Application of ARIMA model on prediction of China’s corn market

Liwen Zhou
Fuzhou University of International Studies and Trade, Fuzhou, China

Abstract. ARIMA is a common model for non-stationary time series modeling, the model is essentially the combination of ARMA model and differentiate operation. This study indicated that most of the non-stationary time series can be differentially stationary by the difference of the appropriate order, and can be fitted by the ARMA model. As a relatively mature statistical method, ARIMA has been widely used in the price prediction of energy, agricultural products and stocks due to its advantages such as solid theoretical foundation and good short-term forecasting effect. In this paper, ARIMA was used to take the monthly corn price data of 23 months from April 2019 to February 2021 as the sample, processed and analyzed the data, and made a forecast analysis of the corn price in March 2021 in China. By comparing the experimental results with the actual price, it can be seen that the model established has a good fitting degree and can predict the corn price in China more accurately.

1. Introduction
In the 21st century, food security, energy security and financial security are also called the three major economic security in the world, among which food security is the weakest link [1]. Corn has a significant role in the world food market because of its diversified use. In China, corn is mainly used as feed except for a small part of it. In 2017, the proportion of corn feed consumption in China was 58.6%. Meanwhile, the industrial consumption of corn increased rapidly. In 2017, the industrial consumption of corn increased by 492.6% compared with 2002. Therefore, the price of corn is closely related to the price stability of necessities such as meat, egg and milk, and its price fluctuation also has a crucial impact on the price level in China [2]. Because the fluctuation of corn price is obvious and disorderly, the prediction research on corn price will provide support for the monitoring and warning work of the whole corn market [1].

In recent years, China has conducted in-depth research on the price forecast of agricultural products. There are several mature models that are widely used in China: arch model is established to analyze the futures price of farm products and find the law of its fluctuation, and on this basis, the factors that affect the price fluctuation are analyzed; X-12-ARIMA seasonal adjustment model is used to conduct empirical analysis on vegetable prices; APSO - AVR model, EMD - AVR model and Grey Prediction GM (1,1) model are used to predict the price of farm products in a short time [3].

Although the above method is feasible, there are still some defects, such as modeling requires a large number of samples, easy to fall into over learning, support vector machine is different from the current traditional statistical methods and so on. ARIMA model has the characteristics of simple prediction method and good prediction effect. It can simplify the regression problem to a great extent,
and does not depend too much on the number of samples. Therefore, based on previous research results, this paper constructs a single prediction model based on ARIMA model to study the change trend of corn price, aiming to provide more reference for corn price prediction methods [3].

2. Principle and algorithm

2.1. Time series with ARIMA model
Time series analysis and prediction opinion has developed into a mature price analysis and forecasting method, among which ARIMA model is one of the common methods for time series analysis and prediction. ARIMA model is a famous time series prediction method proposed by Box and Jenkins in the early 1970s, and it is a common model for nonstationary time series modeling. ARIMA (p, d, q) is called summation autoregressive moving average model, where AR is autoregressive term, p is autoregressive order, MA is moving average term, q is moving average order, d means the number of difference that nonstationary time series will pass through to become stationary sequence [5].

2.2. Establish model
The modeling process of ARIMA model is shown in the Figure1, and the specific steps are as follows:

(1) Test the stationarity of the sequence. According to the sequence diagram, autocorrelation diagram and partial autocorrelation diagram, it can directly judge whether the sequence is stable or not, or use ADF unit root test method to judge whether the sequence is stable from the perspective of statistical test, and use difference or logarithm method to make the non-stationary sequence stable [3].

(2) Model based recognition rules to establish a preliminary model for the time series after stabilization. After the stabilization, the AR model can be established by the judgment of the tail of partial correlation function and autocorrelation function; if the tail of partial correlation function and autocorrelation function are truncated, MA model can be established; when the partial correlation function and autocorrelation function are tail, ARIMA model can be established.

(3) Estimate the parameters of the model, determine the order of the model, and test the feasibility of the specific model. Only feasible models have statistical significance.

(4) The hypothesis test is carried out to determine whether the residual sequence is white noise sequence.

(5) After testing, the specific modeling is carried out with the determined parameters, and the model is used to predict and analyze the time series.
Figure 1. Flow chart of model establishment.

