International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n3.32
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Gross Domestic Product forecasting in Shanghai based on Auto Regressive Integrated Moving Average Model

Yujun Cao *

Shanghai Ulink College, Shanghai, 201615, China
*Corresponding Author: julie07_cyj@163.com

ABSTRACT

GDP is an important reflection of economic development, and the correct prediction of GDP can provide an important reference for predicting economic development. The economy of various provinces and cities is an important influencing factor in China's GDP, and Shanghai, as a province and city with rapid economic development, has played a key driving role in the national economy. Therefore, this study will analyze the GDP value of Shanghai. This paper analyzes the GDP data of Shanghai from 2000 to 2022, establishes the ARIMA (0, 2, 2) model, and uses this model to predict Shanghai GDP data in 2023. Compared with the actual value, it is concluded that the model has high credibility and high prediction accuracy, and can be used for Shanghai. The city's GDP is forecasted in the short term to provide reference for government departments to formulate corresponding response policies.

KEYWORDS

Time Series analysis; Economic forecast; ARIMA; GDP; Shanghai

1. INTRODUCTION

In recent years, the economic popularity of research at home and abroad has been increasing, and many scholars have used the application of the time series model in GDP prediction to conduct research and analysis from different aspects. Gwilym and George [1-5] mainly studied the concept and model introduction of time series, as well as the detailed establishment process of the ARIMA model, which is divided into model identification, estimation, testing and prediction, as well as modeling steps and flowcharts. Zhao Lei [6] and others used the ARIMA model to predict and analyze the GDP of Fujian Province from 1978 to 2004, and concluded that the prediction results of the ARIMA (5, 3, 6) model were the closest to the actual GDP of Fujian Province in 2005. Zhang Wentao [7] and others analyzed the GDP of Henan Province and used the residual autoregressive model to verify that ARIMA (0, 1, 4) can predict the trend of GDP in Henan Province in the short term. Qu Haiqing [8] and others also used the ARIMA model to first predict the GDP of Hubei Province from 1978 to 2017, optimize the model by comparing the predicted GDP values of Hubei in 2018 and 2019 with the actual values, and then predict the GDP of Hubei Province in 2020 and 2021, and finally the error between the predicted value and the actual value obtained by the model ARIMA (0, 2, 3) is the smallest. Li Chenfei [9] and others used STATA statistical analysis software to analyze the GDP data of Hubei Province from 1978 to 2010, established the ARIMA (0, 1, 1) model, and predicted the GDP data of Hubei Province from 2014 to 2015, which basically conformed to the actual value.

Therefore, we know that the optimal prediction models are not the same when predicting the GDP of different provinces and regions. The goal of this study is to find an ARIMA model that can predict the closest to the actual GDP value of Shanghai in recent years. The innovation of this topic is to propose the latest ARIMA model suitable for predicting Shanghai's GDP. By analyzing the GDP of Shanghai in recent years and its influencing factors, the ARIMA model is used to make the prediction closest to the actual value, and the model is optimized. This can help people predict the GDP trend of Shanghai and expect the corresponding policies.

2. INTRODUCTION OF ARIMA MODEL

The ARIMA model is a time series model that combines differential operation and ARMA model. Compared with other traditional models, the ARIMA model has higher fitting accuracy and can predict more accurate values. In the ARIMA model (p, d, q), p is the order of the self-regression term, d is the number of differencing required to make the time series stationary, and q is the number of terms of the moving average term. The ARIMA model consists of three main parts: auto-regression (AR), difference (I) and moving average (MA). These three parts correspond to the parameters p, d and q in the model respectively.

2.1. The Structure of ARIMA Model

The model structure of ARIMA (p, d, q) is as follows:

(1) $\Phi(B) = 1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_p B^p$

(2) $\Theta(B) = 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$

Among them, B represents the meaning of the delay operator; in equation (1), the order is the polynomial of the autoregressive coefficient of p, and in formula (2), it refers to the polynomial of the moving average coefficient of order q.

2.2. ARIMA (p, d, q) Modeling Steps

(1) Stability test: By drawing a time series line chart, ADF test the data and judge its stability.

