

# Research on GDP Forecast of Ji'an City Based on ARIMA Model

Shenghua Zou

The Department of Basic Course Education, Ji'an College, Ji'an, China

Email: s.zou@foxmail.com

**How to cite this paper:** Zou, S. H. (2020). Research on GDP Forecast of Ji'an City Based on ARIMA Model. *Open Journal of Social Sciences*, 8, 353-365. <https://doi.org/10.4236/jss.2020.812029>

**Received:** November 30, 2020

**Accepted:** December 26, 2020

**Published:** December 29, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In this paper, based on the GDP of Ji'an City from 1978 to 2018, we use the ARIMA(0, 2, 1) model to forecast the GDP of Ji'an City from 2019 to 2023. The prediction results can provide scientific reference for Ji'an City to formulate economic development goals.

## Keywords

GDP, ARIMA Model, MATLAB, Forecasting

---

## 1. Introduction

Since the reform and opening up, China has made remarkable achievements in economic development. The living standards of the people are constantly improving, and we are heading toward the goal of completing the building of a moderately prosperous society in all respects. We can feel the rapid development of China's economy in our daily life, but from the national level, we need specific data to analyze the overall state of the national economy. Gross domestic product (GDP) refers to the market value of all final products produced by all resident units of a country or region within a certain period of time. GDP not only plays an irreplaceable role in reflecting a country's national income, consumption capacity and economic development, but also helps people understand the economic condition of a country or region from a macro perspective. It is the key basis for the formulation of national or regional economic policies, as well as a key means to test whether economic policies are effective and scientific. Therefore, if we use appropriate statistical methods to reveal the law of GDP data changes and make high-precision prediction of short-term GDP, then it will be of great practical significance to the overall planning of macro-economy.

Ji'an City, located in central Jiangxi Province, is a famous Red City on the

shore of Gan River. In the past, Ji'an City lagged behind due to inconvenient transportation and weak economic foundation. Ji'an City had 5 national poverty counties, 4 counties in the Luoxiao Mountain area, and 11 counties in the former Central Soviet District, so the task of poverty alleviation is very difficult. According to Jiangxi Daily, after 6 years of hard work, Ji'an City has gone from 350,000 registered poor households in 2013 to only 6000 people remaining in poverty by the end of 2019. In the past 6 years, the people of Ji'an have handed in a vocal answer for poverty alleviation with sweat, hard work and blood. Today, Ji'an City is focusing on consolidating and improving the results of poverty alleviation, advancing continuous poverty reduction, improving the long-term mechanism, and ensuring that a well-off society is built in sync with the whole country. This year is the final year of the 13th Five-Year Plan and the year of decisive victory for poverty alleviation. However, affected by the global spread of the new coronavirus pneumonia and the deterioration of Sino-US relations, Ji'an City is currently facing many uncertainties in the economy like other regions in China. In this context, the research on the GDP of Ji'an City is helpful for the functional departments of Ji'an City Government to grasp the overall situation of economic growth and formulate targeted policy measures to promote better and faster economic development.

In recent years, more and more scholars have realized the importance of accurately forecasting GDP. They used a variety of methods to forecast GDP from multiple perspectives. The main forecasting methods could be divided into four categories: regression analysis (Jin, 2011; Chen, 2012), gray prediction (Li & Li, 2010; Tian & Liu, 2018; Zhang & Xie, 2019), artificial neural network (Huang, 2007; Zhao, 2017; He, Wu, & Xia, 2020) and ARIMA model (Li & Xue, 2013; Zhou, 2015; Yan, 2018). ARIMA model is popular because of its simple operation and high precision. It is one of the most commonly used methods to forecast GDP. Li and Xue (2013) established the ARIMA model and used multiple screening criteria to determine the optimal ARIMA(6, 1, 3) model. This model has been proved to be highly accurate in predicting China's GDP from 2009 to 2011. Zhou (2015) used ARIMA(1, 2, 3) that passed the test to forecast the GDP of Guangxi during the 13th Five Year Plan period, and put forward countermeasures and suggestions to improve Guangxi's ability to achieve sustained economic growth during the 13th Five Year Plan period. Yan (2018) established ARIMA(1, 1, 1) model and used it to predict the GDP of Shandong province. The test results showed that the prediction effect was good. At present, there is no scholar to forecast the GDP of Ji'an City. Based on the GDP of Ji'an City from 1978 to 2018, in this paper, we establish ARIMA model, and forecast the GDP of Ji'an City in the next five years by using the ARIMA model. It is expected to provide reference for Ji'an City to formulate economic development goals and macroeconomic decisions.

This paper is organized as follows. In Section 2, we give some definitions and ARIMA model modeling steps that will be used in this paper. We use the steps

and methods mentioned in the first section to conduct empirical research in Section 3. In Section 4, we analyze the prediction results.

