Development trend analysis and prediction of photovoltaic building integration plate index based on ARIMA model

Zhuoran Wang, Jiayuan Su, Diancong Wu

Maynooth International Engineering College, Fuzhou University, Fuzhou, Fujian, 350108, China

Abstract: The purpose of this paper is to analyze and forecast the future development of the whole industry and the trend of the index by studying the 37 stocks in Shanghai stock index and the photovoltaic construction integration plate index composed of the 37 stocks. Firstly, the traditional ARIMA model is used to predict the period from May 6 to May 28, 2021, and the predicted value is compared with the actual value. The error is analyzed by summing the absolute value of the subtraction between the predicted value and the actual value, and then the model is modified to obtain the step-by-step optimized ARIMA model.

Keywords: index, moving average, ARIMA model, time series analysis

1. Introduction

Photovoltaic building integration (BIPV) is a technology that uses photovoltaic materials to replace traditional building materials and integrates these photovoltaic materials and products into buildings. It plays an important role in realizing green environmental protection and low-carbon environment in China [1]. The stock market is a place where individuals and institutional investors buy and sell stocks together in public places. In the stock market, the daily K-line and moving average (MA) of each stock show the basic trend of the stock. The stock market plate index is a set of stock price changes within the plate, it reflects the overall trend of the plate.

Based on the stock data of 37 photovoltaic construction integration related enterprises in Shanghai and Shenzhen stock markets, this paper forecasts the change trend of BIPV, an emerging industry sector index. According to the data from May 6 to May 28, 2021, the ARIMA model is established to forecast the daily moving average of 20 trading days, the weekly moving average of 3 weeks and the monthly moving average of 2 months after May 28.

2. Model Establishment and Solution

2.1 Model preparation

Box Jenkins method is a relatively perfect and accurate algorithm for time series data analysis and prediction. Its common models include: autoregressive model (AR model), moving average model (MA model), (autoregressive moving average hybrid model) ARMA model [2], and (differential integrated moving average autoregressive model) ARIMA model [3]. ARIMA (P, D, q) model is an extension of ARMA (P, q) model. ARIMA (P, D, q) model can be expressed as:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t$$

Description of error: This paper describes the error by subtracting the predicted data from the actual data and adding the absolute value. Subtraction is obvious because we want to describe the difference between the two data. And for a stock, if the i day according to the model, our forecast data is an upward trend, and the actual data is a downward trend, that is to consider the i day error e_i is a positive value, while considering the situation opposite to the above situation on the j-th day, that is, the j-th error e_j is negative. Obviously, we can't simply let $e_i + e_j$. If so, it can not reflect the real data deviation after

summation, so the method of summation after absolute value is adopted. The specific formula is as follows:

$$e_{sum} = \sum |e_i|$$

2.2 Establishment of ARIMA Prediction Model

First, we select 30 days of data as the training set, and 17 days of data (that is, all the trading day data from May 6, 2021 to May 28, 2021) as the verification set.

After that, AIC and BIC criteria are used to determine the order, and ACF and PACF diagrams are drawn for reference, but the latter two diagrams are not used as the basis for determining the final parameters.
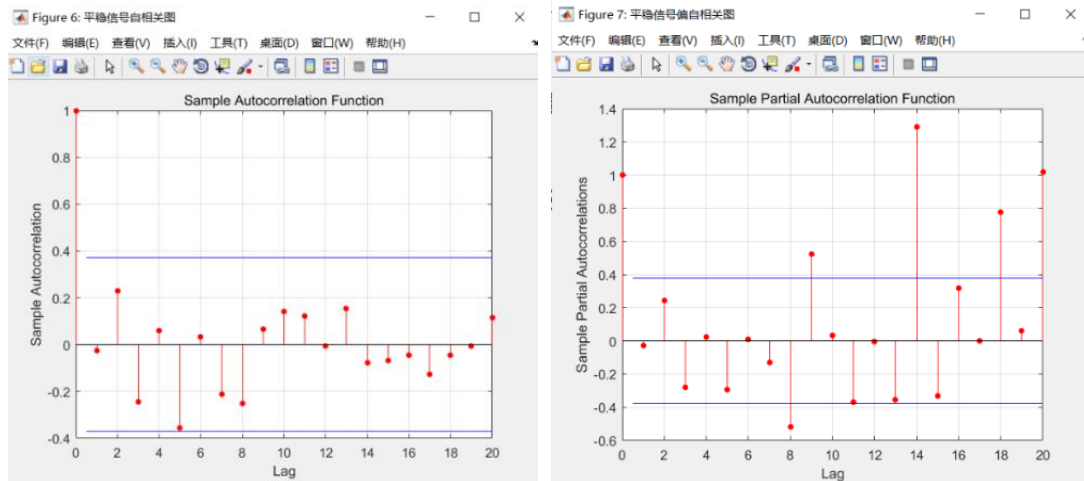

Figure 1: ACF and PACF

As can be seen from the figure above, if ACF and PACF are used as the basis for model selection, the model order is too high and will become too complex, which is not conducive to the subsequent solution, and there is a high possibility of fitting and oscillation. Therefore, AIC and BIC are adopted as the basis for model selection. For convenience, we set the maximum AR order as 3 and the maximum MA order as 3.

