Research on Intelligent Analysis and Prediction Model of Funds Flow Based on Non-stationary Time Series

Qiang Yue¹, Zhongyu Hu² and Dongping Li¹,*

¹School of Information Engineering, Kunming University, Kunming 650214, Yunnan Province, China
²School of Mechanical and Electrical Engineering, Kunming University, Kunming 650214, Yunnan Province, China

*Corresponding author email: lidongping2021@126.com

Abstract. Funds flow prediction is critical to the stable operation of financial service institutions. In view of the non-stationary characteristics of funds inflow and outflow in financial institutions, this paper introduces the key characteristics of time series analysis, compares AR, MA, ARMA and ARIMA models, and proposes a funds flow prediction model based on ARIMA algorithm. The stationary test, pure randomness test and model order determination of time series are explained in detail. Finally, combined with the error and score of the model, the relatively optimal model is selected to predict the test data. The experimental results show that, for non-stationary time series, the model has smaller prediction error and can effectively predict funds flow.

Keywords: Funds flow, Prediction, ARIMA model, Non-stationary time series.

1. Introduction

In the current background of "Smart Finance", funds flow prediction has become an important task for financial service institutions, and the analysis and statistics of funds flow are more and more important to financial service institutions. Users can purchase or redeem at any time according to their own financial needs, if the redemption amount is greater than the reserved amount of institutions, liquidity risk will occur, which will reduce the performance of the institutions. In more serious cases, institutions may face the threat of bankruptcy; When the redemption amount is less than the reserved amount of institutions, liquidity surplus will be generated, which indicates that the institutions have not made full use of the funds and have not maximized their interests. Therefore, it is very important for the operation of financial service institutions to explore a more reasonable funds flow prediction method by analysing the historical behaviour data of users [1].

2. Time Series Analysis

Time series refers to a group of digital sequences that arrange the observed values of random events in time order, such as the stock price, monthly precipitation, rail transit passenger flow, power load and so on, all form a time series. Time series analysis is to observe and study time series, find its change and development regular pattern, and predict its future trend.

Time series have the property of dynamic and random changes. It seems to be chaotic on the surface, but in fact it has certain statistical regularity. In order to establish an appropriate model for the time series studied, we must understand the basic statistical characteristics of the time series and the basic theory and method of time series analysis, so as to ensure the reliability and confidence of the time
series model, and meet certain accuracy requirements. Generally, we can consider the stationarity, pure randomness and seasonality of time series [2].

2.1. Stationarity of Time Series
If the statistical characteristics of a time series do not change with time, the following two conditions are satisfied:
1. For any time t, its average value is constant.
2. For any time t and s, the autocorrelation function and autocorrelation coefficient only depend on the time interval t-s, and have nothing to do with the starting and ending points of t and s.
Such time series are called stationary time series. It can also be considered that if a time series has no obvious upward or downward trend, the observed values fluctuate around its mean value, and the mean value is a constant relative to time, then the time series is a stationary series.

2.2. Pure Randomness of Time Series
If there is no correlation between the sequence values, it means that the sequence is a sequence without memory, and the past behaviour has no influence on the future development. This kind of sequence is called pure random sequence. From the point of view of statistical analysis, pure random sequence has no analysis value. If a time series is a pure random series, it means that there is no regularity in the series and there is no correlation between the terms of the series.

2.3. Differential Operation
Suppose \( \{ x_t, t \in T \} \) is a time series, \( \nabla \) is a differential operator, and 1-order differential operation is the subtraction between two series in a period, i.e.

\[
\nabla x_t = x_t - x_{t-1}
\]

It is called 2-order differential operation to perform another differential operation on the sequence after the 1-order differential operation, i.e.

\[
\nabla^2 x_t = \nabla (x_t - x_{t-1})
\]

By analogy, the sequence after (p-1)-order differential operation is further divided by one order, which is called p-order differential operation, i.e.

\[
\nabla^p x_t = \nabla^{p-1} (x_t - x_{t-1})
\]

An important function of differential operation is to obtain stationary time series. Suppose that the time series, \( x_t = \beta_0 + \beta t + \varepsilon_t \), have a long-term trend, where \( \varepsilon_t \) is white noise with zero mean. Therefore, there is \( E(x_t) = \beta_0 + \beta t \), its mean value changes with time and is not a stationary time series. But after the differential operation, \( \nabla x_t = x_t - x_{t-1} = \beta_t + \varepsilon_t - \varepsilon_{t-1} \), the condition of stationary time series is satisfied.