ARIMA (p, d, q) model of the structure for the [6]:

\[
\begin{align*}
\Phi(B)^d \nabla^d x_t &= \Theta(B) \epsilon_t, \\
E(\epsilon_t) &= 0, \text{Var}(\epsilon_t) = \sigma^2, E(\epsilon_t \epsilon_s) = 0, s \neq t, \\
E(X_t \epsilon_s) &= 0, \forall s < t.
\end{align*}
\] (1)

Type: \( \nabla d = (1 - B)^d \), B for the lag operator. \( \Phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \) to ARIMA (p, d, q) model of regression coefficient polynomials, \( \Theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q \) is ARIMA (p, d, q) model of the moving average coefficients polynomial, \( t \) represents the time, \( \epsilon_t \) for white noise sequence[4]. After difference operation nonstationary sequence shows the nature of the stationary series, then we call the nonstationary sequence difference stationary series, which can be used to sequence the ARIMA model for fitting [7].

3. Data collection and analysis

3.1. Variable selection and data sources
Corn price data from April 2019 to February 2021 are obtained from China corn network. After selection and collation, 23 monthly data from April 2019 to February 2021 are finally determined as the main object of corn future price forecast. As shown in Figure 2, it shows that in recent years, the
domestic corn price fluctuation is relatively fierce, and there is no obvious law from the overall fluctuation trend.

Figure 2. China's corn price trend from April 2019 to February 2021.

3.2. Data preprocessing

3.2.1. Stationarity test. Before analyzing the data, it needs to test the stationarity of the series. When the time series is a stationary series, the mean, variance and covariance of the series should maintain a certain degree of stability in a certain period of time. In this time range, the series should conform to the probability distribution on the relative degree. For non-stationary time series, if the time series is non-stationary, the values of mean, variance and covariance will change with time. As a result, it is difficult to analyze the internal relationship of non-stationary series through econometric model, so it is difficult to predict the future series. In the process of analysis in reality, most of the price series are in a non-stationary state. Therefore, if the sequence studied is a non-stationary sequence, it should be transformed into a stationary sequence by difference or logarithm.

In this paper, the data $Y$ is checked by ADF test method. The test results are shown in Table 1:

| Augmented Dickey-Fuiler test statistic | Statistic   | Prob* |
|---------------------------------------|-------------|-------|
|                                       | -1.874633   | 0.3726|

According to the ADF test results, the ADF test statistic of variable $Y$ is -1.873423, and the corresponding $P$ value is 0.3726 > 0.05. Therefore, the original hypothesis is accepted at the significance level of 5%, that is, the data $Y$ is a non-stationary series. The variable $DY$ is obtained by first-order difference of variable $y$, and the stationarity test of $DY$ is conducted again. The results are shown in Table 2.
Table 2. Test the stationarity of data DY.

| Augmented Dickey-Fuller test statistic | t-Statistic | Prob* |
|---------------------------------------|-------------|-------|
| Test critical values                  |             |       |
| 1% level                              | -3.482628   |       |
| 5% level                              | -2.846736   |       |
| 10% level                             | -2.583362   |       |

According to the ADF test results, the ADF test statistics of variable DY are -9.182684, and the corresponding $P$ value is 0.0000, which is far less than 0.05. Therefore, the original hypothesis is rejected at the level of 5% significance, that is, data DY is a stationary sequence, which meets the condition of establishing ARIMA model.

3.2.2. White noise test. Not all stationary sequences are worth modeling. Only when the sequence values of the sequences are interdependent and the historical data have a certain impact on the future development, it is worth modeling. The sequence values of white noise sequences have no correlation with each other. From the perspective of statistical analysis, there is no value of analysis and modeling. Therefore, white noise test is needed before modeling stationary series. In practice, it often judge the randomness of the sequence by the Q_LB statistics of delaying 6 and 12 periods. The results show that the $P$ values of Q_LB statistics of the new sequence are 0.004783 and 0.04736 respectively. Under the condition of significance less than 0.05, the original hypothesis is rejected and the new sequence belongs to nonwhite noise, so the analysis can be continued.

3.3. Model identification and order
According to the above test results, the time series is not only stable, but also contains relevant information which is worth extracting, so it can be used for model identification and order determination.