## 2 ARIMA Model

### 2.1. ARIMA Model

**Definition 2.1.** For time series  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ , the mean function is defined as

$$\mu_t = EX_t, t = 0, \pm 1, \pm 2, \dots$$

The autocovariance function  $\gamma_{t,s}$  is defined as

$$\gamma_{t,s} = cov(X_t, X_s), t, s = 0, \pm 1, \pm 2, \dots$$

The autocorrelation function (ACF)  $\rho_{t,s}$  is defined as

$$\rho_{t,s} = corr(X_t, X_s) = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t} \cdot \gamma_{s,s}}}, t, s = 0, \pm 1, \pm 2, \dots$$

**Definition 2.2.** (Stationary time series) Time series  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  is called stationary time series, if the following three conditions are satisfied:

- 1)  $EX_t^2 < +\infty, t = 0, \pm 1, \pm 2, \dots$
- 2) The mean function does not change with time.  $EX_t = c, t = 0, \pm 1, \pm 2, \dots$ ,  $c$  is a constant.
- 3) The autocovariance function only depends on the time interval, that is

$$\gamma_{t,s} = \gamma_{t+r, s+r}, t, s, r = 0, \pm 1, \pm 2, \dots$$

Autocorrelation function and autocorrelation function describe the degree of correlation between a random event at two different times. In other words, it describes the impact of one's historical behavior on the current situation. However, this kind of correlation is not pure, because when the degree of correlation between  $X_t$  and  $X_{t-k}$  is measured by calculating the lag  $k$ -step autocorrelation function  $a$ , the influence of  $k - 1$  random variable  $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$  is actually doped. Therefore, we introduce the concept of partial autocorrelation function.

**Definition 2.3.** (Partial autocorrelation function) Let  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  be a stationary time series, we call

$$\phi_k = \rho_{X_t, X_{t-k} | X_{t-1}, X_{t-2}, \dots, X_{t-k+1}} = \frac{E[(X_t - \hat{E}X_t)(X_{t-k} - \hat{E}X_{t-k})]}{E[(X_{t-k} - \hat{E}X_{t-k})^2]}$$

a partial autocorrelation function, where

$$\begin{aligned} \hat{E}X_t &= E[X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k+1}], \\ \hat{E}X_{t-k} &= E[X_{t-k} | X_{t-1}, X_{t-2}, \dots, X_{t-k+1}]. \end{aligned}$$

It can be seen that the partial autocorrelation characterizes the degree of correlation between  $X_t$  and  $X_{t-k}$  that is not affected by  $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$ .

**Definition 2.4.** (White noise) If the time series  $\{\varepsilon_t, t = 0, \pm 1, \pm 2, \dots\}$  meets

the following two conditions:

- 1)  $E\varepsilon_t = 0$ ,
- 2)  $cov(\varepsilon_t, \varepsilon_s) = \begin{cases} \sigma^2, & t = s \\ 0, & t \neq s \end{cases}, t, s = 0, \pm 1, \pm 2, \dots$ ,

then  $\{\varepsilon_t\}$  is called white noise, denoted as  $\{\varepsilon_t\} \sim WN(0, \sigma^2)$ .

White noise is a sequence of uncorrelated random variables with equal zero mean variances. It is the simplest and most basic stationary time series. Its importance lies in the fact that white noise is the “generator” of many important models or sequences. Obviously, the autocovariance function  $\gamma_k$  and the autocorrelation function  $\rho_k$  of white noise  $WN(0, \sigma^2)$  are:

$$\gamma_k = \begin{cases} \sigma^2, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

$$\rho_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

This shows that the items of white noise are uncorrelated.

**Definition 2.5.** (ARIMA( $p, d, q$ ) model) Let  $\{\nabla^d X_t, t = 0, \pm 1, \pm 2, \dots\}$  be a stationary time series, we call the model with the following form as ARIMA( $p, d, q$ ) model:

$$\varphi(B)\nabla^d X_t = \theta(B)\varepsilon_t.$$

where  $B$  is the back shift operator,  $\nabla^d \equiv (1 - B)^d$  is the  $d$ -order difference operator,  $\varphi(B) \equiv 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$  is the autoregressive operator polynomial, and  $\theta(B) \equiv 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  is the moving average operator polynomial,  $\{\varepsilon_t\} \sim WN(0, \sigma^2)$ .

## 2.2. ARIMA Model Modeling Steps

1) To test whether the time series  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  to be studied is stationary.

2) Data preprocessing. If the time series data studied is non-stationary, the data must be preprocessed, such as logarithm operation, difference operation and so on, so that the data can be transformed into stationary. Otherwise, move on to the next step.

3) Check whether a stationary time series is a white noise series. If it is, it means that the data does not have any analytical value and the analysis should be stopped; if it is not, then the next step of modeling can be carried out.

4) ARIMA model structure selection. The autocorrelation and partial autocorrelation function graphs of stationary time series are obtained by using statistical analysis software. The model structure is judged by observing the specific characteristics of these two images. The judgment basis is shown in **Table 1**.

