ISSN 2616-5902 Vol. 4, Issue 11: 7-12, DOI: 10.25236/AJBM.2022.041102

Corn Price Prediction in China's Futures Market during COVID-19

Junxue Lv[1], Xi Wu[2,*]

[1]College of Science, Anhui Agricultural University, Hefei, 230036, China
[2]School of Business Administration, Northeastern University, Shenyang, 110169, China
**Corresponding author: wuxi@mail.neu.edu.cn*

Abstract: *Using data from the Dalian Commodity Exchange from January 2010 to December 2019 as a training set, this study develops an optimal seasonal autoregressive integrated moving average model (SARIMA) in Python to predict the settlement price of corn futures. Further, the model's forecast accuracy and applicability are tested by predicting the price of corn futures from January 2020 to December 2020 and comparing it with the actual settlement price of active corn futures contracts in 2020 after the outbreak of COVID-19. The results show that the SARIMA (2,1,0) $(3,1,1)_{12}$ model can accurately predict the settlement price. Moreover, COVID-19 had a positive short-term impact on the settlement prices.*

Keywords: *Corn Futures, SARIMA Time Series, COVID-19, Price Prediction*

1. Introduction

China is one of the main producers and consumers of corn worldwide, and is a crucial player in international corn trade. Corn futures play an important role in the development and structural adjustments of the corn industry. China's corn futures market has developed in three stages. The first stage is from 2004 to 2008, when China's corn futures market was established. In this period, the trading volume of corn futures on the Dalian Commodity Exchange (DCE) trended up. The second stage was from 2008 to 2015. Influenced by the government's policy of temporary purchase and storage in 2007, and the global financial crisis in 2008, the trading volume of corn futures decreased sharply and hovered at a low level. Finally, the third stage was the period after 2015, where marketization played an important role of intensifying market development and the volume of corn futures again trended up. Still, compared to the development of foreign corn futures markets, China's corn futures market has not been able to fully provide a hedging function. Moreover, the COVID-19 outbreak in early 2020 led to large fluctuations in financial markets and commodity prices. For example, it may have had a negative impact on corn supply, but a positive impact on corn demand. Did COVID-19 actually affect the fluctuation of corn futures prices in China? This study seeks to answer this question by exploring the impact of the emergence of COVID-19 on the volatility of corn futures prices in China. We use time-series panel data to carry out regression analysis on the changes in corn futures price volatility during COVID-19 and then analyze how the futures market can quickly return to the right track in the post-epidemic era.

Studies have used various methods to study corn futures prices. Hua (2021) used the Durbin spatial regression model to empirically analyze the trading price of the corn futures market from three aspects: internal, external, and international factors. The author concluded that corn supply and demand were significantly correlated with price fluctuations in the futures market, and other control variables also had a positive impact. Fang et al. (2020) adopted the Ensemble Empirical Mode Decomposition (EEMD) technology to decompose six different types of agricultural futures prices. Then, the authors used a support vector machine (SVM), neural network (NN), and autoregressive integrated moving average (ARIMA) to predict the decomposition components. Hu (2019) investigated which factors influenced prices from policy, and supply and demand perspectives. The author constructed a corn price model through factor screening and used it to make short-term predictions. Zhang (2018) predicted corn futures prices using SVM with a prediction accuracy rate that reached 65.83%. Wang (2017) used R to construct a time-series and NN model for data mining and use this model to predict corn prices. Wang et al. (2014) examined which decisive factors affected China's corn prices by analyzing the trend and changes in corn futures prices from 2005, when they were listed, to 2014. Monk et al. (2010) used the autoregressive conditional heteroscedasticity (ARCH)/ generalized ARCH (GARCH) methods to quantify the price

ISSN 2616-5902 Vol. 4, Issue 11: 7-12, DOI: 10.25236/AJBM.2022.041102

volatility of white corn futures contracts traded on the South African Futures Exchange (SAFEX); their results highlighted the usefulness of put options as a price risk management tool.

