Application of time series model and deep learning method in measuring the impact of COVID-19 on agriculture in Hubei, China

Yonghong Zou¹, Jiaojiao Wang¹

¹School of Statistics, Chengdu University of Information Technology, Chengdu 610103, China

Corresponding author and e-mail: Yonghong Zou, zyonghong@cuit.edu.cn

Abstract. Taking the agricultural situation of Hubei Province as the research object, this paper uses the time series ARIMA model, ARMA model and deep learning LSTM neural network model to explore the impact of COVID-19 on the agriculture of Hubei Province. Three main indicators are screened out to measure the development of agricultural economy, that is, gross regional product, gross output value of agriculture, forestry, animal husbandry and fishery, agricultural product production price index. Based on the quarterly data of indicators from 2001 to 2019 from the National Bureau of Statistics, three indicators in Hubei Province in the first quarter and the second quarter of 2020 are predicted by using the deep learning-based time series ARIMA model, ARMA model and LSTM neural network model. By comparing the predicted data with the real data, the impact of COVID-19 is measured on the agricultural situation of Hubei Province. It was found that COVID-19 had a great impact on the agricultural situation of Hubei Province in both of the first and second quarters of 2020, with the impact in the first quarter being greater than that in the second quarter. At the same time, the prediction accuracy of the two methods is compared to find that the time series model is more effective and reliable in predicting the agricultural product price index. The LSTM neural network model with a long and short term memory has a good prediction effect on the regional gross product and the total output value of agriculture, forestry, animal husbandry and fishery.

1. Introduction
Since December 2019, China has been hit by a major public health emergency that is spread with fastest rate in the widest scope of infection and suffering the greatest difficulty in preventing and controlling it since the founding of the People's Republic of China. Hubei's economy was the first to suffer [1]. It is necessary to study the impact of COVID-19 on agriculture in Hubei Province.

2. Study theory and method

2.1. Basic theories of ARMA and ARIMA models
ARMA model is a hybrid model of autoregressive moving average (ARMA), which is one of the important analysis and prediction methods of time series. Since this model contains two components, autoregressive and moving average, its order is two-dimensional and consists of two numbers, p and q,
where p represents the order of autoregressive component and q represents the order of moving average component, which is denoted as ARMA (p, q).

ARIMA model is the differential integrated moving average autoregressive model. In ARIMA(p, d, q), AR is autoregressive, and p is the number of autoregressive terms. MA is the moving average, Q is the number of moving average terms, and D is the number (order) of differences made to make it a stationary sequence [2-3]. The word "difference" does not appear in the English name of ARIMA, but it is a key step in the method.

The above two methods are all time series prediction methods, which are one of the prediction methods proposed in this paper.

2.2. Basic theory of LSTM model

Long-Short Term Memory (LSTM) is a special RNN model, which is proposed to solve the problem of gradient dispersion of RNN model. In the traditional RNN, the training algorithm uses BPTT. When the time is relatively long, the residual that needs to be returned will decline exponentially, resulting in the slow update of network weight, which cannot reflect the long-term memory effect of RNN. Therefore, a storage unit is needed to store memory, so the LSTM model is proposed. Compared with traditional RNN, LSTM naturally has a good support for long-term dependence [4]. The core ideas of LSTM model mainly include memory cell and nonlinear gating unit, among which memory cell is used to maintain the state of the system. The non-linear gate unit is used to regulate the inflow and outflow of information from the memory tuple at each time point. Each recursive neural network can be decomposed into an infinite number of basic repeating units, as is the case with traditional RNN and LSTM. In the traditional RNN, the internal structure of the basic repeating unit is very simple, usually with only a simple neural network layer (as shown in Figure 1). In LSTM, four neural network layers are used and interact with each other in a special relationship (see Figure 2). This method is another method to be used in this paper.

