Analysis and Construction of EEMD Smart Model and Fuzzy Forecasting through Improved Bayesian Estimation

Liyao Tang1, *, Zijian Cui2, Kaishuo Liu3

1School of Management and Engineering, Capital University of Economics and Business, Beijing, 100070, China
2School of Reliability and System Engineering, Beihang University, Beijing, 100089, China
3Minzu University of China, Beijing, 100080, China

*Corresponding author: tangliyao@cueb.edu.cn

Abstract. Different portfolio models will have a great impact on the model rate of return. In order to meet the data needs of portfolio investment forecasting, a big data generation method based on incremental Bayesian network model is proposed. Firstly, the fuzzy interval number is constructed by using the global GM (1,1) prediction model, then the fuzzy Mmurv model is established based on the interval fuzzy number, and the model is optimized based on the midpoint, radius and acceptability of the interval number, thus a single objective programming model with parameters is obtained. The future data is partially generated by using the time series generation algorithm, and the Bayesian network model trained by historical data is updated with the newly generated data. So that the updated Bayesian network can reflect the relationship between the variables and the laws contained in the new and old financial data in this period of time. And use the model to predict the amount of foreign direct investment absorbed by China in 2020, test the robustness and analyze the impact of major random events to show the reliability of the prediction. The analysis is carried out on multiple scales. The results show that the integration of IMF1 and IMF2 components through the white noise test can be regarded as the random impact of external factors; the cycle length of IMF3 component is about 3 years, which can be regarded as the inventory cycle caused by the change of enterprise inventory; the cycle length of IMF4 component is 5 years, which can be regarded as political cycle; the cycle length of IMF5 component is about 20 years, which can be regarded as construction cycle. The set of portfolio paths is generated by path search algorithm in Bayesian network, and the big data set with real data characteristics is generated according to the probability distribution of each path. The experimental results show that the method is feasible and ensures a certain accuracy. The forecast results show that the amount of foreign direct investment in China will continue to fluctuate seasonally and in a trend in recent years.

Keywords: Financial portfolio, fuzzy forecasting, stability test, multi-scale
1. Introduction

Theory and practice show that economic growth often presents a cyclical cycle characterized by expansion, contraction, recession and recovery. Both the academia and the political circles are actively exploring the internal mechanism of the economic cycle fluctuations, looking for good measures to deal with the sharp economic fluctuations, so as to reduce its negative impact as far as possible. The research on periodic fluctuation of economic time series focuses on time domain analysis and frequency domain analysis.

Real financial data are generally expressed in the form of time series. At present, the main methods of time series data simulation and generation are: regression moving average model, autoregressive conditional heteroscedasticity model, support vector regression model, long-term and short-term memory network model and so on. As the data generated by the simulation in this paper show obvious quarterly or monthly periodic changes, the seasonal differential autoregressive moving model will be used to automatically simulate and generate the time series data. Supplement part of the future data for the original historical data set and update the subsequent portfolio investment data generation model.

In this paper, the overall GM (1,1) prediction model of the interval fuzzy number is used to predict the return rate and liquidity of stocks, and the corresponding interval fuzzy number is obtained. On this basis, a portfolio selection model based on interval fuzzy number is constructed, and the validity of the model is verified by an example analysis.

The financial data used in this paper is essentially a kind of time series data. So, when it comes to generating extensions to a dataset, the data can be generated by training the time series model. At present, there are time series models and time series analysis methods, which reveal the law of phenomenon change with time through the historical statistical data of the series, and extend the law into the future according to the need. By giving different economic meanings according to the component characteristics, we can identify the different cycles of fixed asset investment fluctuation and study the sub-series of different time scales, so as to find out the hidden cycle fluctuation law of fixed asset investment fluctuation and dig out the deep reason of fixed asset investment fluctuation. The subsequent prediction analysis on this basis can, on the one hand, enrich the application research of EEMD model in economic sequence analysis and prediction in theory, and on the other hand, improve the accurate understanding of China's fixed asset investment fluctuation cycle in practice, and provide more scientific support for the formulation and introduction of macro-control policies.

