Financial Index Prediction based on ARIMA Model

Zhuohong Han
Mathematics and statistics, Central South University, Changsha, 410083, China.
8201180913@csu.edu.cn

Abstract
In this paper, the Shanghai Stock Exchange Index is used as the training data, and the time series prediction method is used. First, detect the stationarity of the Shanghai Stock Exchange Index time series data, and then the time series which do not satisfy the stationarity are processed by difference method. After that, establish an ARIMA model, where the figure of ACF and PACF offers the probable values of parameter and the parameter is chosen through AIC and BIC the minimum criterion, to further fit historical stock data. Eventually, through training, the predicted values are divided into long-term, medium-term and short-term forecasts which are compared together so as to conduct segment forecast. Experimental results show that the ARIMA model can extract historical stock information well, can make short-term forecast for stock trend, and has certain reference value for enterprises, investors or market regulators.

Keywords
Time Series Analysis; ARIMA Model; Shanghai Stock Index Forecast; Segment Forecast.

1. Introduction
With the development of global economy, more and more people participate in financial investment, especially with the rapid development of Chinese economy, China’s financial world has become more exciting and complex. Stock is an important path of financial investment, attracting a large number of retail investors and institutions. In order to improve the trading winning rate, it is of great significance to predict the stock price. Stock price prediction is a process of quantitative analysis of stock market. At present, the main research methods include traditional investment analysis method and modern statistical method, such as time series analysis [1], trend extrapolation method, exponential smoothing method and artificial intelligence method [2]. Based on time series analysis, a ARIMA model is established to predict Shanghai stock index. The Shanghai stock index reflects the changes in the price of listed shares on the Shanghai Stock Exchange and is of great significance to the overall evaluation of A and B shares.

2. Arima model brief
Autoregressive Integrated Moving Average Model is composed of Differential Autoregressive Moving Average Model, Auto Regressive Model and Moving Average Model. The formula of ARMA(n,m) is as follows

\[ X_t - \sum_{i=1}^{n} \phi_i X_{t-i} = \alpha_t - \sum_{i=1}^{m} \theta_i \alpha_{t-i} \], among which \( X_t \) is a time series, \( \phi \) are autoregressive coefficients, \( \alpha_t \) is a white noise sequence.

When the time series is not stable, it is necessary to smooth the sequence by differential processing, so it is necessary to introduce the ARIMA (n, d, m) model, in which the d order difference is \( \nabla^d X_t \), and the difference sequence is a stationary time series, m is the autoregressive order, and n is the moving average.
3. Data acquisition and preprocessing

3.1. Data acquisition

Stock is an important type of securities, its quantity is huge, variety. When constructing the time series of stock data, it is faced with the situation that there are many selection objects and a wide range. This situation is not conducive to targeted research. Therefore, when selecting stock data, screening stock market composite index is a good choice [3]. This paper analyzes the Shanghai Stock Exchange Index as time series data by using tushare financial website with python interface.

![Figure 1. Sequence diagram of the closing price of 2752 trades](image1)

![Figure 2. Sequence diagram of the closing price of 1295 trades](image2)
3.2. Data preprocessing

The stationarity test [4] of data is the premise of establishing ARIMA conditions. If it is not satisfied, it is necessary to adjust and correct the stationarity of time series data.

By using python software to visualize the Shanghai Stock Exchange index sequence, this paper selects the closing price of Shanghai Stock Exchange index from January 2010 to April 2021. The result is shown in figure 1.

Here we see that there are large abnormal fluctuations between 2015 and 2016, so we select the stable sequence after 2016 as the training set to establish the model.

The selected Shanghai Stock Exchange index sequence is shown in figure 2.

In order to smooth the sequence, the difference sequence is obtained by first order difference. Sequence Visualization is shown in figure 3.

![First Order Difference](image)

Figure 3. First Order Difference

The first order difference sequence of Shanghai stock index is stable, so the parameter d of the ARIMA (n, d, m) model is determined to be 1.

4. Building arima model to forecast shanghai stock index

4.1. Building arima models

We can see from [5] relevant literature that the ARIMA model has the following properties in selecting autocorrelation coefficient and partial autocorrelation coefficient, as shown in Table 1.