![Figure 1. RNN unit structure drawing.](image)

![Figure 2. LSTM unit structure drawing.](image)

3. Estimation of the impact of COVID-19 on China's agricultural and rural economy based on time series

3.1. Selection of variables

In this paper, the gross regional product of Hubei Province, the total output value of agriculture, forestry, animal husbandry and fishery in Hubei Province, and the production price index of agricultural products in Hubei Province are selected as three representative indicators to measure the impact of the epidemic on agricultural and rural economy in Hubei Province. Due to the limited data available on the website of the National Bureau of Statistics, the time range of selecting the gross regional product of Hubei Province is from 2005 to the second quarter of 2020, the time range of selecting the gross output value of agriculture, forestry, animal husbandry and fishery of Hubei Province is from 2001 to the second quarter of 2020, and the time range of producing price index of agricultural products in Hubei Province is from 2004 to the second quarter of 2020. The forecast range is for the first and second quarter of 2020.
3.2. The modeling process

3.2.1. Stationary test of the model. First of all, according to the three indexes in the first quarter of 2005 to the fourth quarter of 2019 data of time sequence diagram, the gross area of Hubei province and fishing output value of agriculture and forestry have obvious growth trend, thus can judge both the sequence for non-stationary time series, and Hubei province agricultural producer price index figure has no obvious growth trend but not smooth.

Eviews7.0 statistical software was used to conduct ADF test on these three time series. In the case of intercept and trend selection, ADF values of series {API} were all less than the critical values of 1%, 5% and 10%, respectively, as shown in Table 1. Therefore, both the original time series {GDP} and {NLMY} are judged to be non-stationary time series of random trend, while the original series {API} is non-stationary time series of deterministic trend.

In order to remove the randomness trend of {GDP} and {NLMY} from becoming stable, the first-order difference of these two original sequences is needed to obtain the sequences {GDP1} and {NLMY1}. The ADF test on {GDP1} and {NLMY1} was continued, and the results showed that, as shown in Table 2, {GDP1} and {NLMY1} rejected the null hypothesis and did not have unit roots. Therefore, it can be determined that {NLMY1} and {GDP1} after first-order difference are stationary time series, both of which belong to first-order integral series. Therefore, d=1 in the model. Finally, the correlation test of these two sequences shows that {GDP1} and {NLMY1} are both stationary and non-white noise time series.

In order to remove the deterministic trend of the original sequence {API} and become stable, the trend removal method is needed. First, t is used to test the significance of constant c and time t to generate the residual sequence {API2} of {API}, and then unit root test is conducted on the sequence {API2}. The test results show that, as shown in Table 3, the null hypothesis is rejected and there is no unit root \[5\]. Therefore, it can be judged that the residual sequence {API2} is a stationary time series.

Finally, the correlation test of these three sequences shows that {GDP1}, {NLMY1} and {API2} are all stationary and non-white noise time series. Satisfy the conditions of establishing the ARIMA model.

| Table 1. Time series ADF test value of gross regional product and total output value of agriculture, forestry, animal husbandry and fishery in Hubei Province. |
|---|---|---|---|---|
| variable | ADF value | 1% critical value | 5% critical value | 10% critical value |
| GDP | -0.733 | -4.134 | -3.494 | -3.176 |
| nlmy | -2.558 | -4.091 | -3.473 | -3.164 |
| api | -4.807 | -4.118 | -3.487 | -3.172 |

| Table 2. First order difference ADF test value of gross regional product and total output value of agriculture, forestry, animal husbandry and fishery in Hubei Province. |
|---|---|---|---|---|
| variable | ADF value | 1% critical value | 5% critical value | 10% critical value |
| GDP1 | -4.139 | -3.555 | -2.916 | -2.596 |
| nlmy1 | -2.598 | -2.599 | -1.946 | -1.614 |

| Table 3. Residual series ADF test value of time series of agricultural product price index in Hubei province. |
|---|---|---|---|---|
| variable | ADF value | 1% critical value | 5% critical value | 10% critical value |
| api2 | -4.923 | -2.604 | -1.946 | -1.613 |
3.2.2. Model identification and hierarchy. The model was identified and graded according to the obtained time series samples, and the autocorrelation (AC) and partial autocorrelation (PAC) diagrams of time series \{GDP1\}, \{NLMY1\} and \{API2\} were drawn, as shown in Figure 3.

![Figure 3](image)

Figure 3. Autocorrelation and partial autocorrelation graphs of sequences \{GDP1\}, \{NLMY1\} and \{API1\}.

As shown in Figure. 3 (1) and sequence \{GDP1\} become stable after a first difference, the model can be preliminarily determined as ARIMA \((5, 1, 0)\). After the significance test and residual autocorrelation test of parameters, only ARIMA\((5,1,0)\) passed the test, so the model is finally determined as ARIMA\((5,1,0)\).The (2) and sequence \{nlmy1\} in Fig. 3 become stable after a first difference, and the model is preliminarily determined as ARIMA\((3,1,1)\), which is finally determined as ARIMA\((3,1,1)\) after testing. According to (3) in Figure 3, it can be judged that this time series combination is suitable for ARMA \((p, q)\) model. It is preliminarily confirmed that only model ARMA\((2,3)\) can pass the significance test and residual autocorrelation test at the same time, so the model is finally determined as ARMA\((2,3)\).