ARMA model establishment this section establishes ARIMA model for y data. ARIMA (p, d, q) model determines the order by AC and PAC minimum information criterion. In this paper, common values of p and q are compared to determine the final lag order. Therefore, ARIMA (3,1,3) is the final model in this paper. Using this model for regression simulation, the test results show that all regression coefficients are significant at the 5% significance level, the fitting coefficient of the model is 0.1004%, the F statistic of the model is 2.84, and the corresponding $p$ value is 0.05, which indicates that the whole model is significant at the 5% significance level.

3.4. Residual test
The original hypothesis assumes that there is no correlation between the series of m periods, so it is impossible to get accurate and reasonable connection from the price data. The alternative hypothesis assumes that there is correlation between the price series, and the series are interrelated. The autocorrelation partial correlation test is carried out on the residual of the model, and the test results are used to verify the price time series, so as to determine whether the price time series has correlation.

Model residual test is used to test the autocorrelation and partial autocorrelation of the residuals of ARIMA (3,1,3) model. The test results show that the autocorrelation and partial autocorrelation values of the model residuals are all smaller than the critical value, and the $p$ value corresponding to $q$ statistic is larger than 0.05, which indicates that there is no autocorrelation in the residuals of ARIMA model at the 5% significance level, which indicates that the model can better describe the characteristics of the data model.
3.5. Model prediction results

After the series passed the test, the prediction was started according to ARIMA model. According to the above analysis, ARIMA (3,1,3) is used to predict the future price based on China's corn price data from April 2019 to February 2021, thus confirming the ARIMA (3,1,3) model. It has good credibility. According to the forecast, the real price is in the confidence interval of the forecast. This result proves that the established model has a good fitting degree, and can accurately predict the corn price in Hebei Province. According to the relevant values of the above results, the prediction results of the model can be scientifically and reasonably expressed from the perspective of econometrics, in which the Theil inequality coefficient is 0.1015475, the value is obviously < 0.05, and the partial correlation coefficient is 0.000004, which means that the prediction has a very high degree of fit, and the variance ratio is 0.027354. The above three can effectively prove that the results have high accuracy. The value of covariance ratio in the correlation value is 0.974765, which is very close to 1. This result strongly proves that the model has high credibility. But when the value of covariance ratio is very high, it means that the deviation is very small, which verifies that the prediction results are very similar to the real price data.

4. Conclusions

In this paper, ARIMA model of time series analysis was used to predict the price change of China's corn market, and ARIMA (2,1,2) model was established. The model was used to analyze China's corn market price from April 2019 to February 2021, and the change of China's corn market price can be predicted. Because the price series will change with the change of many factors in the actual analysis, this model can well predict the change of corn price in Hebei Province, and all the prices in the series were generally random. When using the real price data in the past to forecast, there will inevitably be some errors. Therefore, it was more suitable for short-term forecasting, and the established model should be supplemented and modified by constantly updating the data to make the forecasting more accurate.

References

[1] Zhang, wang chuan, chun-ying Yang, king just. Corn prices prediction research based on time series SVR model [J]. Journal of Chinese farmers, 2020, 4 (1): 115-120.
[2] Zheng Xuyun, li-chuan chuang, Qiu Zehui. China's corn prices fluctuate distribution and influencing factors of the empirical [J]. Journal of statistics and decision, 36 (2): 52-56.
[3] Populus euphratica, Charles. Based on the ARIMA model to predict the price of corn in Hebei province [J]. Journal of agriculture and technology, 2020, 40 (23): 149-152.
[4] Wen-ling huang, Zheng Xiaoying, Breda McCarthy, da-bin zhang. ARIMA model based on the prices of live pigs in guangdong's short-term prediction [J]. Chinese journal of animal husbandry, 2018, (12): 119-123.
[5] Hu-sheng lu, liu wei, chun-li li. Based on the ARIMA model of baiyun obo ore price prediction [J]. Journal of rare earth oxides of rare earth, 2019, 40 (2): 148-158.
[6] What book yuan. Application of time series analysis [M]. Beijing: Peking University press, 2014.
[7] Zhang Yifan Fan Meihua. ARIMA model based on the price of eggs of the national market analysis and forecast [J]. Chinese poultry, and 2020 (01): 82-86.