<table>
<thead>
<tr>
<th>model</th>
<th>ACF</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(n)</td>
<td>Decay to zero</td>
<td>n rear end</td>
</tr>
<tr>
<td>MA(m)</td>
<td>m rear end</td>
<td>Decay to zero</td>
</tr>
<tr>
<td>ARMA(n,m)</td>
<td>After the n order, the decay tends to zero</td>
<td>After the m order, the decay tends to zero</td>
</tr>
</tbody>
</table>

When estimating the parameters of the ARIMA model, the above properties can be combined. When the n and m satisfies the above description in the ACF diagram and the PACF diagram, it can be considered that this parameter is more suitable for the model, in which the truncated representation falls into the confidence interval.

The corresponding figure of ACF and PACF are obtained according to the sequence, shown in figure 4.
The observed obtained auto correlation coefficient tends to zero after the third order, which is valued in the [11,13] interval. The most appropriate order n and m are further determined by AIC and BIC the minimum criterion[6]. The result is as follows.

Table 2. Result of AIC and BIC for different models

<table>
<thead>
<tr>
<th>model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(11,1,2)</td>
<td>12872.454</td>
<td>12949.937</td>
</tr>
<tr>
<td>ARIMA(11,1,3)</td>
<td>12865.521</td>
<td>12948.169</td>
</tr>
<tr>
<td>ARIMA(12,1,3)</td>
<td>12868.407</td>
<td>12956.220</td>
</tr>
<tr>
<td>ARIMA(13,1,3)</td>
<td>12866.215</td>
<td>12959.194</td>
</tr>
</tbody>
</table>

Therefore we can draw the conclusion that it is better to choose ARIMA(11,1,3) according to AIC and BIC the minimum criterion.

For the convenience of calculation and expression, the B of back-shifting operator is introduced [7]. Finally we obtain the formula:

\[ (1-B)X_t = c + \sum_{i=1}^{11} \phi_i B^i + \left(1 - \sum_{j=1}^{3} \theta_j B^j\right)\alpha_t \]

Where "c=0.16
\[ \phi_i = (-0.88,-0.93,-0.81,0.02,-0.02,-0.13,-0.12,-0.11,0.00,0.01,0.04) \]
\[ \theta_j = (0.84,0.97,0.87) \]

### 4.2. Model prediction

In verifying the forecast effect, the data from 2016 to 2020 are used as the training set to predict the trend of the Shanghai Stock Exchange index in 2021, to observe its trend and to observe its accuracy.

Visualization of predictive is shown in figure 5.

The entire time span is from September 2019 to May 2021. The orange line is the true value of Shanghai Stock Index, while the blue line is the prediction curve. It is obvious to see that the forecast results of ARIMA model can predict the trend of Shanghai stock index closing price to some extent. There is an upward trend in general, but it climbs slightly from around 3050. During the period between January and March in 2020, the prediction curve fits the true value well, then the true value goes below the prediction line from March to approximately July in 2020. Besides, the curve of true value reaches the lowest point at about 2650. After that, the true value surpasses the prediction line as it soars to around 3440 in July. Next it remains fluctuant till January 2021 with mean about 3300. Next, between January and March, the figure rises dramatically at first and then suffers a significant drop. Eventually it remains stable and fluctuant at approximately 3400. However, the curve has been higher than the prediction line since July 2020, therefore, according to the result of prediction, the Shanghai index is expected to drop.
In order to think throughly, it is necessary to compare different categories of training set and prediction interval. The following figure shows the prediction from 2018 to 2021.
In the second forecast map, we obtain a similar prediction curve with similar trend (slight increase). However, the mean value of the latter is around 250 higher than that of the former. In addition, the final value of prediction fits the true value well since both of them are about 3400.

To sum up, the second means of prediction is better than the first one, but they both produce the same conclusion that in the short term the Shanghai index is expected to fall with greater probability.

At last, to predict recent trend in short term, the third figure is provided.

![Shanghai Index Forecast Map](3)

In the third map, the prediction interval is between 15 April and 30 April 2021. Here we obtain the prediction curve has the same tendency as the true value. This model fits quite well. Therefore, we are able to forecast the rise or fall of the Shanghai index in the coming days according to the ARIMA(11,1,3) model.

5. Conclusions

This paper analyzes the time series of the obtained Shanghai stock index based on the Python language, and fits its historical stock data, and finally establishes a ARIMA model that can not only predict the future value, but it can also forecast the rise or fall in the coming days in the short term. In the process of prediction, the training set of different time periods is modeled and predicted, which is divided into prediction of medium-term sequences, a long-term sequence prediction and a short-term sequence prediction. The results show that the final value of long-term prediction and short-term prediction can better fit the latest test set and the application effect is excellent. As a result, the ARIMA model can predict the stock closing price as well as its ups and downs in the future, and to some extent can provide some suggestions for enterprises, investors or market regulators.
References