As the observation of images is subjective, the judgment may not be so accurate. Therefore, when determining the structure of the model, it is appropriate to select several values around the determined  $p$  value and  $q$  value to construct the ARIMA model.

**Table 1.** ARIMA model structure decision table.

Model	Autocorrelation function	Partial autocorrelation function
AR( $p$ )	Trailing	P-order truncated
MA( $q$ )	Q-order truncated	Trailing
ARMA( $p, q$ )	Trailing	Trailing

5) Determine the specific parameter values in ARIMA model. In the process of building multiple ARIMA models, we investigate whether they can pass the significance test, specifically for the model and for the parameters. The models that fail to pass the two types of tests are discarded, and then the model corresponding to the minimum BIC value is determined as the optimal model by comparing their specific BIC values.

6) White noise test of residual sequence. If the residual sequence passes the white noise test, the modeling can be terminated; if the residual is not white noise, it means that there is useful information in the residual and the model needs to be modified.

7) Use the finally established ARIMA model to make short-term forecasts. The predicted value is obtained by substituting the known data into the relation established by the above model.

The visual operation process of the ARIMA model modeling steps is shown in **Figure 1**.

### 3. Empirical Analysis

#### 3.1. Data Selection and Preprocessing

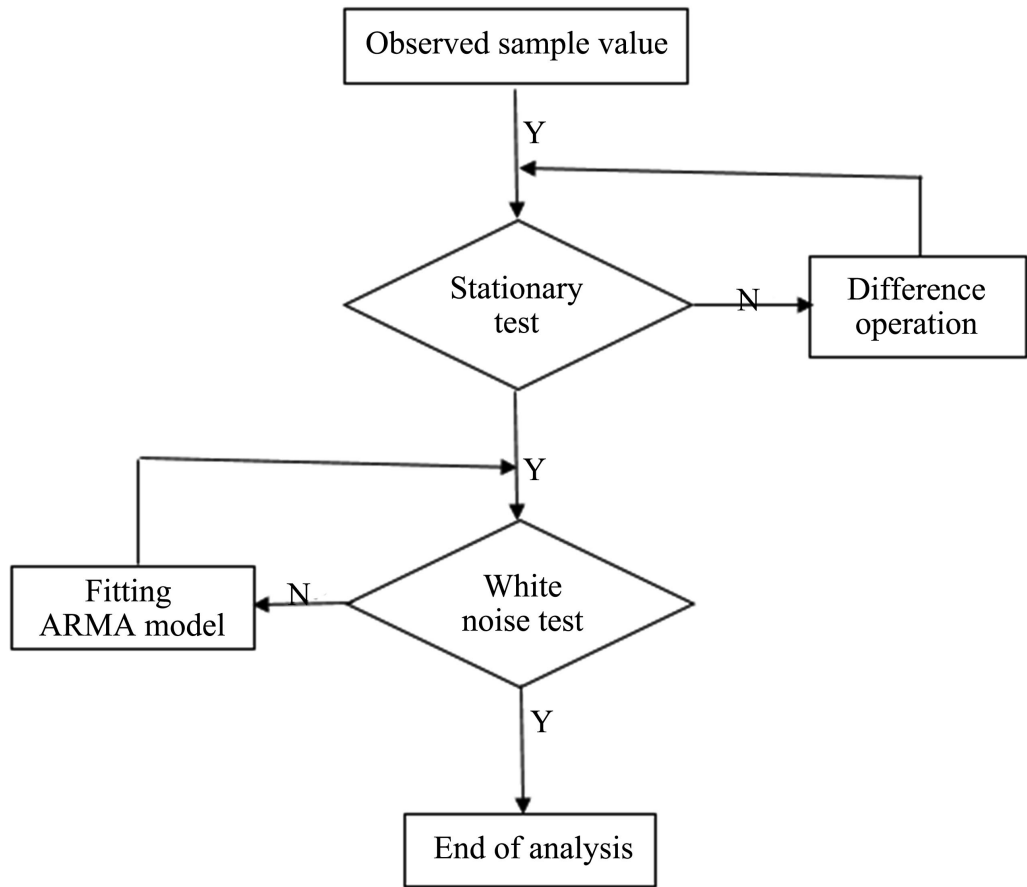
In this article, we use the Ji'an Yearbook released by Ji'an City as the data source, and select the Ji'an City GDP data from 1978 to 2018 as a sample to establish an ARIMA model to predict the Ji'an City GDP from 2019 to 2021.

#### 3.2. Stationarity Test

According to the data in **Table 2**, we use MATLAB software to draw the GDP time series diagram of Ji'an City from 1978 to 2018, as shown in **Figure 2**. It can be clearly seen that the series has an exponential increasing trend and is a non-stationary time series.

In order to eliminate this trend of exponential increase, we perform logarithmic operations with base  $e$  on the original data, and records the processed sample time series as  $\ln\text{GDP}$ , and then use MATLAB again to make a sequence diagram of the new series obtained after taking the logarithm. As shown in **Figure 3**.