Because only stationary sequence can be used for ARIMA modeling, it is found that only the data after first-order difference becomes stationary sequence, so the value of d in ARIMA (P, d, q) is 1.
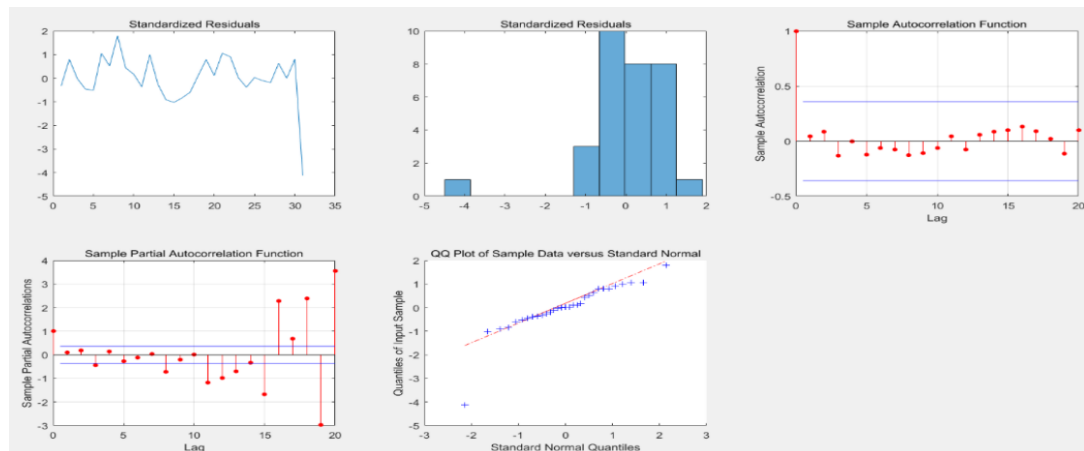
Next, the residual test is performed.

Figure 2: Residual analysis

By matlab calculation, P value is 1, Q value is 0. So we build ARIMA (1,1,0) model. The model is used to forecast the data from May 6, 2021 to May 28, 2021, and is plotted in the same chart with the actual data.
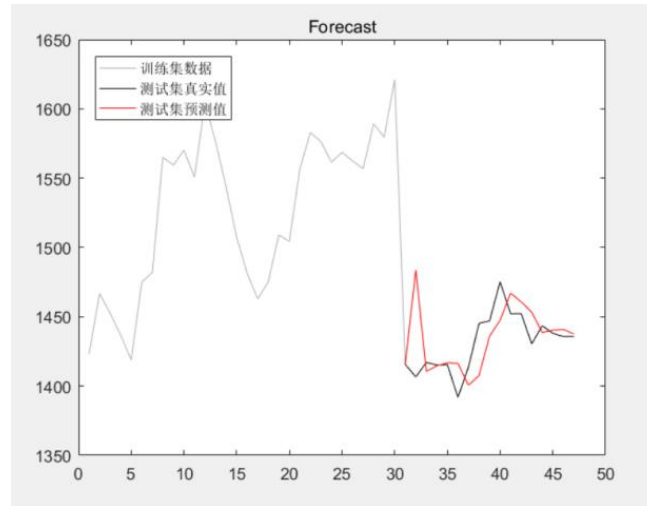
Figure 3: Forecast chart

Intuitively, it is found that the difference between the predicted data and the actual data is small, that is, the model has a good fitting effect.
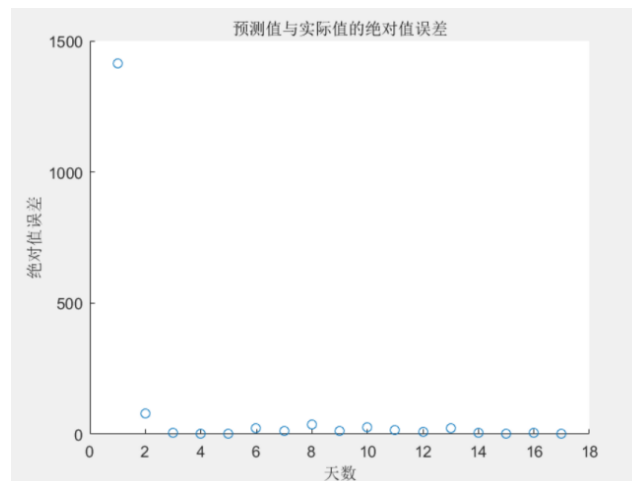

Figure 4: Absolute error between predicted value and actual value

It is observed that the error of the first point is large. Considering the lag effect of ARIMA model, We don't take it into account when we calculate e_{sum}. Using the above formula, the model error is calculated $e_{sum} = 259.8341$.

