. With the globalization of financial economy and the continuous improvement of financial technology, the measures of financial market risk have become more and more extensive. Among them, VaR has become the most important and the most common benchmark to measure risk in the international financial market. Since VaR also has certain limitations and the description of tail information is not perfect, the researchers proposed a spectrum risk measurement model based on ES model based on consistent risk measurement [8].

According to the definition of VaR, the following mathematical formula is used to represent the maximum possible loss of the investment portfolio:

\[ \text{Prob} (\Delta p \leq -V a R) = 1 - \alpha \]  

Although VaR plays an important role in the measurement of risks, in actual research, VaR is not the most reasonable and effective method. It ignores the difference of extreme value losses. However, in the research and analysis of financial crisis events, it is a crucial point in the research and analysis of the occurrence of low-probability events. Aiming at the defects of VaR method, many scholars want to seek a new risk measurement, which can not only have better performance than VaR but also make up the deficiency of VaR. In view of this, an improved model based on consistent risk measurement framework, Expected Shortfall model, is proposed to overcome the shortcomings of VaR and replace risk measurement, which is also called conditional VaR [9]. Its expected loss ES is:

\[ ES = -E (R|q(\alpha)) = -\alpha^{-1} \int_{-\infty}^{q(\alpha)} r f(r) dr \]  

Return on financial Assets:

\[ r_t = \ln(1 + R_t) = \ln \frac{p_t}{p_{t-1}} \]  

2.3. Tail Dependence
The copulas connect theory in the process of continuous development and improvement, MiaoBai its people find copulas connect models such as particularly effective in capturing the tail end of the time series correlation, TsafackG Copul et al., this paper proposes a state transition model, what kind of copulas connect model can be used to measure the extreme correlation between international stock market and bond market structure, studies have shown that the tail end of the same type of international asset has a strong dependence, especially in the asymmetric state. Chollete analyzes the correlation and tail dependence of international stock markets, and studies show that Latin American stock markets have asymmetric tail dependence. Okimoto established a multi-state smooth transition CopulaGARCH model and studied the asymmetric growth trend of tail dependence in the international stock market through this model. Liang Feng Zhen etc further stock and external financial assets yield sequence for the asymmetric volatility clustering characteristics of the study, results show that compared with the foreign exchange market, the stock market volatility asymmetry more apparent Su studies focus on the Taiwan and South Korea and the United States, Japan, China and European stock markets tail dependence. Reboredo and Ugando used a VineCopula model to study whether there was a correlation between the prices of gold, silver, platinum and gold. The results showed that different precious metals exhibit different average correlations and tail correlations [10].

Some domestic scholars have also applied Copula model to the measurement of tail dependence. In previous studies, li and cheng used a variety of Copula models to study the tail correlation between the sse composite index and the hang seng index, using gumbel-hcopula to calculate the upper tail dependency and ClaytonCopula to calculate the lower tail dependency. Jiang Hongli, He Jianmin et al. combined the extreme value theory with GARCH and Copula models to study the dynamic tail correlation between China's real estate industry and banking industry, and indicated that THE SJC model could better model and analyze the tail dependence. They also found that the tail dependence
had autocorrelation and ARCH effect. Yi Rong combined the mixed Copula model and the GARCH model, and studied the tail dependence between the futures of 9 major agricultural products in China. In addition, in order to solve the asymmetry problem of tail dependence, Wu Jin-lin et al. combined the mixed Copula function with the multi-state smooth transition function and put forward a mixed Copula model. The research shows that the tail dependence among China's stock markets presents obvious asymmetry.

3. Forecast and Research Analysis of Financial Time Series

3.1. Design of Self-trending Flow Network Structure

The core idea of the self-trending flow network is to extract from the stock's own historical data the depth characteristics that are helpful to predict the future stock price trend. Therefore, the Wavelet module is used to deal with the basic market data in stock history. Meanwhile, the DAE module is used to extract the depth characteristics of common technical indexes in quantitative trading. Finally, the processed data is put into the LSTM network to predict the closing price of the next day.

3.2. Noise Reduction Autoencoder Module

Due to the variety of technical indicators, and the need to carry out complex analysis of them in order to make a forecast of the possible trend of stock prices in the future. Therefore, it is necessary to extract in-depth features of technical index data to effectively reduce the complexity of analysis, and it is hoped that potential, essential and predictive trend features of stock prices can be extracted from technical index data.

As a neural network model commonly used for feature extraction, noise reduction autoencoder has been widely used in text translation, image recognition, behavior recognition and other fields. The noise abencoder attempts to reconstruct the original "clean" data from the "contaminated" data. Therefore, the noise abencoder can usually be used to extract strong robustness features from the original data. This section first introduces the common structure of the auto-encoder to explain the principle that the auto-encoder can achieve feature extraction. Then on the basis of ordinary autoencoder, the introduction of noise reduction autoencoder is introduced.