2. Time series generation algorithm

As one of the manifestations of data, time series data truly records all kinds of important information at different time points (or time slices), which contains rich and valuable knowledge. This paper uses SARIMA to generate multiple columns of financial time series data. SARIMA model is derived from the Differential Autoregressive Moving Average (ARIMA) model. In the ARIMA (p, d, q) model, p is the number of autoregressive terms, q is the number of moving average terms, and d is the number of difference times for time series to become stationary series. If the time series \( \{Y_t\} \) is a non-stationary series, the ARIMA model can be expressed as:

\[ \Phi(B) \Delta^d Y_t = c + \Delta(B) \epsilon_t \quad (1) \]

Wherein, \( \Delta^d = (1 - B)^d \), B represents the hysteresis operator, \( BY_t = Y_{t+1} \), and \( \Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots \) is the autoregressive coefficient, c is the constant term, \( B = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \) is the moving average coefficient, the random term \( \epsilon_t \) is an independent white noise sequence, and obeys the normal distribution \( N(0, \sigma^2 \epsilon) \), and the random term is correlated with the hysteresis variables \( Y_{t-1}, Y_{t-2}, \ldots, Y_{t-P} \) is not correlated.

SARIMA (P, D, Q) (P, D, Q) S model is mainly used to analyze the time series with periodic changes caused by periodicity (including weekly, monthly, quarterly, annual, etc.) or other factors, and to carry out seasonal difference based on periodicity on ARIMA model. Let the change period of seasonal series be s, and the seasonal difference operator is defined as
\[ \Delta s = 1 - B^s \]  

\( B_s \) is s step lag operator. If the seasonal time series is expressed by \( \{Y_t\} \), the primary seasonal difference is expressed as

\[ \Delta s \cdot y_t = (1 - B^s)y_t = y_t - y_{t-s} \]  

If \( \{Y_t\} \) is non-stationary seasonal time series, D-order seasonal difference is required, and a P-order autoregressive Q-order moving average seasonal time series model with a period of S is established

\[ \Phi(B) \lambda_p(B^s) \Delta^d \psi(B^q) Y_t = c + \Theta(B) \psi(B^q) \varepsilon_t \]  

In the existing SARIMA model building methods, the stationarity test and processing of the data are usually carried out first: to observe whether the sequence is a stationary time series; if the sequence is non-stationary time series, ordinary difference is required for the original sequence and seasonal difference, and thereby determine the parameters d, D; By observing the truncation or trailing characteristics of the autocorrelation and partial autocorrelation functions of time series. Most of the above steps require manual observation to determine the value of parameters, and for different time series data, the parameters of the model will be different. However, through literature reading and many experiments, it is found that the range of values of p, q, d, p, q and d is relatively conventional. Therefore, grid search method is used in this paper to determine the optimal parameter combination of the model. Then, the determined parameter combination is used to check and test the model. If the relative error between the predicted value and the actual value obtained from the data in the practical test set is less than 5%, it indicates that the established model has a high accuracy, and the model can accurately simulate and generate future data.

3. Bayesian Network algorithm for data Generation

The incremental learning method is used to update the Bayesian network model. The main idea of this method is that a certain amount of data samples are randomly selected from the historical data and combined with the new data set to train the Bayesian network structure, and then the Bayesian network structure is combined with the historical model to perfect the Bayesian network structure.

![Figure 1. Historical data generated by a Bayesian network](image)

Due to the inconsistency of time and frequency of data, we resampled the mean value according to the month. In addition, due to the need of constructing Bayesian network, the data is discretized. First, make first-order difference for data. For time series data \( \{X_t\} \), the p-order difference operation formula is
\[ \Delta^p X_t = \Delta^{p-1} X_t - \Delta^{p-1} X_{t-1} \]  

(5)

4. Model result

The data values of central bank deposit, loan interest rate and deposit reserve ratio are relatively stable, and the weight random generation method is adopted, which is more in line with the reality than random generation method. In this experiment, data of 3 years are generated.

![Data generation results](image1)

Figure 2. Data generation results

SARIMA algorithm is used for the remaining data to generate new data for subsequent Bayesian network updates. Since there are many types of data generated, the first column of data currency and the change amount of quasi-money supply \( M_2 \) are taken as examples to show the generation process of new data in the next 3 years through learning historical data by using SARIMA time series algorithm. Since the processed data is relatively stable but the value is large, Min-Max standardization is used to process the data in the experiment.

After data processing, parameters of Sarima \((p, d, q)\) \((p, d, q)\) s model were determined. Since the first-order difference has made the data stable, \(d = d = 1\), and \(s\) is the period of time series, monthly data are used in this experiment, and \(s\) is 12. Algorithm 1 is used to determine the remaining parameters, and the determined parameter combination is put into the model to predict the two-year data.

