Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 3, No. 2, 2022

Forecast and Analysis of National GDP in China Based on ARIMA Model

Yiqun Liu[1], Xiao Chen[2], Qiqi Wang[3,*], Lipeng Zhu[4], Zejiong Zhou[1]

[1]School of Economics, Anhui University of Finance and Economics, Anhui, China
[2]School of Finance, Anhui University of Finance and Economics, Anhui, China
[3]School of Economics and Management, Anhui University of Science and Technology, Anhui, China
[4]School of Geographic Information and Tourism, Chuzhou University, Anhui, China
[*]Corresponding author: 2816513846@qq.com

Abstract: Gross domestic product (GDP) is an important indicator to measure the economic situation and development level of a country or region, which is of great significance in promoting steady economic development. Therefore, this paper selects the national GDP data from 1978 to 2019, uses Eviews 9.0 software to build a model for the selected time series, and finally determines ARIMA (1, 1, 1) as the optimal model, and forecasts and analyzes the national GDP in 2020 and 2021. The prediction results show that the gap between the predicted GDP in 2020 and 2021 and the actual value is small. Therefore, ARIMA model can better reflect the development trend and short-term forecast of national GDP.

Keywords: GDP, Time series, ARIMA model.

1. Introduction

Gross domestic product (GDP) is the final outcome of production activities of all permanent residents of a country (or region) in a certain period of time. GDP is not only the core indicator of national economic accounting, but also an important indicator to measure the economic situation and development level of a country or region. On the one hand, forecasting the future GDP can provide guidance for promoting macroeconomic development; On the other hand, it is helpful for the government or other financial institutions to make relevant decisions in a timely manner.

At present, many domestic scholars conduct empirical analysis and prediction of GDP from different perspectives and using different methods. The first part is the research on the time series analysis of ARIMA model. Huang et al. (2019) used the ARMA model and OLS to estimate the parameters, studied the residents' consumption level in Hunan Province from 1978 to 2017, and finally concluded that the gap of residents' consumption level in Hunan Province in the next three years is still expanding in general, but the range is decreasing [1]. Wang (2017) selected the monthly report data of OTO e-commerce sales of counties and districts in Longnan from December 2014 to July 2017, and used Eviews8.0 software to establish a stable time series ARMA model to conduct an empirical analysis on the development of agricultural e-commerce in Longnan [2]. The second is the research on the prediction of the results of time series analysis. During the prediction phase, Wang et al. (2019) compared ARIMA model with Holter Winter prediction model and found that the relative errors of prediction values were 0.053 and 0.072 respectively, which proved that ARIMA model had high reliability [3]. In time series analysis and prediction, actual value and predicted value are often inconsistent. When Wang (2017) predicted the e-commerce sales, there was a large difference between the actual value and the predicted value. He believed that the data deviation was caused by the small amount of data and the lack of obvious seasonality [2]. Tang et al. (2017) used Eviews software to study population growth, birth rate, gender ratio between men and women, agricultural and non-agricultural proportion and other relevant demographic indicators, established corresponding time series models such as AR model, MA model and ARIMA model, and predicted the demographic change trend of a region in the next 20 years without the implementation of the two child policy [4]. Hao (2021) used the quarterly time series data from 2005 to 2018 to analyze the impact of financial industry development on the cultural industry from three aspects: direct finance, indirect finance and financial efficiency, and put forward suggestions to promote the coordinated development of financial and cultural industries according to the empirical results [5]. The last part is the research on the combination of ARIMA model and GDP. Qu et al. (2021) studied the GDP data of Hubei Province from 1978 to 2019 using ARIMA model, and predicted the GDP data in 2020 and 2021. The research results show that ARIMA model can better reflect the development trend and short-term forecast of GDP [6]. Wu (2019) took the GDP data of Hangzhou from 1978 to 2017 as a sample, made a first-order difference on the GDP series after selecting natural logarithms and fitted the AR model. After testing the fitted model, the ARIMA model was used to predict the LnGDP of Hangzhou from 2018 to 2022, and then the software was used to restore the GDP of Hangzhou in the next five years. Finally, the conclusion was drawn that the GDP of Hangzhou would increase year by year from 2018 to 2020, with a relatively stable upward trend [7].