3. ARIMA Model
Time series analysis is divided into stationary time series analysis and non-stationary time series analysis according to its stationarity. AR (Autoregressive) model, MA (Moving Average) model and ARMA (Autoregressive Moving Average) model are applied to stationary time series; ARIMA (Autoregressive Integrated Moving Average) model and seasonal model applied to non-stationary time series. AR model, MA model and ARMA model are only suitable for modeling and forecasting based on stationary time series, ARIMA model is suitable for modeling and forecasting based on non-stationary time series [3]. ARIMA model is the combination of AR model, MA model and ARMA model. AR model uses the previous data to predict the future trend. MA model reduces the prediction error as much as possible. AR model and MA model are combined to form ARMA model [4].
3.1. AR, MA and ARMA Model

P-order autoregressive model is a very effective stochastic model describing time series, which is abbreviated as AR (p), as shown in (4):

\[ x_t = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t \]  

(4)

where \( \phi_0 \) is a constant, \( p \) is the order of AR model, \( \phi_i \) is autoregressive coefficients, \( \epsilon_t \) is white noise series with a mean value of zero, that is \( \epsilon_t \sim WN(0, \sigma^2) \).

Another important model to describe observation time series is called q-order moving average model, which is abbreviated as MA (q), as shown in (5):

\[ x_t = \mu + \sum_{i=1}^{q} -\theta_i \epsilon_{t-i} + \epsilon_t \]  

(5)

where \( \mu \) is a constant, \( q \) is the order of MA model, \( \theta_i \) is moving average coefficients, \( \epsilon_t \) is white noise series with a mean value of zero, that is \( \epsilon_t \sim WN(0, \sigma^2) \).

The model with the following structure is called autoregressive moving average model, which is abbreviated as ARMA (p, q), as shown in (6):

\[ x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} \]  

(6)

In particular, when \( p = 0 \), it is the MA (q) model; when \( q = 0 \), it is the AR (p) model.

3.2. ARIMA Model

The above models can only deal with stationary time series. But in reality, most time series are non-stationary. Therefore, the analysis of non-stationary series is more common and important. The analysis methods of non-stationary time series can be divided into deterministic time series analysis and stochastic time series analysis. There are two main problems in the deterministic factor decomposition method. One is that the deterministic factor decomposition method can only extract strong deterministic information, which is a serious waste of stochastic information. The other is the deterministic factor decomposition method, which attributes the changes of all sequences to the comprehensive influence of the factors, and always fails to provide a clear and effective method to judge the exact relationship between the major factors. The stochastic time series analysis method is to make up for the deficiency of deterministic factor decomposition method, and provide people with more abundant and more accurate time series analysis tools.

If the d-order differential operation \( W_t = \nabla^d x_t \) of a time series \( \{ x_t \} \) is ARMA(p, q), i.e.

\[ W_t = \phi_0 W_{t-1} + \phi_1 W_{t-2} + \ldots + \phi_p W_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} \]  

(7)

Then equation (7) is called autoregressive integrated moving average model, which is abbreviated as ARIMA (p, d, q) model. In particular, when \( d = 0 \), it is the ARMA (p, q) model.

4. Modeling and Experiments

For the model establishing and data prediction, we use R language as backend of data analysis. The experimental dataset comes from the funds flow of a financial service institution in China, the institution has accumulated a large number of records of user funds inflow and outflow. Based on the institution's demand to accurately forecast the amount of funds inflow and outflow, we set the goal of the prediction model as: to predict the total redemption amount of the institution in August, and compare it with the real value to get the relatively optimal model. The process of prediction model is shown in Figure 1, which mainly includes the following procedure.
4.1. Data Extraction
The data set contains the purchase and redemption information of the institution, and we extracted the records from July 1, 2013 to August 31, 2014, with a total of more than 2.8 million records. It was divided into two sets, the training set and the test set. A new data set was created to count the redemption amount of funds on the same date.

4.2. Testing Stationarity
There are two methods to test the stationarity of time series. One is the graph test method based on the real characteristics of time series graph and autocorrelation graph, and the other is hypothesis-testing of constructing test statistics. This paper mainly uses the graph test method to test the stationarity of funds outflow data.

1. Sequence diagram checking
According to the property that the mean and variance of stationary time series are constant, the time sequence diagram of stationary time series should show that the series always fluctuates randomly near a constant value. The sequence diagram of redemption data is shown in Figure 2.