4. Analysis of Test Results

4.1. Research on the Stock Risks of Small and Medium-sized Enterprises Sector

| Table 1. Basic Statistical Analysis of Four Stock Return Series |
|----------------|--------------|----------------|------------------|
|                | Huadong technology | FAW car | Tuopai Qujiu | Shuanghe pharmacy |
| average value  | 0.0734       | 0.2593 | 0.1983    | 0.3011           |
| Standard deviation | 3.7922  | 3.3947 | 3.4241 | 3.3859 |
| skewness        | -0.4011 | 0.0034 | -0.2481 | -0.1367 |
| kurtosis        | 5.6511 | 4.3050 | 4.8224 | 4.1773 |
| J-B Inspection  | 194.9022 | 43.2640 | 90.6498 | 37.1117 |
| Q(20)           | 11.9970 | 27.9239 | 24.9771 | 16.1234 |
| D-W value       | 1.9259 | 1.8886 | 2.0081 | 2.0624 |

As can be seen from Table 1, during the observation period, the average returns of the four stocks are all positive; Except for FAW car, the skew value of the other three stocks is negative, which
indicates that the earnings of the other three stocks except FAW car are likely to fall; According to the kurtosis values in Table 1, it can be seen that the income distribution of these four stocks shows the feature of thick tail. The value of J-B test statistics in the four stocks is far greater than the critical value 6.7325, indicating that the return sequence does not obey the normal distribution. As the critical value of LJung-Box-Pierce test is 32.1342, it can be seen from the Q (20) value in the table that these four stocks all have conditional heteroscedasticity. The d-W statistic values of the four stocks are all close to 2, indicating that the autocorrelation of their return series is very weak, so the mean value equation of the return series below does not consider the autoregression term.

**Table 2.** Normal Copula Correlation Matrix

|        | HDKJ | YQJC | TPQJ | SHYY |
|--------|------|------|------|------|
| HDKJ   | 1.000| 0.4845| 0.5285| 0.4863|
| YQJC   |      | 1.000| 0.4439| 0.4468|
| TPQJ   |      |      | 1.000| 0.4564|
| SHYY   |      |      |      | 1.000|

According to the sample points, Copula parameters were estimated and simulated 10,000 times to obtain the mean, maximum variance and minimum variance of simulated sample values of Copula function, as shown in Figure 1 below.

**Figure 1.** Statistical Values of Simulated Samples under Different Dependent Structures
4.2. Expected Earnings Statistics

Figure 2.

Conclusion 1: Under the same dependency structure, the VaR value of the asset portfolio at the same confidence level is significantly lower than the ES value at the same confidence level. This is because ES measures the average loss over VaR.

Conclusion 2: Under the same dependency structure and at the same confidence level, VaR and ES of portfolio assets will increase with the increase of portfolio return rate. This also suggests that the greater the expected return, the greater the risk of a portfolio.

Conclusion 3: At the same confidence level, the dependence structure among financial assets is a nonlinear dependence structure. If only the lower tail correlation structure of financial assets is considered and the upper tail is not considered, the risk may be underestimated; on the contrary, if only the upper tail correlation structure is considered and the lower tail is not considered, the risk may be overestimated.

5. Conclusion

Financial risk plays an important role in the investment portfolio. Studying the risk-dependent structure, especially the tail-dependent structure among variables, is helpful to improve the accuracy of decision making. The study shows that Copula function can effectively measure the dependence between variables. Based on Copula function, this paper conducts empirical research on four stocks of small and medium-sized enterprises, real estate and financial industry, and Shanghai and Shenzhen stock markets respectively. As a new statistical analysis tool, Copula function has become a key word in recent years by virtue of its unique excellent performance. The more Copula functions are applied to various domains. On the basis of Copula theory, this paper discusses the risk management among financial markets in a small scope.

References

[1] Haan L D, Cécile Mercadier, Zhou C . Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. Finance and Stochastics, 2016, 20(2):321-354.
[2] Cheng C H, Wang S H. A quarterly time-series classifier based on a reduced-dimension generated rules method for identifying financial distress. Quantitative Finance, 2015, 15(12):1979-1994.

[3] Zhu F, Quan W, Zheng Z, et al. A Bayesian Learning Method for Financial Time-Series Analysis. IEEE Access, 2018, PP (99):1-1.

[4] Chung S S. A class of semiparametric volatility models with applications to financial time series. Dissertations & Theses - Gradworks, 2016, 123(3):439-444.

[5] Aryal S, Nadarajah D, Rupasinghe P L, et al. Comparative Analysis of Deep Learning Models for Multi-Step Prediction of Financial Time Series. Journal of Computer Science, 2020, 16(10):1401-1416.

[6] He H, Hong Y, Liu W, et al. Data mining model for multimedia financial time series using information entropy. Journal of Intelligent and Fuzzy Systems, 2020(1):1-7.

[7] Li H, Cui Y, Wang S, et al. Multivariate Financial Time-Series Prediction With Certified Robustness. IEEE Access, 2020, PP (99):1-1.

[8] Marchenko G, Gagliardini P, Horenko I. Towards a Computationally Tractable Maximum Entropy Principle for Nonstationary Financial Time Series. SIAM Journal on Financial Mathematics, 2018, 9(4):1249-1285.

[9] Assis C A S, Machado E J, Pereira A C M, et al. Hybrid deep learning approach for financial time series classification. Revista Brasileira De Computação Aplicada, 2018, 10(2):54-63.

[10] Bozic J, Babic D. Financial time series forecasting using hybrid wavelet-neural model. The international arab journal of information technology, 2018, 15(1):50-57.