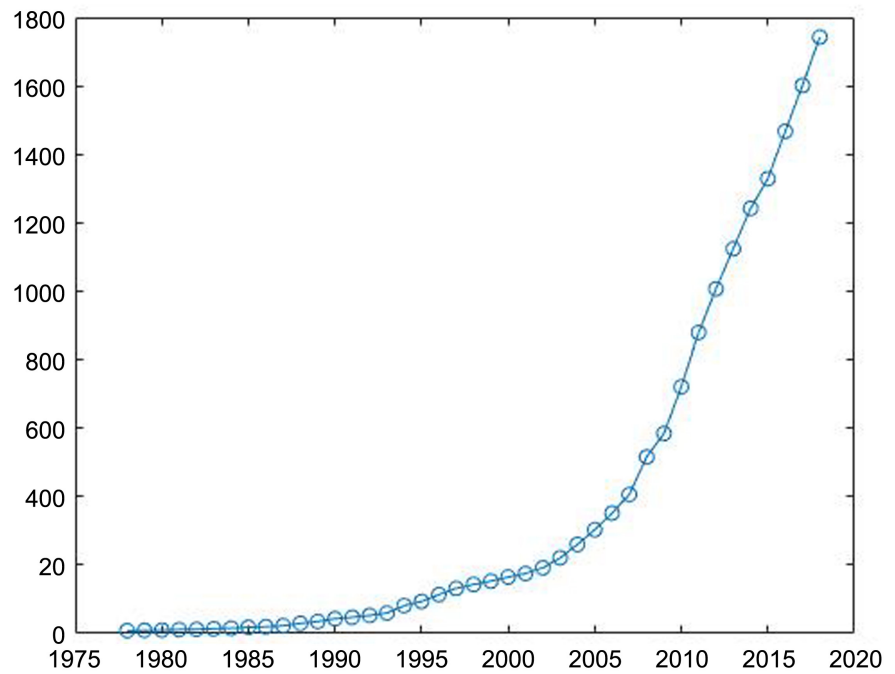
It can be seen from **Figure 3** that the upward trend of time series  $\ln\text{GDP}$  is more gradual, but still presents a linear upward trend. The first-order difference of the time series  $\ln\text{GDP}$  is denoted as  $\nabla\ln\text{GDP}$ , and its sequence diagram is shown in **Figure 4**.



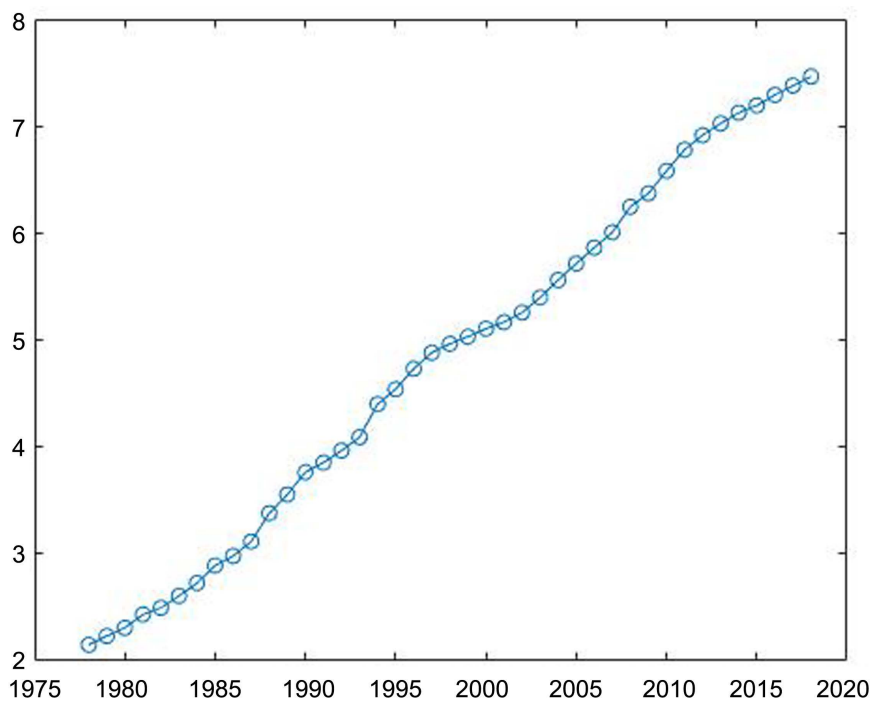
**Figure 1.** ARIMA model modeling process.