The seasonal ARIMA (SARIMA) model, based on the Python language, is another useful for forecast analysis. In a SARIMA model represented by SARIMA(p,d,q)(P,D,Q)$_n$, d and D represent the number of non-seasonal and seasonal differences, respectively, and p, P, q, and Q represent the maximum lag order of non-seasonal, seasonal, autoregressive, and moving average operators, respectively. Using this model, a non-stationary time series with obvious seasonality can be converted into stationary time series after finite seasonal and non-seasonal differences. Here, we used the SARIMA model to analyze the prices in China's corn futures market. The settlement price of active corn futures on the DCE from January 2010 to December 2017 was taken as the training set. Then, the effect of SARIMA model was evaluated using data from January 2018 to December 2019. Finally, to study the impact of corn futures market price fluctuations before and after COVID-19, the settlement price from January 2020 to December 2020 was predicted and compared with the actual settlement price in 2020 after the COVID-19 outbreak.

2. Materials and methods

We build a SARIMA model for the settlement price of corn futures based using Python to predict short-term price fluctuations in the corn futures market.

2.1. Data Selection

The futures settlement price data of yellow corn active contracts from December 2009 to December 2020 were extracted from the WIND database. To maintain the continuity of price data, we chose the futures trading data of active contracts with large trading volumes and strong representativeness, which can objectively reflect the price of the corn futures market.

Before model prediction, data cleansing and format conversion were undertaken in Python. The dataset was imported into Python JupyterLab, and simply screened and filtered. We checked the attributes of the data and missing values for subsequent visualization operations in order to determine the type of the established time-series model.

2.2. Model Specification

First, we tested the stationarity of the time series by focusing on the mean value and standard deviation. We also used the Augmented Dickey Fuller (ADF) test whether the time series was stable. If the original time series is not stable, then the series should be stabilized. Thus, the second step was to use the difference method (DM) to stabilize the sequence. Then, a white noise test was performed to observe the p-value of the Q statistic of the test results. If the test statistic is less than the critical value, the post-stabilization sequence is considered a non-white noise sequence at the given significance level modeling analysis can be carried out in the next step. d and D were determined according to the difference in times, and p, q, P, and Q were determined according to autocorrelation function (ACF) and partial ACF (PACF); then, we chose the optimal model. Parameter estimation and residual analysis were carried out to observe the Ljung-Box statistic P-value of the residuals of the model. If the residual is a white noise sequence, the model fits well. Finally, the Mean-square Error (MSE) and Root MSE (RMSE) were used to measure the forecast accuracy and goodness of fit of a model. The smaller the MSE and RMSE, the better the prediction accuracy and goodness of fit a model.

2.3. Parameter Estimation and Test

The time series diagram of the original data shows that the settlement price of corn futures has obvious seasonal and long-term trends, and that it is a non-stationary time series. However, the residual of the time series is stable after extracting the trend and seasonal effects. The ADF test result (-1.545) also shows that the original time series is non-stationary, as shown in Figure 1.

One of the most common methods for eliminating trends and seasonal effects in a time series is the difference method. Here, the trends and seasonal effects of the original time series were extracted using the first-order 12-step difference method. After using the difference method, the moving average fluctuates around zero and the standard deviation is stable; this indicates a stationary time series. Further, the ADF test value (-3.929) of this transformed time series also shows its stationarity. In addition, the white noise test shows that after the first-order twelve-step difference, the Q statistic p-values were all less than 0.01. Therefore, the sequence is a stationary non-white noise sequence and the model can be established as $\mathrm{SARIMA}(p,1,q)(P,1,Q)_{12}$.