Based on the current domestic scholars' short-term GDP forecast, most scholars use the time series model to forecast it, but there are often deviations between the predicted value and the actual value. Therefore, this paper selects the data from 1978 to 2019 and uses ARIMA model to forecast GDP in 2020 and 2021. After comparing the predicted value with the real value, the result shows that the deviation between the predicted value and the real value is small.

2. Introduction to ARIMA Model

The time series analysis method is used to solve the

stochastic, seasonal and stable time series models, which was proposed by Boxing Jenkins. Its basic models mainly include the following four types: autoregressive quadrature moving average model (ARIMA for short), autoregressive model (AR for short), regression moving average model ARMA (p, q) and moving average (MA) model [8].

2.1. ARMA (p, q) model

ARMA (p, q) model is a mixed form of autoregressive model (AR) and moving average model (MA), so it is also called autoregressive moving average mixed model. The equation form is:

$$X_t = C + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \beta_q \varepsilon_{t-q} \quad (1)$$

Where, C represents the constant, and 1, 2, ... , p represents the coefficient value of the autoregressive model. p represents the order value of the autoregressive model. β_1, β_2, ... , β_q refers to the corresponding coefficient value of q order moving average model. ε_t refers to a white noise sequence whose mean and variance are 0 and σ^2.

2.2. ARIMA (p, d, q) model

ARMA model can only be used in stationary time series, but it is no longer applicable to non-stationary time series. However, most economic time series belong to non-stationary time series, and ARIMA model is mainly used to solve non-stationary time series problems. There are two common methods to transform non-stationary series into stationary time series: one is to take logarithms of variables. The second is to do differential processing on the sequence. Only after the non-stationary series are converted into a stationary time series can ARMA model be used for subsequent analysis and prediction.

Let X_t be a stationary sequence of order d, $X_t \sim I(d)$, then $W_t = \Delta^d X_t$. W_t is a stationary sequence, and then ARMA (p, q) model can be constructed. The equation form is as follows:

$$X_t = C + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \ldots + \theta_p W_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \beta_q \varepsilon_{t-q} \quad (2)$$

When $d = 0$ in ARMA (p, d, q) model, it is equal to ARMA model (p, q). ARMA model (p, q) is generally used for modeling and analysis of raw data without difference, that is, stationary time series analysis. However, for the differential data, in most cases, the ARMA (p, d, q) model is used to analyze the unstable time series. It is found that the GDP data selected in this paper from 1978 to 2019 belongs to a non-stationary time series when the unit root test is conducted using the ADF method below. Therefore, this paper uses ARIMA model for subsequent research.

3. Establishment of ARIMA (p, d, q) Model

In order to establish this model, the values of p, d, q must be determined. First, we need to judge whether the selected time series is a stable time series, mainly through two methods: time sequence diagram and unit root test. If the variable increases with time, the series must be non-stationary. Only when the variable floats up and down around a certain value, the series may be a stationary time series, and then the unit root test method is used for further confirmation. There are two common methods to transform non-stationary series into stationary time series: one is to take logarithms of variables. The second is to do differential processing on the sequence. The value of d is the number of times the difference needs to be performed. The values of p and q need to be determined by autocorrelation coefficients (ACF) and partial autocorrelation coefficients (PACF). The selection principles are shown in Table 1 below.

Table 1. Selection principle of ARIMA (p, q) model

ACF	PACF	Model
Tailing	q order truncation	ARMA (p, 0)
q order truncation	Tailing	ARMA (0, q)
Tailing	Tailing	ARMA (p, q)

The second step is to estimate the model parameters, generally including least square method, maximum likelihood estimation method, etc. The third step is to test the model, mainly through whether the parameters of the model have a significance level and whether the estimated residual sequence is a white noise sequence. If these two aspects are met at the same time, the model passes the test, otherwise the model needs to be rebuilt. The fourth step is to make dynamic prediction according to the model, compare the actual value with the predicted value, and further judge the feasibility of the model.

