Forecasting Analysis of Guangzhou's Gross Domestic Product Based on ARIMA Model

Minjing Peng[a], Zhiyu Liu*[b]

[a] School of Economics and Management, Wuyi University WYU SEM Jiangmen, China;
[b] School of Economics and Management, Wuyi University WYU SEM Jiangmen, China

* Corresponding author: liuee214@163.com

ABSTRACT . ARIMA model is a common method for time series analysis and fore-casting. GDP is an essential indicator of the social and economic development of a country or region. It plays an instrumental role in understanding the economic strength of a country or region as well as the trend and status of economic development. This paper establishes an ARIMA model by selecting the GDP data of Guangzhou from 1990 to 2021 to analyze the economic development of Guangzhou. First of all, the stationary test and processing are carried out on the data. Then, the parameters are determined based on the model and the suitable ARIMA model is chosen. Finally, use the model to predict the GDP data of Guangzhou in the next three years, and provide scientific reference and suggestions for its future urban development according to the urban characteristics of Guangzhou.

Keywords: ARIMA model; GDP; time series model; predictive analysis

1 INTRODUCTION

GDP is the core indicator of national economic accounting, reflecting the economic strength and market size of a country or region. After the reform and opening up, Guangzhou's economy has developed vigorously, and now it has entered a new economic normal, and the high-speed economic growth in the past is no longer applicable. Facing the new economic normal, Guangzhou needs to actively explore and think about how to continue to develop in a new economic model. On this basis, establishing an effective model to predict the development trend of Guangzhou's future GDP will provide a reference for future economic development and is conducive to the healthy development of the macro economy.

ARIMA models, also called autoregressive moving average models, are commonly used to predict prices and sales[12]. Contemporary forecasting of GDP is mainly achieved through econometric models, machine learning models and gray system theory. Longyue Liang and Yuxia Chen study quarterly GDP data forecasting in China based on a deep learning long short-term memory (LSTM) neural network model[3]. Dong Zhonghao and Gao xingyu used the gray forecast model to predict the GDP of

A. Rauf et al. (eds.), *Proceedings of the 3rd International Conference on Management Science and Software Engineering (ICMSSE 2023)*, Atlantis Highlights in Engineering 20,
https://doi.org/10.2991/978-94-6463-262-0_41

Shandong Province in the next five years, and the results show that the GDP of Shandong Province will maintain a stable growth trend in the next five years. [4]. Musundi Sammy Wabomba et al. selected the annual GDP data of Kenya from 1960 to 2012 for the study based on ARIMA model and established ARIMA(2,2,2) model to forecast Kenya's GDP for the next 5 years, and the results showed that the relative values and forecasts are within 5% and the forecasts are relatively effective[5].

To sum up, this article will use the ARIMA model to analyze and predict Guangzhou's GDP, select Guangzhou's GDP data from 1900 to 2020, establish an ARIMA model suitable for Guangzhou's GDP, and predict the development trend of Guangzhou's future GDP. Provide a reference for future economic development. Conducive to the healthy development of the macro economy.

2 Methodology

2.1 ARIMA Model Structure

ARIMA model is a time series forecasting method based on stochastic theory. It is a stochastic theory based time series forecasting method, which was introduced by Box and Jenkins in 1976[6]. The ARIMA (p, d, q) model is determined by three important parameters: p means that the sequence value lags p order[7]. d means that the time series data needs to be differentiated at least d times to become a stationary sequence. q means that the error term lags by q order.

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

$$\begin{cases} \Phi(\mathrm{B})(1\text{-}\mathrm{B})^d Y_t = \Theta(B)\alpha_t \\ E(a_t) = 0, Var(a_t) = \sigma^2, E(a_t, a_s = 0, t \neq s) \\ E(Y_t, a_t) = 0, \forall s < t \end{cases} \tag{1}$$

$$\Phi(\mathrm{B})=1\text{-}\varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \tag{2}$$

$$\Theta(\mathrm{B})=1\text{-}\theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p \tag{3}$$

In the structure, B represents the meaning of the delay operator. Equation 2 represents the p-order autonomous coefficient polynomial. Equation 3 represents the q-order moving average coefficient polynomial. $\{\alpha_t\}$ represents the white noise sequence.

2.2 Modeling steps of ARIMA model

First, determine whether the sequence is stationary. It is necessary to judge the original data, draw the time series diagram of the original data, and judge whether the sequence is stable. If the sequence is not stable, it needs to be processed by some mathematical methods, such as differential transformation, so that the sequence becomes a stationary sequence. The parameter d in the ARIMA model is determined by the number of differences.