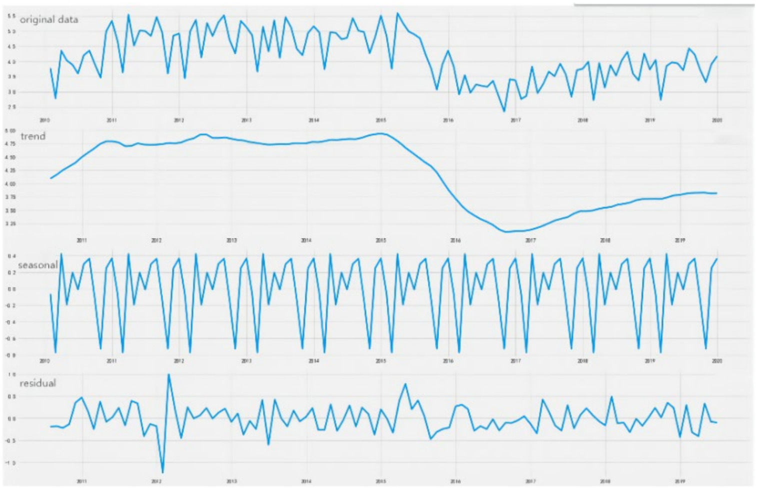

Figure 1: Trend, seasonality, and randomness effects of the initial time series

2.4. SARIMA Model Construction

The optimal parameters of the model can be found in the autocorrelation and partial autocorrelation graphs of the time series after the first difference. Fig. 2 demonstrates that the autocorrelation coefficients within order 12 are truncated and the partial autocorrelation coefficients are trailing, suggesting that the seasonal model uses the SARIMA (1,1,1) $(3,1,1)_{12}$ model to extract the short-term autocorrelation information of the sequence after difference.

We use the grid search method to systematically select the optimal parameter value. For each group of parameters, a new seasonal ARIMA model can be fitted using the SARIMAX () function in the Statsmodels module in Python, and its overall quality can be assessed using the Akaike Information Criterion (AIC)/ Bayesian Information Criterion (BIC) values. Specifically, when the grid search traverses the entire parameter environment, the parameters with the best performance are selected from the parameter set according to the criterion of evaluating the time-series model.

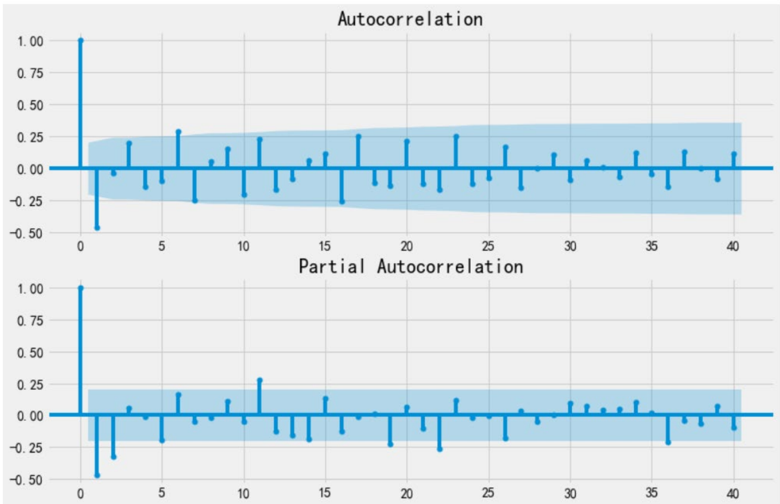

Figure 2: ACF and PACF diagram of the time series after 1-order 12-step difference

According to the grid search and AIC criterion, the AIC value of SARIMA (0,1,1) $(0,1,1)_{12}$ was the lowest at 97.576. Therefore, we believe that this is one of the best choices for all parameter combinations. Meanwhile, two reference models – SARIMA (2,1,3) $(3,1,1)_{12}$ and SARIMA (2,1,0) $(3,1,1)_{12}$ – were obtained by using separate AIC/BIC criterion. The best minimum AIC/BIC model was SARIMA (2,1,0) $(3,1,1)_{12}$.