**Table 2.** GDP data of Ji'an City from 1978 to 2018 (unit: 100 million yuan).

year	GDP	year	GDP	year	GDP
1978	8.5886	1992	52.7341	2006	351.7803
1979	9.3293	1993	59.8116	2007	406.0052
1980	10.0649	1994	81.4528	2008	515.6101
1981	11.3964	1995	93.648	2009	584.1087
1982	12.1314	1996	113.5076	2010	720.5251
1983	13.5649	1997	131.8252	2011	879.0619
1984	15.2919	1998	143.2171	2012	1006.261
1985	18.0235	1999	153.0699	2013	1123.9013
1986	19.6977	2000	164.8636	2014	1242.1109
1987	22.5486	2001	175.2254	2015	1328.5198
1988	29.3866	2002	191.7888	2016	1467.0327
1989	34.9773	2003	220.8465	2017	1601.1313
1990	43.1361	2004	259.9126	2018	1742.2292
1991	47.1531	2005	303.1422		

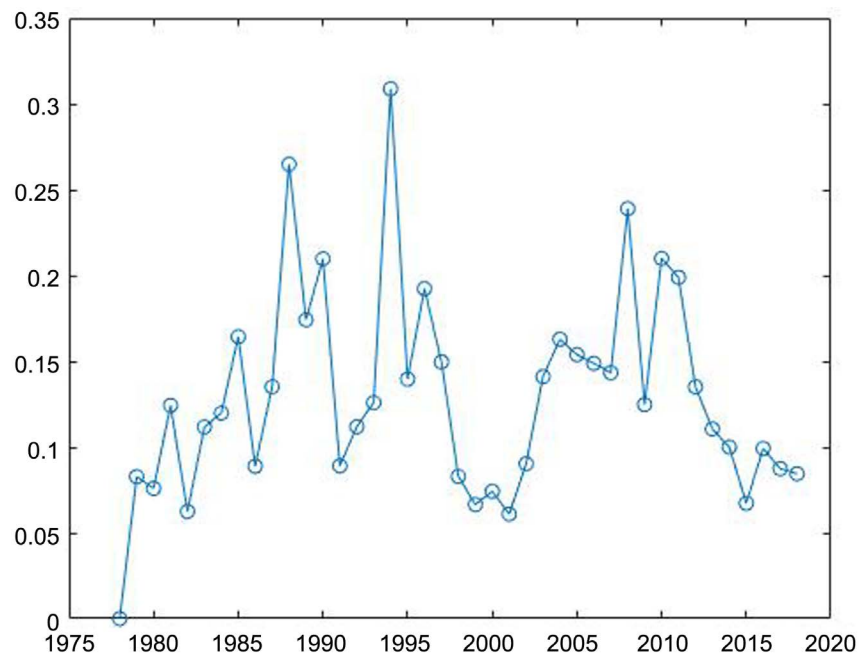


**Figure 2.** Sequence diagram of GDP of Ji'an City over the years.



**Figure 3.** Sequence diagram of lnGDP of Ji'an City over the years.

It can be seen from the image that the sequence obtained by the first difference always oscillates around a certain constant value. It follows that the new sequence is roughly stable. In order to avoid the interference of subjective consciousness and obtain more rational and accurate conclusions, it is necessary to continue to carry out the unit root test, which has become the most widely used



**Figure 4.** Sequence diagram of  $\nabla \ln \text{GDP}$ .

statistical test method of stationarity by establishing the one-to-one correspondence between the existence of unit roots of characteristic equations and the stationarity of sequences. ADF test was carried out on the  $\nabla \ln \text{GDP}$  data of Ji'an City from 1978 to 2018, and the running result of MATLAB program is

$$\text{adftest}(\nabla \ln \text{GDP}) = 0.$$

The results show that,  $\nabla \ln \text{GDP}$  is a non-stationary series. Therefore, in order to obtain the stationary sequence, we still need to do further difference operation. The second-order difference is made for the time series  $\ln \text{GDP}$ , which is denoted as  $\nabla^2 \ln \text{GDP}$ . The time sequence diagram is shown in **Figure 5**.

From **Figure 5**, it can be preliminarily judged that the second-order difference  $\nabla^2 \ln \text{GDP}$  of the time series  $\ln \text{GDP}$  is stationary. Through the ADF stationarity test of the sequence  $\nabla^2 \ln \text{GDP}$ , the MATLAB program running result is

$$\text{adftest}(\nabla^2 \ln \text{GDP}) = 1,$$

which confirms the stationarity of  $\nabla^2 \ln \text{GDP}$ .

### 3.3. Model Recognition and Order Determination

Through the above stationarity analysis, the logarithmic difference order  $d = 2$  has been determined. In order to determine the  $p$  and  $q$  in the  $\text{ARIMA}(p, d, q)$  model, autocorrelation and partial autocorrelation analysis are performed on the second-order difference sequence  $\nabla^2 \ln \text{GDP}$ . The results are shown in **Figure 6** and **Figure 7**. It can be seen from the figure that the autocorrelation function graph is truncated after order 1, and the partial autocorrelation function graph is tailed. According to the characteristics of autocorrelation function and partial autocorrelation function,  $\text{ARIMA}(0, 2, 1)$  is considered for fitting.