(2) Model recognition: Observe the sequence diagram and estimate its p and q.

(3) Model test: analyze and test the model according to the values of AIC.

(4) Model prediction: Use the established ARIMA model for data prediction.

(5) Comparison: Compare the predicted GDP value with the actual GDP data of Shanghai in 2023.

3. ESTABLISHMENT OF ARIMA MODEL

3.1. Source of Data

This study selected the GDP of Shanghai from 2000 to 2023 as the research object, and established the ARIMA (0, 2, 2) model. The data are all from the Shanghai Municipal Bureau of Statistics. Table 1 lists the GDP of Shanghai from year 2000 to 2022.

Table 1. 2000~2022 Shanghai GDP (unit: 100 million yuan)

Year	GDP	Year	GDP
2000	4812.15	2012	21305.59
2001	5257.66	2013	23204.12
2002	5795.02	2014	25269.75
2003	6804.04	2015	26887.02
2004	8101.55	2016	29887.02
2005	9197.13	2017	32925.01
2006	10598.86	2018	36011.82
2007	12878.68	2019	37987.55
2008	14536.90	2020	38963.30
2009	15742.44	2021	43653.17
2010	17915.41	2022	44652.80
2011	20009.68		

3.2. Data Stability Test

According to the data in Table 1, we use R language software to make a time path map of GDP data, Figure 1. We generate the dat into time series to see how GDP of Shanghai change during the selected period. The image generally shows an upward trend, so the time series is not stable.

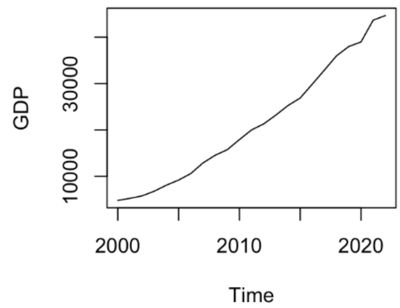

Figure 1. Time path map of Shanghai GDP in 2000~2022

In order to determine whether the data is stable, we conduct an ADF unit root test for the data. The results are shown in Table 2. After the difference, the data after the second-order difference is basically smooth and passed the ADF test. As shown in Figure 2, the sequence of second-order difference is basically stable. From this, we can get the differential order d=1 in the ARIMA (p, d, q) model.

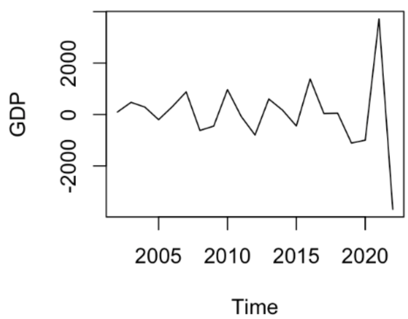

Figure 2. Second-order differential line chart of GDP

Table 2. ADF test results

Variable	Differential order	t	p	AIC	Critical value		
					1%	5%	10%
GDP (100 million yuan)	0	3.3283	0.7239	351.51	-2.66	-1.95	-1.60
	1	0.0338	0.3655	351.75	-2.66	-1.95	-1.60
	2	-5.9539	0.03887	351.75	-2.66	-1.95	-1.60

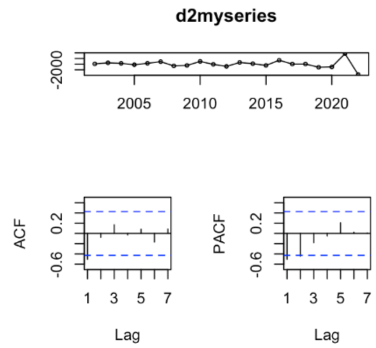

Figure 3. ACF and PACF diagram

According to the drawn ACF diagram and PACF diagram as shown in Figure 3, combined with R code, automatically find the optimal solution and obtain the ARIMA model based on second-order differential data (0, 2, 2)