![Data prediction results](image2)

Figure 3. Data prediction results

![Sequence diagram of the sequence DY after difference](image3)

Figure 4. Sequence diagram of the sequence DY after difference
In order to make the sequence stable, the original sequence Y is processed by first-order difference, and the sequence DY is obtained. From figure 4, we can see that the sequence DY after difference always fluctuates up and down a constant value, and the trend is eliminated. Further unit root test shows that the original hypothesis that there is a unit root is rejected at this time, and the sequence DY is stable.

It can be seen from the seasonally adjusted time series diagram that the fluctuations affected by seasonal factors decrease, but the truncation and trailing trends of autocorrelation and partial autocorrelation graphs are still not obvious, and attempts to fit the ARIMA model have poor results. It is observed that AC and PAC are still large at the 12th order delay, which is still obvious in the figure. It can be considered that there is still seasonal effect in the seasonally adjusted sequence. Therefore, the short-term correlation of this sequence has complex correlation with seasonal effect, which cannot be extracted simply.

![Timing diagram of seasonally adjusted sequence DY12](image1)

**Figure 5.** Timing diagram of seasonally adjusted sequence DY12

![Forecast effect of each annual model](image2)

**Figure 6.** Forecast effect of each annual model

It can be seen that although the actual value is different from the predicted value, the relative error is small, the average error is 3.5%, and the prediction accuracy is good. For the sake of safety, the monthly data of FDI from 2002 to 2019 are predicted by ARIMA model to see the overall fitting effect. Due to the difference of the data, the range of predicted samples is reduced. It can also be seen that the covariance ratio of the model is 0.9865), which is much larger than the bias ratio and variance ratio, so the fitting effect of the model for each annual historical data is also ideal as a whole.

5. Discussion and analysis

In the short term, in the face of major random events such as the epidemic, it is imperative to support and help foreign enterprises to resume normal production and operation. We need to solve the difficulties
of resuming work and production of foreign enterprises in a timely manner, and implement preferential policies such as taxation to offset the impact of the epidemic. At the same time, we will ensure the implementation of major landmark foreign investment projects, make full use of various investment agencies and platforms, and continue to promote investment promotion and investment attraction. In the long run, the first step is to further expand the scope of foreign investment encouragement and guide the industrial restructuring of foreign direct investment. While the outbreak has impacted international economic and trade activities, it has also brought opportunities for China to accelerate the development of science and technology and promote industrial optimization and upgrading. It is conducive to promoting foreign investment in bio-medical and high-tech fields.

6. Conclusions
The prediction results in the sample show that the fitting effect of the model is good, and the model is used to predict the amount of foreign investment in China in recent months, and the results show that the future FDI inflow in China will maintain a relatively stable combined fluctuation state of seasonality and trend, and the time series generation algorithm can generate multi-column time series data automatically and accurately. The incremental Bayesian network is constructed, so that the Bayesian network can be updated dynamically with the generation of new data, and it is convenient to generate long-term accurate data later; the path set of each node is generated by path search of Bayesian network, and the big data set is generated on the basis of the value of each path node in the set and its corresponding probability. EEMD model is used to decompose the fixed asset investment series in multi-scale, and each component is divided into random shock, inventory cycle, political cycle, construction cycle and long-term trend items according to its fluctuation frequency or cycle length. Finally, the robustness test is carried out, and the impact of major random events is analyzed to further test the reliability of the prediction results.

References
[1] Yang Long Hao, Wang Suhui, Ye Fei Fei, Liu Jun, Wang Ying Ming, Hu Haibo. Environmental investment prediction using extended belief rule-based system and evidential reasoning rule [J]. Journal of Cleaner Production, 2021289 (prepublish).
[2] Liu Yin, Fan Ti. Research on Intelligent Prediction of Science and Technology Innovation Board Enterprise Investment value-- Analysis from the Perspective of Financial performance [J]. Price Theory and practice, 2020 (06): 109-112.
[3] Wang Mianbin, Qi Xia, an Lei, Geng Pengyun, Wang Jianjun. Science and Technology Investment Forecast of Power Grid Technical renovation Project based on VMD-SVM under the condition of small sample [J]. Science and Technology Management Research, 20200.40 (16): 231-235.
[4] Ciaran Driver, Nigel Meade. Enhancing survey-based investment forecasts [J]. Journal of Forecasting, 2019, 38 (3).
[5] Shuang-qing LIN, Rui WANG, Ming-wei LI, Jian-jia ZHOU, Yan DENG, Tao LUO, Ji-yang ZHANG. Power Investment Prediction Based on the Improved GM (1 dint 1) Model [P]. DEStech Transactions on Engineering and Technology Research, 2018.