2.3 Error Analysis and Correction

In data science, increasing the order of model, increasing the amount of data and adding regularization term are often considered in error analysis. Through experiments, we find that higher order can not make the model more accurate, so we consider increasing the amount of data. Here, we use 125 groups of data as the total amount of data, in which the last 17 groups of data are validation sets, and the other data are training sets.

At the same time, in order to make the order and parameters of the model better fit and predict the data, after each new prediction value is calculated, the ARIMA model selection algorithm is called again to determine the new model order, and the new model order (that is, the new model) is used to fit the data and determine the parameters. Finally, the prediction data is added to the data set. The next prediction data will be obtained from the original training data and prediction data.

In this way, we establish a step-by-step optimization ARIMA model.

Recalculate e_{sum} , get $e_{sum} = 206.3264$. It shows that the accuracy of the model is significantly improved without sacrificing the simplicity.

2.4 Forecast

According to the distribution optimization ARIMA model, the daily moving average of 20 days after May 28, 2021, the weekly moving average of 21 days and the monthly moving average of 2 months are calculated.
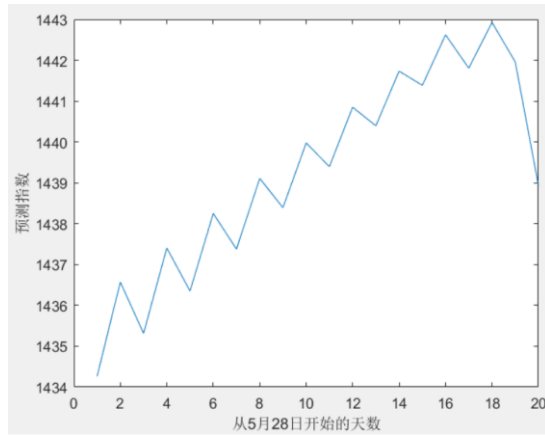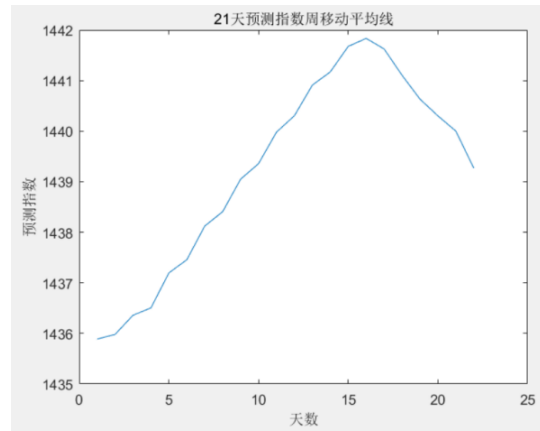

Figure 5: 20 daily moving average
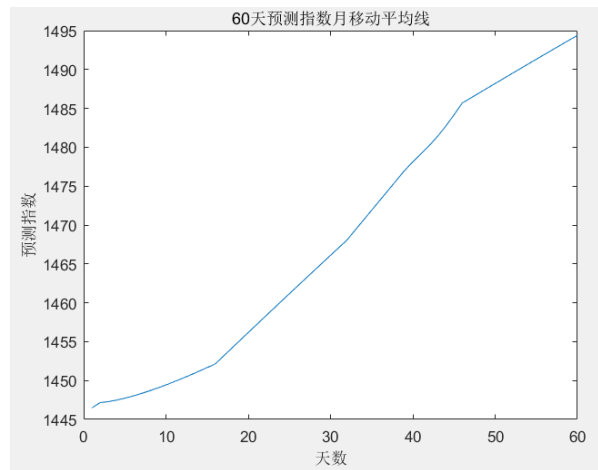

Figure 6: Weekly moving average

Figure 7: Monthly moving average

3. Advantages and Generalization of the Model

Different from the traditional ARIMA model of time series, the parameters and order of this model are dynamic. In the case that the parameters and order may change with the addition of new data (prediction data), this model is more accurate than the traditional ARIMA model with fixed parameters and order, and the generated prediction data is more stable, which can reflect the overall trend of the data rather than local noise.

ARIMA model has good adaptability to all kinds of time series. The model established in this paper can not only forecast index and stock data in econometric finance or financial engineering level, but also fit and forecast GDP in a few years, temperature change in a local area, talent demand in a certain area The fitting prediction of sales volume of a product in a period of time has good universality and popularization.

The model has strong nonlinear mapping ability, can work normally even in the case of local damage, and has a certain fault tolerance.

References

[1] Baidu Encyclopedia, photovoltaic building integration, https://baike.baidu.com/item,2021.05.29
[2] Xia Li, research on regional power consumption prediction method based on ARIMA model and

regression analysis [D], Nanjing University of technology, 2013

[3] Hou Lu, short term analysis and prediction of oil price based on ARIMA model [D], Jinan University, 2009