ISSN 2616-5902 Vol. 4, Issue 11: 7-12, DOI: 10.25236/AJBM.2022.041102

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                         No. Observations:                  120
Model:             SARIMAX(2, 1, 0)x(3, 1, [1], 12)   Log Likelihood                 -25.364
Date:                              Tue, 20 Apr 2021   AIC                             64.728
Time:                                      21:51:24   BIC                             80.366
Sample:                                  01-31-2010   HQIC                            70.932
                                       - 12-31-2019
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6111      0.120     -5.108      0.000      -0.846      -0.377
ar.L2         -0.2174      0.126     -1.727      0.084      -0.464       0.029
ar.S.L12      -0.1340      0.167     -0.804      0.421      -0.461       0.193
ar.S.L24      -0.2961      0.118     -2.503      0.012      -0.528      -0.064
ar.S.L36      -0.2775      0.123     -2.247      0.025      -0.519      -0.035
ma.S.L12      -0.6925      0.312     -2.217      0.027      -1.305      -0.080
sigma2         0.1104      0.022      4.942      0.000       0.067       0.154
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                 0.88
Prob(Q):                              0.89   Prob(JB):                         0.64
Heteroskedasticity (H):               0.50   Skew:                            -0.25
Prob(H) (two-sided):                  0.11   Kurtosis:                         3.25
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Figure 3: Modeling results of SARIMA (2,1,0) (3,1,1)12 model

The modeling results of the SARIMA (2,1,0) $(3,1,1)_{12}$ model in Fig. 3 show that the residual Ljung-Box statistic P value of this model is 0.89, indicating that the residual is a white noise sequence and that the model fits well.

2.5. Model Fitting Test

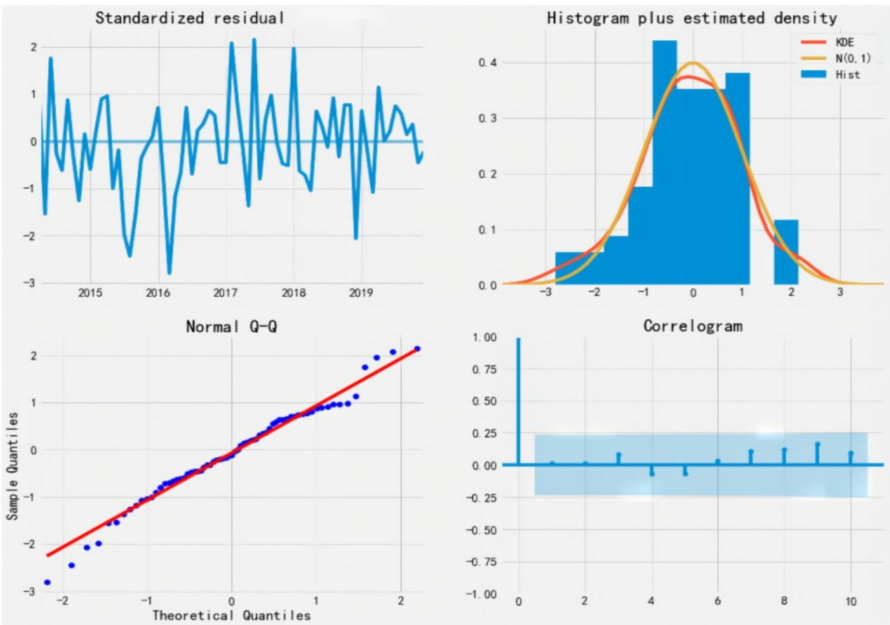

Figure 4: SARIMA model fitting test results

Model fit testing ensures that the assumptions made by the model are not violated. If the residuals in the SARIMA model are correlated and do not follow a zero-mean normal distribution, this indicates that the model can be further improved. In contrast, when the model fits well, this indicates that the model has fully extracted the information in the sequence. We use the Ljung-Box test to test the residual white noise. The results of the SARIMA fitting model show that when the residual sequence delay is 1-12, the P-values of the Q statistics are all greater than 0.01. Therefore, the residual is a white noise sequence, indicating that the information in the time series has been completely extracted by the fitting model.

Regarding the residual distribution problem, the plot_diagnostic function in Python can quickly generate a model-fitting diagnosis and investigate the model’s abnormal behavior. Fig. 4 demonstrates that the timing sequence diagram of the residual is stable. As indicated by the red kernel density estimate

ISSN 2616-5902 Vol. 4, Issue 11: 7-12, DOI: 10.25236/AJBM.2022.041102

(KDE) line adjacent to the N (0,1) line with a mean of zero and standard deviation of 1, the timing sequence is normally distributed. Furthermore, the autocorrelation diagram of the residual indicates a white noise sequence. Therefore, the SARIMA (2,1,0) $(3,1,1)_{12}$ model has a good fit, and can help us in analyzing and predicting the original time series data of the future settlement price more accurately.