4. Forecast and Analysis of National GDP Level

4.1. Model Identification

4.1.1. Unit Root Test

In this paper, the national GDP data from 1978 to 2019 are downloaded from the website of the National Bureau of Statistics. It can be seen from the time sequence diagram in Figure 1 that the GDP is increasing with time, which can be roughly judged as a non-stationary sequence. Subsequently, ADF method was used to carry out unit root test. The results are shown in Table 2. The P value is equal to 1, that is, the original hypothesis is accepted at the level of 1%, 5% and 10%. Therefore, this series is non-stationary.

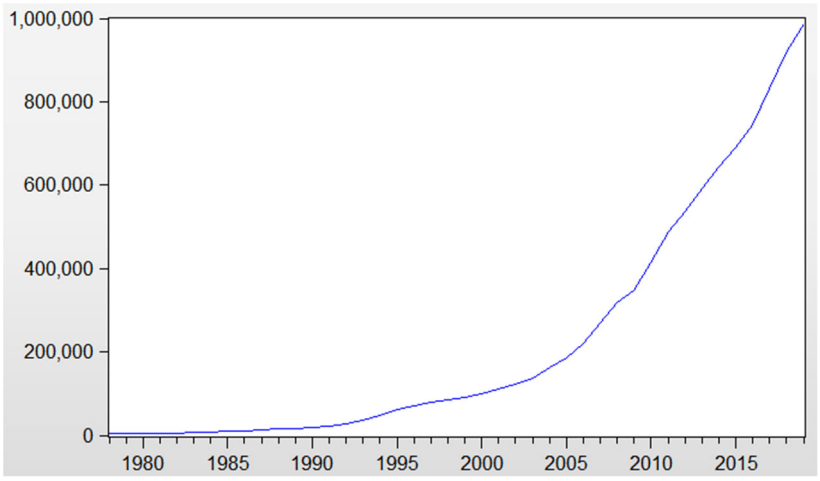

Figure 1. GDP Time Series

Table 2. ADF Test of GDP Series

	t-Statistic	Prob.*
ADF test statistic	4.143889	1.0000
Test critical values: 1% level	-3.646342	
5% level	-2.954021	
10% level	-2.615817	

4.1.2. Stabilization Treatment

In this paper, the original data GDP is logarithmically processed and recorded as LnGDP. According to the above steps, the time series diagram of LnGDP is made and the ADF method is used for testing. The results show that LnGDP increases with time, and the P value of the ADF test result is still at the level of 1%, 5%, and 10%, which are subject to the original hypothesis, and are non-stationary series. Next, take the first order difference of GDP LnGDP and record it as DLnGDP. The stability test results are shown in Figure 2 and Table 3. It can be roughly judged from Figure 2 that this data series belongs to a stationary series. The unit root test results show that the P value is less than 0.05, that is, it is significant at the level of 5% and 10%. This data series has no unit root, that is, DLnGDP belongs to a stationary series. ARMA model can be established.

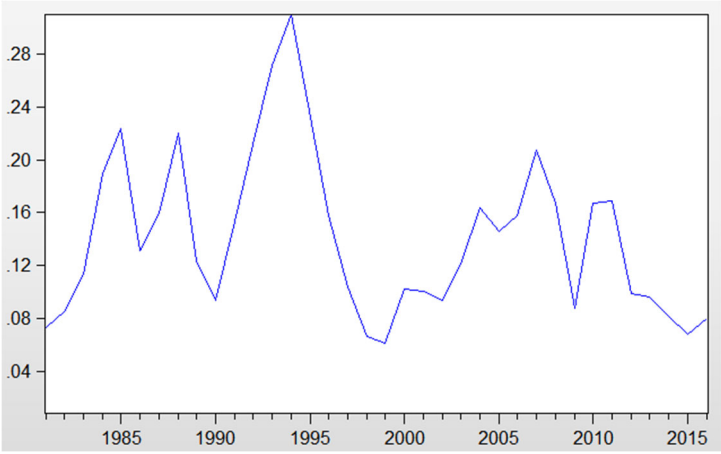

Figure 2. DLnGDP Time Series

Table 3. ADF test of DlnGDP series

	t-Statistic	Prob.*
ADF test statistic	-3.4254	0.0163
Test critical values: 1%level	-3.6210	
5%level	-2.9434	
10%level	-2.6103	