2. Autocorrelogram checking
Autocorrelogram is a plane two-dimensional coordinate hanging line graph. The abscissa represents the number of delay periods, the ordinate represents the autocorrelation coefficient, and the hanging line represents the autocorrelation coefficient. The autocorrelogram of redemption data is shown in Figure 3.

Due to the requirements of the model, it is necessary to transform non-stationary time series into stationary time series. There are two common methods.
- Capture the stationary part of the data through the sequence diagram
- Differential operation

It can be seen from the data sequence diagram that the data after March 2014 fluctuates randomly near a value. We decided to select the data from March to July 2014 as the training data and the data from August 2014 as the test data. The sequence diagram and the autocorrelogram of the intercepted training data are shown in Figure 4 and Figure 5.

Make a differential operation for redemption with a period of 7 days, and the sequence diagram and autocorrelation diagram are shown in Figure 6 and Figure 7.
4.3. Establishing ARIMA Model

4.3.1. Determining order of Model. Before establishing ARIMA model, we need to determine the order of the model, that is, the value of p and q. The BIC criterion is used to determine the order. The BIC diagrams of the original sequence and the differential sequence are shown in Figures 8 and 9.

4.3.2. Model checking. After constructing 10 series models based on BIC diagram, it is necessary to check the pure randomness of residuals to ensure that the fitting model is effective. The residual values of p calculated of the models are greater than the significance level value of 0.05, indicating that the residual values are independent, and all models have passed the check.

4.3.3. Model evaluation. The average error of each day is calculated as shown in (8):

\[
\text{error} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{x}_i - x_i|}{x_i}
\]

(8)

where \(x_i\) is the actual value on day \(i\) and \(\hat{x}_i\) is the predicted value on day \(i\).

In order to judge the effect of the model, the score of daily predicted results is defined, as shown in (9):
The daily average error and score of the models are shown in Table 1.

Table 1. The average error and score of the models.

| Model No. | p   | q   | P   | Error | Score | Model No. | p   | q   | P   | Error | Score |
|-----------|-----|-----|-----|-------|-------|-----------|-----|-----|-----|-------|-------|
| 1         | 0   | 3   | 0   | 0.2169| 4.6472| 6         | 0   | 4   | 3   | 0.2133| 4.5988|
| 2         | 0   | 3   | 1   | 0.217 | 4.6342| 7         | 3   | 3   | 0   | 0.2037| 4.8902|
| 3         | 0   | 3   | 3   | 0.2133| 4.6833| 8         | 3   | 3   | 1   | 0.2172| 4.4566|
| 4         | 0   | 4   | 0   | 0.2148| 4.6391| 9         | 3   | 3   | 3   | 0.2091| 4.6415|
| 5         | 0   | 4   | 1   | 0.2147| 4.6329| 10        | 3   | 4   | 0   | 0.2177| 4.4306|

4.3.4. Predicted results. The predicted results of the relative optimal model are shown in Figure 10.

5. Conclusion
AR, MA and ARMA models are applied to stationary time series, but in reality, most of the time series are non-stationary. This paper starts with the historical records of users' purchase and redemption, and establishes a funds flow prediction model by analysing the non-stationary series of purchase and redemption. This model can not only effectively predict the large amount of fund redemption in the future and prevent liquidity risk, but also avoid excessive funds idle and liquidity surplus. It is an important application in the background of "smart finance".

Acknowledgments
This work was supported in part by the Joint Special Foundation Project of Basic Research in Local Universities of Yunnan Province (Project No. 202001BA070001-197) and the Science Research Foundation Project of the Education Department of Yunnan Province (Project No. 2016ZDX097).

References
[1] Jadhav V, Chinnappa Reddy B V, Gaddi G M 2017 Application of ARIMA Model for Forecasting Agricultural Prices Journal of Agricultural Science and Technology 5 pp 981-992
[2] Wadi S A L, Almasarweh M and Alsaraireh A A 2018 Predicting closed price time series data using ARIMA Model Modern Applied Science 12 11
[3] Chen P and Lu W Q 2016 Analysis on the Application of ARIMA Model in Regional Population Prediction Contemporary Economics 6 pp 111-113
[4] Valipour M, Banihabib M E and Bebahan S M R 2013 Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir Journal of Hydrology 476 pp 433-441
[5] Liu H, Tian H Q and Li Y F 2012 Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction Applied Energy 98 pp 415-424