3. Forecast and results

Before using the SARIMA (2,1,0) $(3,1,1)_{12}$ model to predict the corn futures prices, we first verify the model's accuracy with static and dynamic predictions.

3.1. Static Prediction

The original data from 2010 to 2017 were used as the test set, and the fitted model was used to carry out static predictions from 2018 to 2019. We then drew a time series diagram of the real and static predicted values from 2018 to 2019. Generally, the predicted value of the static prediction fits well with the actual value, as shown in Figure 5.

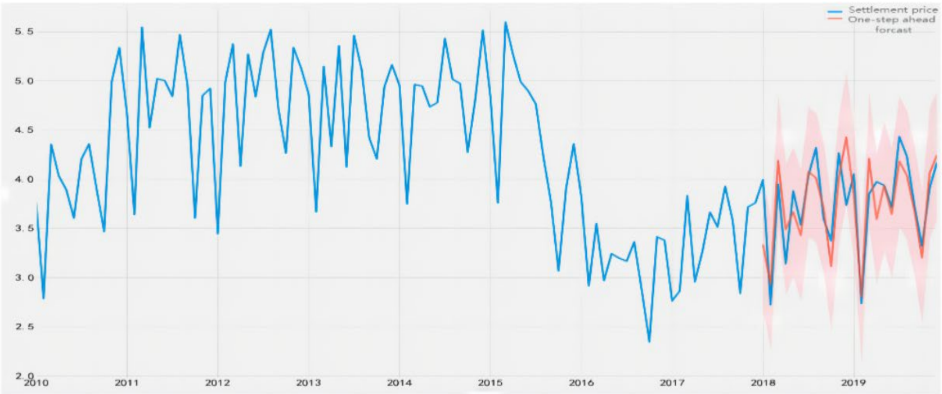

Figure 5: Static model prediction fitting time series diagram

Furthermore, MSE and RMSE were used to quantify the accuracy of the model's static prediction. The MSE and RMSE were 0.08 and 0.282, respectively, indicating a reliable static prediction effect.

3.2. Dynamic Prediction

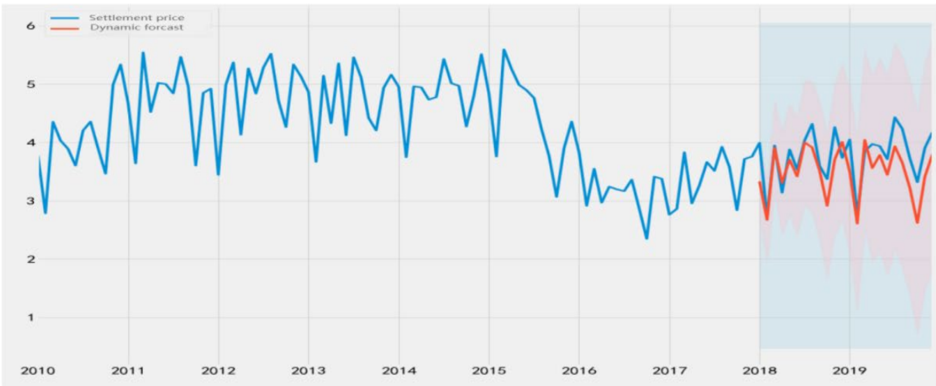

Figure 6: Dynamic model prediction fitting time series diagram

Next, as in the static prediction, the fitted model was used for the dynamic prediction of the data from 2018 to 2019. According to Fig. 6, the dynamic prediction fits the time-series diagram and the overall prediction is accurate. The predicted values (red line) of all dynamic predictions agree well with the original observed values (blue line) and are within the predicted confidence interval.