4.1.3. Determining the order

From the stationarity test and stability processing above, we know that d=1. Next, we determine the values of p and q in the ARMA model through the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF). The results of autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) are shown in Table 4. It can be seen from Table 4 that the autocorrelation coefficient is first-order truncation. The partial autocorrelation coefficient shows the first order truncation. However, it may be inaccurate to judge the values of p and q in this way, and it is necessary to compare several different values of p and q. Therefore, the ARMA (1, 3), ARMA (1, 2) and ARMA (1, 1) models are roughly constructed. The results are shown in Table 5-7.

Table 4. Analysis of n order differential autocorrelation and partial autocorrelation

	AC	PAC	Q-Stat	Prob
1	0.665	0.665	19.466	0.000
2	0.274	-0.300	22.860	0.000
3	0.123	0.169	23.561	0.000
4	-0.102	-0.414	24.055	0.000
5	-0.258	0.110	27.308	0.000

Table 5. Results of ARMA (1, 3) model

Variable	Coefficient	Std.Error	t-Statisitic	Prob
AR(1)	0.6666	0.1175	5.6694	0.0000
MA(3)	0.1140	0.1486	0.7672	0.4478
R-squared	0.4665		Mean dependentvar	0.1364
Adjusted R-squared	0.4233		S.D.dependent var	0.0593

Table 6. Results of ARMA (1, 2) model

Variable	Coefficient	Std.Error	t-Statisitic	Prob
AR(1)	0.8816	0.1127	7.1876	0.0000
MA(2)	-0.6156	0.2415	-2.5487	0.0151
R-squared	0.5561		Mean dependentvar	0.1363
Adjusted R-squared	0.5202		S.D.dependent var	0.0593

Table 7. Results of ARMA (1,1) Model

Variable	Coefficient	Std.Error	t-Statisitic	Prob
AR(1)	0.3806	0.1585	2.4022	0.0214
MA(1)	0.7511	0.1548	4.8524	0.0000
R-squared	0.6015		Mean dependentvar	0.1364
Adjusted R-squared	0.5692		S.D.dependent var	0.0593

Table 8. AIC, SC and HQ Values of Three Models

Model Type	AIC	SC	HQ
ARMA (1, 3)	-3.253667	-3.084690	-3.192790
ARMA (1, 2)	-3.426376	-3.259198	-3.365499
ARMA (1, 1)	-3.525932	-3.358754	-3.465055

It can be seen from the above table that the coefficient t statistics of the parameters AR (1) and MA (1) in the ARMA model are very significant, and the values of AIC, SC and HQ obtained from the table are smaller than those of the other two models. Therefore, ARMA (1, 1) model can be selected as the optimal model.

4.2. Model Estimation and Test

According to the above model comparison and analysis, ARIMA (1, 1, 1) is the best model. The parameters are estimated by the least square method, as shown in Table 7. From the data in the table, on the one hand, p values are less than 0.05, indicating that the original hypothesis is rejected at the level of 5%; On the other hand, the adjusted decision

coefficient is 0.5692, which indicates that the model is well fitted. It can be seen from the residual sequence in Table 9 that the estimated residual sequence is stationary and white noise sequence, which means that the model test is passed.

Table 9. Residual Sequence

	AC	PAC	Q-Stat	Prob
1	0.018	0.018	0.0145	——
2	-0.053	-0.053	0.1414	——
3	0.157	0.160	1.2901	0.0256
4	-0.018	-0.029	1.3053	0.521
5	-0.301	-0.291	5.7417	0.125

4.3. Time Series Prediction

Next, Eviews 9.0 measurement software will be used to forecast the national GDP in 2020 and 2021. The predicted series is named LnGDPf. Open LnGDPf and LnGDP as an array and draw the corresponding image. The result is shown in Figure 3. The results show that the difference between the predicted and actual values of LnGDP in 2020 and 2021 is small, indicating that the predicted results are basically in line with the actual situation.