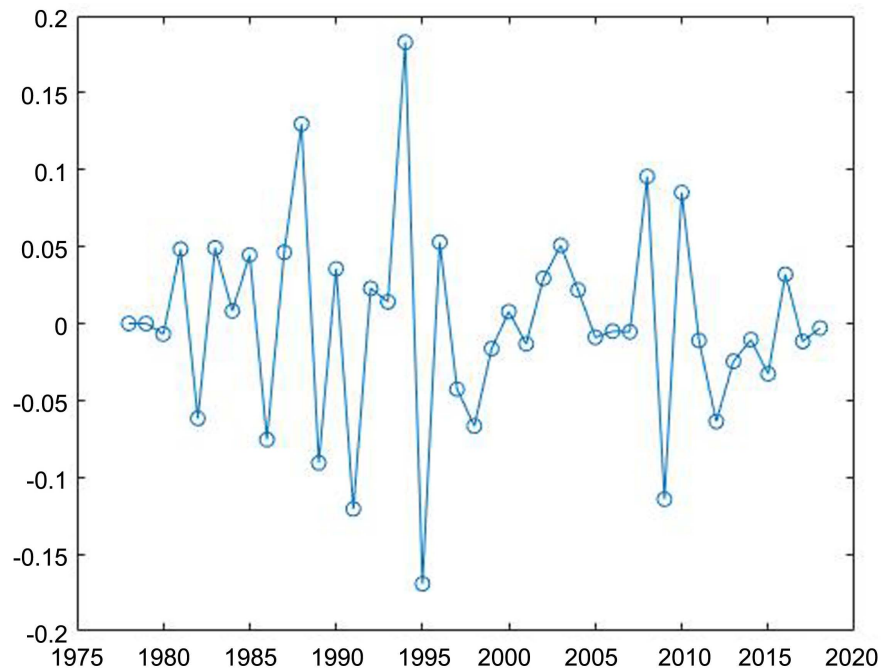


Figure 5. Sequence diagram of  $V^2 \ln GDP$ .

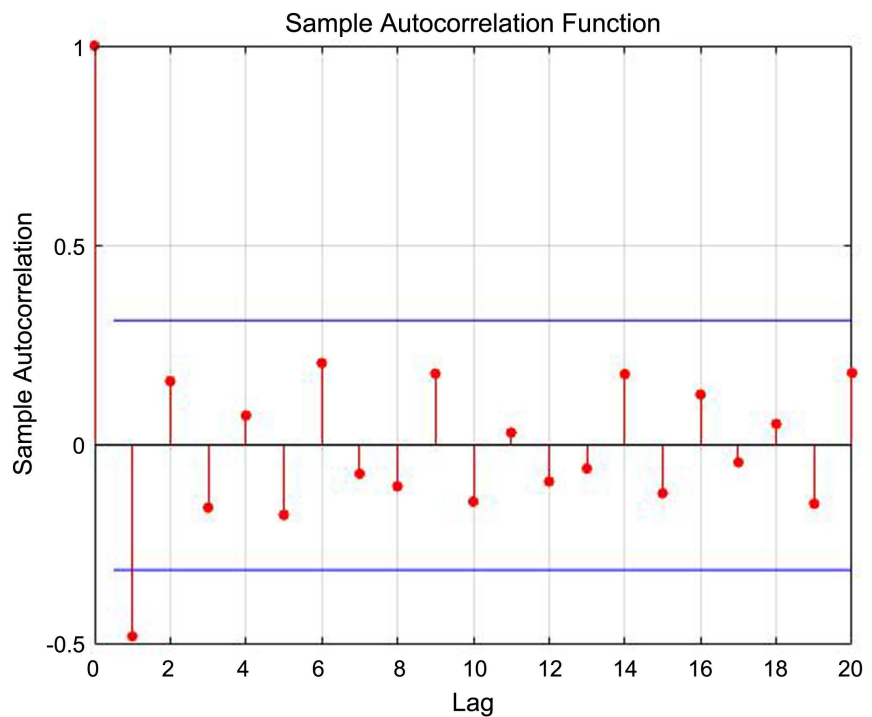


Figure 6. ACF diagram of  $V^2 \ln GDP$ .

It is not accurate to determine the values of  $P$  and  $Q$  only through **Figure 6** and **Figure 7**. According to the BIC criterion, the BIC function values of a limited number of models are investigated within a certain range, and then the relative optimal order is selected. The corresponding BIC values of each model are shown in **Table 3**.