Meanwhile, the MSE and RMSE are 0.15 and 0.388, respectively, indicating that dynamic prediction is feasible. Notably, the MSE and RMSE of the dynamic prediction are both larger than those of the static prediction because the dynamic predictions are less dependent on historical data in the time series. In conclusion, the SARIMA (2,1,0) $(3,1,1)_{12}$ model has a very good prediction ability and the fitted model can be used to predict future corn futures settlement prices.

3.3. Model Prediction

Finally, we use the SARIMA (2,1,0) $(3,1,1)_{12}$ to predict the settlement price of corn futures in 2020, as illustrated in Fig. 7. The model predicts that corn futures prices (red line) should decrease in 2020. However, the actual corn futures price (yellow line) on the DCE increases in 2020. This shows that COVID-19 has a positive short-term impact on the settlement prices of active corn futures contracts.

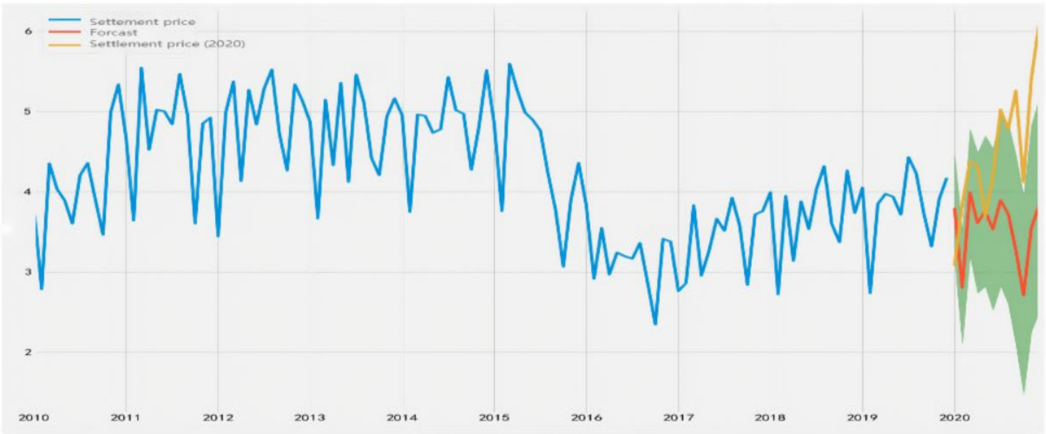

Figure 7: A time series diagram of the actual and predicted values of the 2020

4. Conclusions

This study develops a SARIMA (2,1,0) $(3,1,1)_{12}$ time series model to predict the change in corn futures prices in DCE. The empirical modelling results show that the model can better predict changes in corn futures prices. However, because of various factors, real corn futures prices can be stochastic in reality. Thus, using the real corn futures price from the past for prediction may inevitably lead to errors. Thus, the model may be more suitable for short-term prediction. Moreover, when a major public event (like COVID-19) occurs, a single quantitative model is not sufficient and external factors should be incorporated into the model. Future research can also consider using other ARIMA mixed models combined with various neural network models to improve prediction accuracy.

References

[1] Hua Yutao. Analysis of price fluctuation factors in agricultural futures market -- a case study of corn futures [J]. Commercial Economics Research,2021 (05): 182-184.
[2] Fang Y, Guan B, Wu S. Optimal forecast combination based on ensemble empirical mode decomposition for agricultural commodity futures prices [J]. Journal of Forecasting, 2020, 39 (6).
[3] Hu Zhongxing. Analysis and prediction of influencing factors of corn price fluctuation in Heilongjiang Province [D]. Northeast Agricultural University, 2019, 1-80.
[4] Zhang Shouping. Research on corn futures price Prediction method based on machine learning [D]. Northeast Agricultural University, 2018, 1-83.
[5] Wang Wenjing. Prediction of corn market price based on data mining [D]. Qingdao University, 2017, 1-54.
[6] Wang Yanqing, Wang Xiaoshu, Wu Laping. Analysis on price trend and influencing factors of corn futures market [J]. Agricultural outlook, 2014, 10 (11): 30-35.
[7] Monk, M. J., Jordaan, Henry B. Factors affecting the price volatility of July futures contracts for white maize in South Africa [J]. Agrekon, 2010, 49 (4)