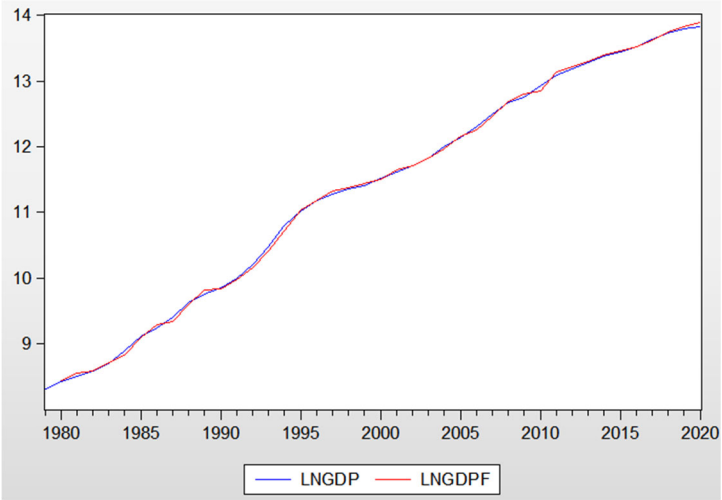

Figure 3. Image of LnGDPf and LnGDP

Next, Eviews software is used to convert LnGDP into GDP, and the actual value in 2020 and 2021 is further compared with the predicted value. The results are shown in Table 10. The results show that there is little difference between the actual value and the predicted value, which is basically in line with expectations.

Table 10. Real and Forecast GDP in 2020 and 2021

Year	Real GDP (100 million yuan)	Estimated GDP (100 million yuan)
2020	1013576	1076224
2021	1143669.7	1063174.7

5. Conclusion

This paper selects the GDP data from 1978 to 2019, makes a first-order difference on the LnGDP series and simulates the corresponding model. At the same time, it forecasts the GDP in 2020 and 2021 through ARIMA model and obtains the final results. The goodness of fit of the model shows that the model used this time has a high degree of fit. Comparing the predicted data with the actual data, the difference is small, which means that the model used this time has a good effect on short-term prediction.

References

[1] Huang Bing, Li Wei. Analysis and prediction of residents' consumption level based on ARMA model -- taking Hunan Province as an example [J]. Marketing Circle, 2019 (46): 157-158.

[2] Wang Shengqing. Analysis and forecast of e-commerce sales of agricultural products based on ARMA model - Eviews8.0 demonstration from Longnan data [J]. Journal of Gansu Normal University, 2017, 22 (09): 24-28.

[3] Wang E, Zhang Ting. Application of Time Series in GDP Forecast of Hunan Province -- Based on ARIMA Model [J]. Journal of Qingdao University (Natural Science Edition), 2019, 32 (03): 136-140.

[4] Tang Jieru, Zhu Xiaxia, Zhou Lijian. Prediction and research of population structure based on time series [J]. Journal of Jiamusi University (Natural Science Edition), 2017, 35 (04): 696-699.

[5] Hao Yuxing. Empirical Research on Financial Support to the Development of Cultural Industry -- Time Series Data Analysis Based on Beijing [J]. Business News, 2021 (15): 87-88.

[6] Qu Haiqing, He Xianping. Research on Hubei GDP Forecast Model Based on Time Series Analysis [J]. Journal of Hubei University of Economics (Humanities and Social Sciences Edition), 2021,18 (09): 37-39.

[7] Wu Boni. GDP Forecast of Hangzhou City Based on Time Series Model [J]. Enterprise Reform and Management, 2019 (23): 212-213.

[8] Guzarati. Fundamentals of Econometrics [M]. Beijing: China Renmin University Press, 2006.