Second, ARIMA model fitting is performed on the post-difference stationary series. Estimating the parameters of the ARIMA model and analyzing the partial autocorrelation of the stabilized series, as shown in Table 1, according to the truncation situation and estimated parameters p and q values, the model fitting is carried out.

Finally, carry out model testing, observe the fitting effect of the model through visual charts, test the significance of parameters. The residual series is determined by testing whether the residual series of white noise series.

Table 1. Parameters p, q and determination conditions of the model

MODEL	ACF	PACF
AR(p)	Trailing	Truncation after p-order
MA(q)	Truncation after q-order	Trailing
ARMA(p,q)	Tailing after q-order	Tailing after p-orde

3 Empirical Analysis of ARIMA Model

3.1 Data Collection and preprocessing

In this paper, annual GDP data of Guangzhou City from 1990 to 2021 are chosen for empirical analysis. The data are sourced from the Guangzhou Statistical Yearbook.

From Figure 1, the original time series of Guangzhou's GDP, it can be seen that Guangzhou's GDP shows an exponential trend, and the growth trend has been stronger since 1998. It can be preliminarily judged that the sequence is non-stationary.

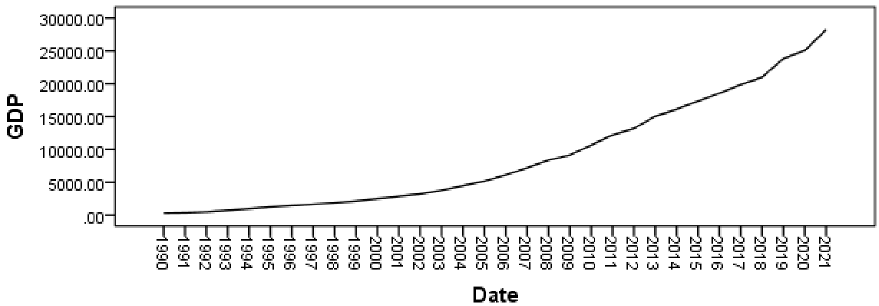

Fig. 1. The original time series of Guangzhou's GDP from 1990 to 2021

The ARIMA model requires a stationary time series, and then the ACF and PACF graphs are used for testing, and the results are shown in Figure 2. The ACF graph looks like a tail, and the subsequent data has an increasing trend, and there is no fluctuation, indicating that the original sequence has a monotonous trend. However, the PACF graph shows a first-order tail, and the coefficient fluctuates in a small range under the zero axis, so it can be judged that the original time series is not a stationary sequence.

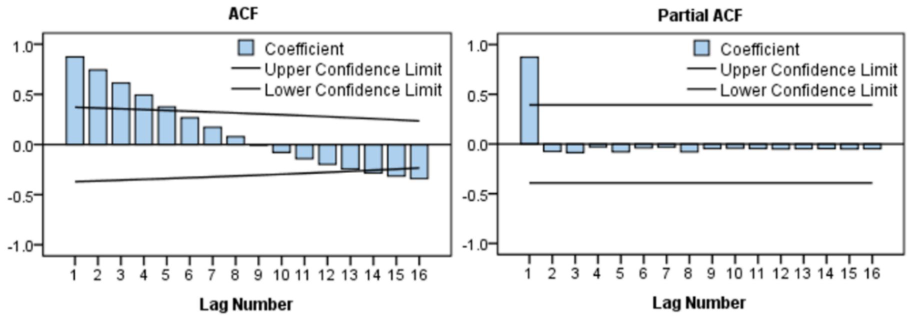

Fig. 2. ACF and PACF diagram of the original sequence

Because the original series is not smooth, the original series is differenced. The sequence is still not smooth after first-order differencing, so second-order differencing is performed. As shown in Figure 3, compared with the original sequence, the sequence after second-order differencing basically conforms to the characteristics of a stationary time series. The values are fluctuating around a fixed value, and the growth trend of the series is basically eliminated.