3.2.3. Estimation and diagnosis of model parameters. AC and PAC diagrams of residual sequences of ARIMA\((5,1,0)\) model, ARIMA\((3,1,1)\) model and ARMA\((2,3)\) model were drawn respectively, namely the autocorrelation test diagram of residual, as shown in Figure 4 (1), (2) and (3).

![Figure 4](image)

Figure 4. (1), (2) and (3) are AC and PAC diagrams of residual sequences of AR\((5)\) model of sequence \{GDP1\}, ARIMA\((3,1,1)\) model of sequence \{NLMY1\} and ARMA\((2,3)\) model of sequence \{API2\}, respectively.
It can be seen from the above three figures that the residual sequences of the three selected models are all white noise, which has passed the test. Finally, the best prediction model of time series \{GDP1\} was determined as ARIMA(5,1,0). The best prediction model of time series \{NLMY1\} is ARIMA(3,1,1). The best prediction model of time series \{API2\} is ARMA(2,3).

### 3.3. Model prediction results

#### 3.3.1. Assessment

**Table 4.** Forecast result and comparison of each index in Hubei province.

| time    | indicators                                      | Gross Domestic Product of Hubei Province (100 million yuan) | Gross Output Value of Agriculture, Forestry, Animal Husbandry and Fishery in Hubei Province (100 million Yuan) | Agricultural Products Producer Price Index of Hubei Province (%) |
|---------|-------------------------------------------------|-----------------------------------------------------------|----------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|
| Q3 2019 | The actual value                                | 11664.85                                                   | 2446.89                                                                                                   | 110.30                                                         |
|         | Predictive value                                | 11437.59                                                   | 2317.40                                                                                                   | 108.74                                                         |
|         | Relative error/\%                              | 0.019                                                      | 0.052                                                                                                     | 0.014                                                          |
| Q4 2019 | The actual value                                | 12858.64                                                   | 1882.18                                                                                                   | 123.20                                                         |
|         | Predictive value                                | 12976.67                                                   | 1659.73                                                                                                   | 109.75                                                         |
|         | Relative error/\%                              | 0.009                                                      | 0.118                                                                                                     | 0.109                                                          |
| Q1 2020 | The actual value                                | 6379.35                                                    | 970.14                                                                                                    | 152.00                                                         |
|         | Predictive value                                | 10975.23                                                   | 1125.14                                                                                                   | 108.98                                                         |
| Q2 2020 | The actual value                                | 11101.16                                                   | 1406.05                                                                                                   | 126.00                                                         |
|         | Predictive value                                | 11642.78                                                   | 1182.89                                                                                                   | 106.78                                                         |

As you can see the relative error size in table 4, GDP and the relative error of agricultural producer price index model is more reasonable, and the agriculture, forestry and fishing output model of relative error is bigger, so gross area of Hubei province and agricultural producer price index model is more reliable [6-7], predicted results believable degree is higher. By comparing the predicted data of the first quarter and the second quarter of 2020 with the real data affected by the epidemic, we can estimate the impact of the epidemic on the gross regional product (GDP), gross output value of agriculture, forestry, animal husbandry and fishery, and the production price index of agricultural products in Hubei Province.

#### 3.3.2. Analysis of prediction results

As can be seen from the prediction results in Table 4, the gross regional product (GDP), total output value of agriculture, forestry, animal husbandry and fishery, and producer price index of agricultural products of Hubei Province are all affected the most in the first quarter of 2020. The difference between the predicted GDP under normal circumstances and the true value after the outbreak of the epidemic reached 459.588 billion yuan, that is, under the impact of the
epidemic, it was estimated that the GDP of Hubei Province in the first quarter of 2020 lost 459.588 billion yuan, and the loss of 54.162 billion yuan in the second quarter.

In the first quarter of 2020, the actual output value of agriculture, forestry, animal husbandry and fishery in Hubei Province affected by the epidemic was 97.014 billion yuan, while the predicted output value of agriculture, forestry, animal husbandry and fishery in the first quarter not affected by the epidemic was 112.514 billion yuan, with a loss of 15.5 billion yuan. Compared with the loss of Hubei's gross regional product is small.