3.3. Model Prediction

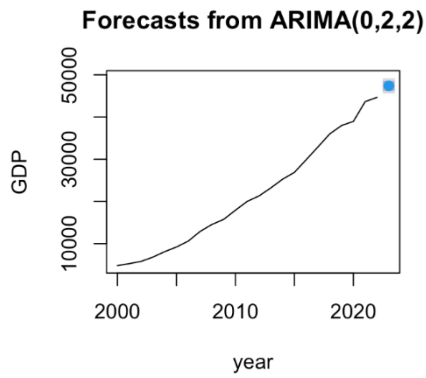

Figure 4. Predicted trend of GDP of Shanghai in 2023

As shown in Figure 4, the blue dot range represents the range of the predicted value. The predicted GDP of Shanghai in 2023 is 4,739.239 billion yuan. Table 3 compares the GDP values and forecast values of Shanghai in 2023 recorded by the Shanghai Municipal Bureau of Statistics. We can conclude that the error of the predicted data and actual values of the ARIMA (0, 2, 2) model is 0.368%. The difference is very small, this reflects the high accuracy of the ARIMA model for short-term data prediction.

Table 3. Comparison between predicted GDP and real GDP of Shanghai in 2023

Year	Predicted GDP (100 million yuan)	Real GDP (100 million yuan)
2023	47392.39	47218.66

4. CONCLUSION

The above study analyzes and predicts the GDP data of Shanghai from 2000 to 2022. First, the data is divided by second-order, and then the ADF test is carried out to test the stability of the sequence. Then, combined with ACF and PACF charts, a model ARIMA (0, 2, 2) suitable for predicting Shanghai's GDP is obtained. Finally, the model is used to predict the value of Shanghai's GDP in 2023, which has a small error with the actual value.This shows that the ARIMA model has a good predictive effect in the short term. Moreover, this model can be used to study the GDP data forecast of Shanghai in the next few years, and the forecast value is consistent with the actual economic development. Therefore, ARIMA (0, 2, 2) has certain practical significance, which can provide reference for the economic development of Shanghai, and can also help people predict the GDP trend and expect corresponding policies.

The ARIMA model is the most recent and advanced time series prediction model. The model is simple, suitable for stable time series prediction, and can predict the trend of time series more accurately. However, the disadvantage of ARIMA is that it requires the stability of data. And it is impossible to deal with other multiple factors that cause changes in time series. This study only predicts the GDP data of Shanghai in 2023, because ARIMA will have large errors in the long-term forecast. Since the model cannot reflect the factors affecting economic development, this study uses the predictions made by ARIMA for reference only.

REFERENCES

[1] George Edward Pelham Box. Science and Statistics [J]. Journal of the American Statistical Association, 1976 (71)791-799.

[2] George Edward Pelham Box. Empirical Model-Building and Response Surfaces [M]. State of New Jersey: John Wiley & Sons, Inc, 1987.

[3] Gwilym Meirion Jenkins. Case Studies in Time Series Analysis [M]. State of New Jersey: John Wiley & Sons, Inc, 1983.

[4] Gwilym Meirion Jenkins. Practical Experience with Modelling and Forecasting Time Series [M]. State of New Jersey: John Wiley & Sons, Inc, 1979.

[5] George Edward Pelham Box, Gwilym Meirion Jenkins. Time Series Analysis: Forecasting and Control [M]. State of New Jersey: John Wiley & Sons, Inc, 1970.

[6] Zhao Lei, Chen Meiying. Application of ARIMA model in GDP forecasting in Fujian Province [J]. Technology and Industry, 2007, 7(1):45-48.

[7] Zhang Wentao, Li Yingqi. Analysis of GDP index in Henan Province based on ARIMA model [J]. Journal of Luoyang Normal University, 2019, 38(2):11-14.

[8] Qu Haiqing, He Xianping. Research on the GDP prediction model of Hubei Province based on time series analysis [J]. Journal of Hubei University of Economics (Humanities and Social Sciences Edition), 2021, 18(09):37-39.

[9] Li Chenfei, Chang Jie, Shen Yan. Application of ARIMA model in GDP forecasting of Hubei Province [J]. Journal of Hubei Normal University (Philosophy and Social Sciences Edition), 2015, 35(04):62-66.