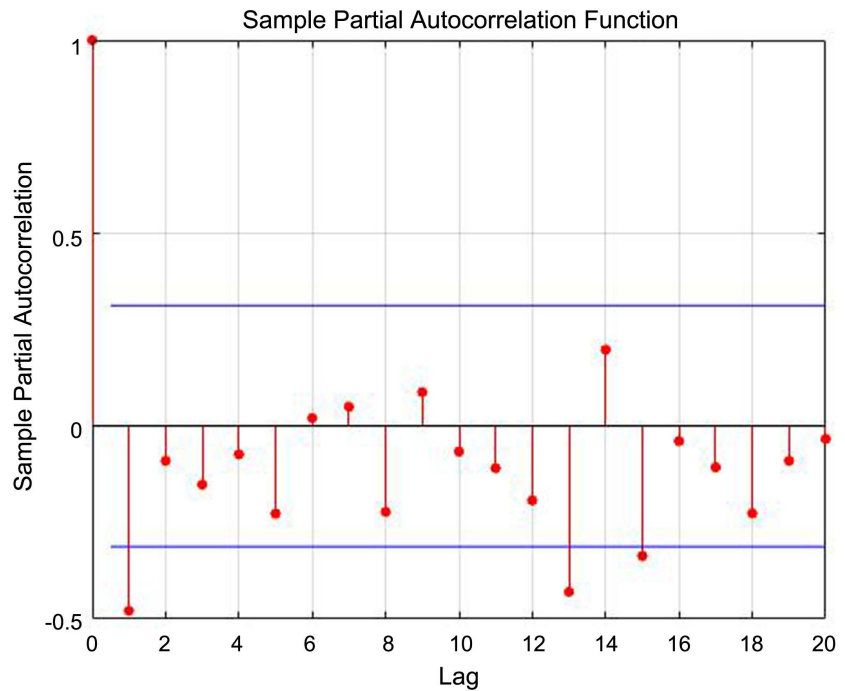


Figure 7. PACF diagram of  $\nabla^2 \ln GDP$ .

Table 3. BIC value of each model.

Model	BIC	Model	BIC
		ARMA(2, 0)	-107.844
ARMA(0, 1)	-112.464	ARMA(2, 1)	-108.042
ARMA(0, 2)	-108.766	ARMA(2, 2)	-104.505
ARMA(0, 3)	-107.402	ARMA(2, 3)	-100.901
ARMA(1, 0)	-111.227	ARMA(3, 0)	-105.08
ARMA(1, 1)	-109.142	ARMA(3, 1)	-104.501
ARMA(1, 2)	-105.645	ARMA(3, 2)	-100.891
ARMA(1, 3)	-102.908	ARMA(3, 3)	-100.834

According to BIC criterion, BIC value is the smallest when  $p = 0$  and  $q = 1$ , which is consistent with the above image analysis results. Therefore, the optimal fitting model of time series  $\nabla^2 \ln GDP$  is ARIMA(0, 0, 1), that is, the optimal fitting model of time series  $\ln GDP$  is ARIMA(0, 2, 1). The results of fitting the model with MATLAB are as follows:

ARIMA(0, 0, 1) Model (Gaussian Distribution):				
	Value	Standard Error	TStatistic	PValue
Constant	0	0	NaN	NaN
MA{1}	-0.59172	0.15305	-3.8661	0.00011058
Variance	0.0031447	0.00055498	5.6663	1.4596e-08

The model expression obtained by ARIMA(0, 2, 1) model fitting is:

$$\nabla^2 \ln \text{GDP}_t = (1 - 0.5917B)\varepsilon_t.$$

Removing the difference, we get the final model expression:

$$\ln \text{GDP}_t - 2 \ln \text{GDP}_{t-1} + \ln \text{GDP}_{t-2} = \varepsilon_t - 0.5917\varepsilon_{t-1}$$

Therefore, the forecast formula of GDP time series  $\text{GDP}_t$  is:

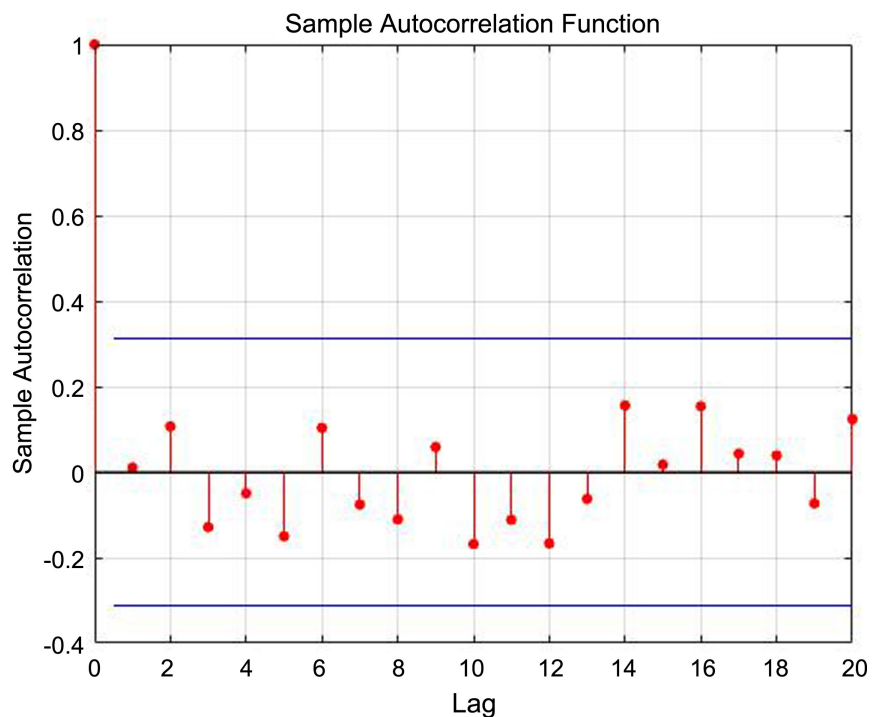
$$\text{GDP}_t = \frac{\text{GDP}_{t-1}^2}{\text{GDP}_{t-2}} e^{\varepsilon_t - 0.5917\varepsilon_{t-1}}$$