In the first quarter of 2020, the production price index of agricultural products is normally predicted to be 108%, but due to the impact of the epidemic, its actual value is 152%. Similarly, in the second quarter of 2020, without the impact of the epidemic, the predicted value of the production price index of agricultural products is 106.78%, while the real value affected by the epidemic is 126%. The actual value affected by the epidemic was higher in both quarters than would have been predicted in the absence of the epidemic, thus indicating a large increase in the price of agricultural products in the first and second quarters due to the epidemic.

4. LSTM neural network-based Measurement of the impact of COVID-19 on agricultural and rural economy in China

4.1. The modeling process
In the case of less available indicators, in order to use the data of the past few years to predict the data of 2020, the data utilization must be maximized, that is, the input and output are a unified sequence. Therefore, the gate control loop network -- LSTM neural network is selected in this paper. In this paper, the LSTM neural network model is constructed from four aspects: data, model, training and evaluation.

4.1.1. Data selection
The use of historical quarterly data of 2020 years ago respectively two quarters of agricultural producer price index, regional GDP and ecological-economic to forecast GDP, considering the data sample size is less and simplify the model structure, this paper will put the calendar year data direct input in the model, at the same time in order to convenient data into periodic prediction, joined in and quarter two dimensions, Finally, the data are standardized.

4.1.2. Model training
Hope that through historical data, this paper makes LSTM forecast data from 2020 years ago two quarters, and so on model predicting the training with historical data, namely training LSTM according to a few months before the ability of data to predict the next month's data, such as after completion of training, we let the LSTM according to historical data to predict the next month's data. The data for the first two quarters of 2020 are iterated by taking the forecast data just output as input. This time, there is only one set of training data, namely the index data of the past years. In order to maximize the data utilization, it is necessary to train the LSTM to predict the data of different months in the same batch. Therefore, the length of the training sequence in the same batch is not uniform, so use the PAD_SEQUENCE tool of TORCH. At the same time, considering that the shorter the data is, the greater the error is, weight is set for loss in the model. The shorter the length of the input sequence is, the smaller the weight of the error is.

4.2. Model prediction results

4.2.1. Assessment. In order to compare and evaluate the predicted data more directly, the graph output is set in the model, where the blue represents the real data curve, the red represents the predicted data curve, and the black represents the dividing line between the training set and the predicted data. Error
represents the degree of deviation between prediction and reality, so the error output item is set in the model.

4.2.2. Forecast of GDP index. It can be seen from the figure that the black line divides the data into two parts. The left part of the black line is the training set and the right part is the predicted result. It can be seen from the figure that the predicted result has a high degree of fit with the actual situation and has a high reference value. See Table 5 for detailed data.

![Figure 5](image)

**Figure 5.** Regional gross domestic product (GDP) LSTM neural network forecast results.

| Time  | The actual value | Predictive value |
|-------|------------------|------------------|
| 2019Q2 | 11219.24         | 11275.156        |
| 2019Q3 | 11664.85         | 11156.9          |
| 2019Q4 | 12838.84         | 12541.088        |
| 2020Q1 | 6379.35          | 10335.7705       |
| 2020Q2 | 11101.16         | 12193.107        |

**Table 5.** LSTM neural network forecast results of regional gross domestic product.

4.2.3. Forecast of gross output value of agriculture, forestry, animal husbandry and fishery. It can be seen from the figure that the black line divides the data into two parts. The left part of the black line is the training set, and the right part is the predicted result. It can be seen from the figure that the predicted result has a high degree of fit with the actual situation and has a high reference value. See Table 6 for detailed data.

![Figure 6](image)

**Figure 6.** Total output value of agriculture, forestry, animal husbandry and fishery is predicted by LSTM neural network.

| Time  | The actual value | Predictive value |
|-------|------------------|------------------|
| 2019Q3 | 2446.89          | 2499.0805        |
| 2019Q4 | 1882.18          | 1723.5587        |
| 2020Q1 | 970.14           | 1161.0413        |
| 2020Q2 | 1406.05          | 1333.3306        |

**Table 6.** Forecast results of total output value of agriculture, forestry, animal husbandry and fishery by LSTM neural network.

4.2.4. Producer price index of agricultural products forecast. It can be seen from the figure that the black line divides the data into two parts. The left part of the black line is the training set, and the right part is the prediction result. It can be seen from the figure that the prediction result has a low degree of fit with the actual situation and is of low reference value.
Table 7. The results of LSTM neural network prediction of agricultural product price index.

| Time   | The actual value | Predictive value |
|--------|------------------|------------------|
| 2019Q3 | 110.3            | 94.674           |
| 2019Q4 | 123.2            | 93.945           |
| 2020Q1 | 152              | 96.85            |
| 2020Q2 | 126              | 92.37            |

Figure 7. Agricultural product price index LSTM neural network forecast results

4.2.5. Forecast of the ratio index between the gross output value of agriculture, forestry, animal husbandry and fishery and the gross local product. It can be seen from the figure that the black line divides the data into two parts. The left part of the black line is the training set, and the right part is the prediction result. It can be seen from the figure that the prediction result has a high degree of fit with the actual situation and has a high reference value.

Table 8. The proportion of LSTM neural network prediction results.

| Time   | The actual value | Predictive value |
|--------|------------------|------------------|
| 2019Q3 | 0.2098           | 0.2135           |
| 2019Q4 | 0.1464           | 0.1361           |
| 2020Q1 | 0.1521           | 0.1205           |
| 2020Q2 | 0.1267           | 0.1165           |

Figure 8. The proportion of LSTM neural network prediction results.

5. Conclusions

Through the prediction of the above two methods, different methods have different prediction effects for different index data. For the time series model, the prediction of total output value of agriculture, forestry, animal husbandry and fishery is ineffective, while the prediction effect of regional gross product and agricultural product price index is better, and the result is more reliable. For the LSTM neural network model of long and short-term memory, the prediction effect of other indicators is good except for the agricultural production price index. Therefore, in general, in order to better measure the impact of the epidemic on China's agricultural and rural economy, the combination of the two methods in this paper has the best combined effect. The prediction results of traditional time series ARIMA model for agricultural product price index can make up for the prediction results of machine learning LSTM long and short memory neural network model for agricultural product price index, and the prediction results of other indicators are suggested to choose the prediction results of LSTM long and short memory neural network model.

Influenced by COVID-19, the GDP of Hubei Province in the first quarter fell seriously, nearly 38.3%, out of the expectation. In the second quarter, there were signs of recovery, but there was still a low decline compared with the expectation. In the first quarter, the gross output value of agriculture, forestry, animal husbandry and fishery was less affected by COVID-19 and suffered relatively low losses. In the second quarter, the output value of agriculture, forestry, animal husbandry and fishery recovered rapidly and even increased by nearly 5.5% compared with the forecast. The producer price
The index of agricultural products was greatly affected by the COVID-19 in both the first and second quarters, and the price growth exceeded the expected level. The total output value of agriculture, forestry, animal husbandry and fishery and the ratio of GDP to GDP in the first and second quarters were also affected by the epidemic, while the ratio was significantly higher than expected.

References
[1] Marcus, R. K., Richard, D. S. The Economic Impact of SARS, How does the Reality Match the Predictions [J]. Health Policy, 2008, 88(1): 110-120.
[2] Huber, C., Finelli, L., Stevens, W. The Economic and Social Burden of the 2014 Ebola Outbreak in West Africa [J]. The Journal of infectious diseases, 2018, 218(5): 698-704.
[3] Monterrubio, J. C. Short-Term Economic Impacts of Influenza A (H1N1) and Government Reaction on the Mexican Tourism Industry: An Analysis of the Media [J]. International Journal of Tourism Policy, 2010, 3(1): 1-15.
[4] Wu Lingyun, Zhao Yuandong. Analysis of the impact of emergency events on economy with time series model [J]. Statistics and Decision, 2004(04):111-112.
[5] Mo Zixing, Wang Jie, Wang Jingyao, Mo Ziyan. An improved SARIMA model for quantitative analysis of economic impact of emergencies [J]. China's Collective Economy, 2009(36):76-77.
[6] Zhou Xinhui, Li Yuzhe, Li Fufu. Evaluation of the impact of COVID-19 on small and medium-sized service-oriented enterprises and countermeasures: Analysis and prediction based on regression algorithm optimization model [J/OL].Economic Review: 1-19[2020-08-11].
[7] Xie Jiazhi, Che Sifang, Lin Yong. Evaluation and prediction of economic losses of earthquake disasters based on stochastic weight neural network [J]. Journal of Catastrophology, 2017, 32(01):1-4+10.