### 3.4. White Noise Test

After establishing ARIMA(0, 2, 1), it is necessary to investigate whether ARIMA(0, 2, 1) can extract the information contained in historical data sufficiently. Therefore, the white noise test should be performed on the residual sequence obtained by ARIMA(0, 2, 1). We use the autocorrelation function of the residual sequence to examine its independence, and the result is shown in **Figure 8**.

Obviously, the autocorrelation function of any order is smaller than the standard deviation, so the residual sequence can be determined to be independent of each other. We used the Ljung-Box test to determine the adaptability of the model. The MATLAB program running results are as follows:

$h=$	$p$ Value= $h$	stat= $h$	c Value= $h$
0	0.8450	13.7082	31.4104



**Figure 8.** Autocorrelation function graph of residual sequence.

**Table 4.** GDP forecast of Ji'an city in 2019-2023 (unit: 100 million yuan).

year	2019	2020	2021	2022	2023
Predicted Values	1908.00	2089.55	2288.36	2506.10	2744.55

The results show that  $h = 0$  and  $p = 0.8450 \ll 0.05$ , which indicates that the residual sequence is a white noise sequence. Therefore, the ARIMA(0, 2, 1) model we established is feasible and can be used to forecast the GDP of Ji'an City.

### 3.5. Model Prediction

Now, we use the above ARIMA(0, 2, 1) model to forecast the GDP of Ji'an City in the next 5 years, the results are shown in **Table 4**.

So far, by smoothing the GDP data of Ji'an City from 1978 to 2018, we have established an ARIMA(0, 2, 1) model. On this basis, we obtain the GDP forecast value of Ji'an City in 2019-2023.

## 4. Conclusion

In this paper, ARIMA(0, 2, 1) model is established by using the GDP data of Ji'an City from 1978 to 2018. The adaptability test of the model shows that this model can be used for the short-term forecast of GDP of Ji'an City, which can provide decision-making reference for Ji'an City to make economic plans. According to the predicted results, the GDP growth of Ji'an City in 2019-2023 is good. In this situation, Ji'an City should increase investment in education, health and other livelihood fields to increase the attraction of talents. In the short term, this will increase the financial burden of Ji'an City, but in the long run, the increase of population will make Ji'an more competitive in the future competition, and Ji'an city will achieve more long-term and stable economic growth. However, due to the novel coronavirus epidemic and the relationship between China and the United States, whether the GDP of Ji'an City can achieve the predicted value remains to be tested in practice.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- Chen, M. (2012). Analysis on the Influencing Factors of China's GDP Based on Linear Regression Model. *Guide to Business*, 9, 20-21.
- He, G., Wu, W., & Xia, J. (2020). Application of Grey Neural Network Based on Simpson Formula in GDP Forecast. *Statistics and Decision*, 2, 43-47.
- Huang, X. (2007). *Research on Application of Artificial Neural Network in GDP Forecast*. M.S. Dissertation, Changchun: Jilin University.
- Jin, Y. (2011). *Research on Regression Estimation Method of GDP Accounting*. M.S. Dis-

- 
- sertation, Shanghai: Shanghai Jiao Tong University.
- Li, D., & Li, Z. (2010). Research on GDP Forecast of Guangxi Based on GM (1,1) Model and Grey Correlation Analysis. *Anhui Agricultural Sciences*, *38*, 19704-19706+19710.
- Li, M., & Xue, J. (2013). China's GDP Growth Forecast Based on Optimal ARIMA Model. *Statistics and Decision*, *9*, 23-26.
- Tian, Z., & Liu, M. (2018). An Empirical Study of GDP Forecast Based on Improved Grey GM (1,1) Model. *Statistics and Decision*, *34*, 83-85.
- Yan, Y. (2018). Analysis and Forecast of Shandong Province GDP Based on ARIMA Model. *Mathematics in Practice and Theory*, *48*, 285-292.
- Zhang, H., & Xie, X. (2019). Research on Combination Optimization Model Based on Grey Correlation Degree. *Statistics and Decision*, *35*, 19-23.
- Zhao, L. (2017). *GDP Forecast Based on Particle Swarm Optimization Neural Network*. M.S. Dissertation, Beijing: North China Electric Power University.
- Zhou, L. (2015). Forecast and Analysis of Guangxi's Economic Growth during the 13th Five Year Plan Period Based on ARIMA Model. *Business Economic Research*, *15*, 137-139.