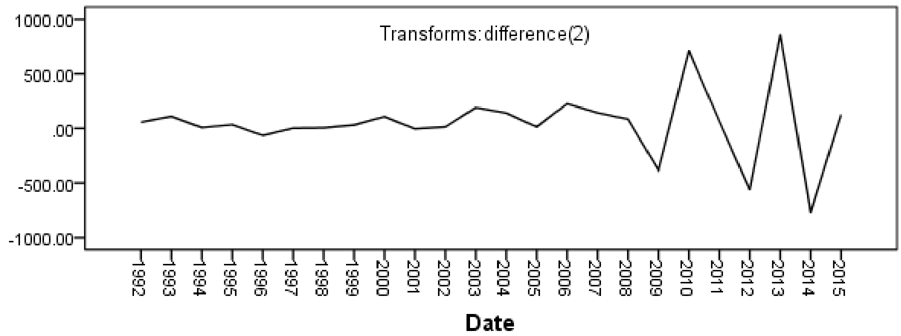

Fig. 3. The original time series of Guangzhou's GDP from 1990 to 2021

3.2 ARIMA model identification

According to the selection principle of the parameters of the ARIMA model, according to the above stationarity test, the original data is divided by the second order, and then the corresponding parameter d value in the AIRMA model can be determined to be 2.

Then, the parameters p and q in the model are established by examining the autocorrelation after the second-order difference. The ACF and PACF diagrams after the second order difference of the original sequence are shown in Figure 4. According to the ACF graph showing truncated phenomenon after the first order, there is no tailing, and the judgement sequence is truncated in the first period, so the p-value PACF graph taken from the regression order shows truncated phenomenon after the second order, and there is no dragging tail, it can be determined that the value of the parameter q is 1, and the value of the parameter p is 2. Therefore, the model in this paper is ARIMA(2, 2, 1).

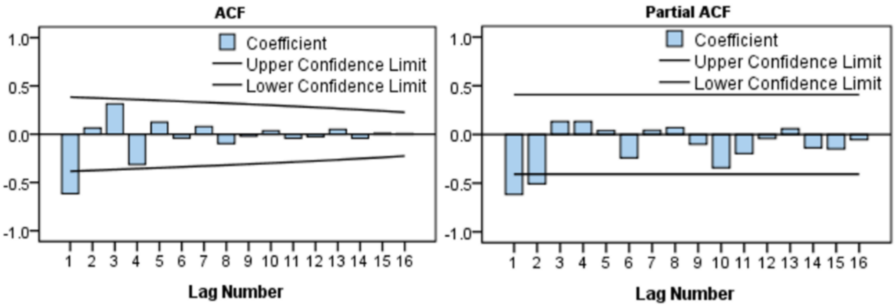

Fig. 4. ACF and PACF diagram of second-order difference sequences

3.3 Result and discussion

According to the definite model ARIMA (2, 2, 1) determined in the previous section, the modeling is carried out in SPSS software, and the GDP of Guangzhou is selected as the dependent variable to create the ARIMA model. From Table 2, it can be shown that the stable R-squared is 0.695 (greater than 0.5), and the value of R square reaches 0.999. The significance of Ljung-Box Q(18) after the second order difference is greater than 0.05, indicating that there is no autocorrelation in the ARIMA model (2, 2, 1).

After the ARIMA model is built, it is indispensable to perform a white noise test on the residual of the model to ensure that the model does not have serial correlation. The test results are shown in Figure 5. All the autocorrelation coefficients of the residuals fall within the confidence interval and tend to approach 0 gradually. The null hypothesis that the residuals are uncorrelated cannot be rejected.

In summary, it can be determined that the residual item after fitting the model is a white noise sequence, and the ARIMA (2, 2, 1) model has passed the test.

Table 2. Model Statistics Table

Model Fit statistics		Ljung-Box Q(18)		
Stationary R-squared	R-squared	Statistics	DF	Sig.
0.695	0.999	7.804	15	0.931

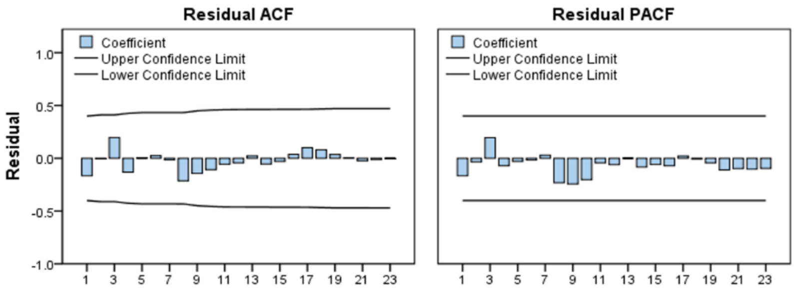

Fig. 5. Residual ACF and PACF diagrams after ARIMA(2,2,1) model fitting

The ARIMA (2, 2, 1) model was used to forecast the GDP of Guangzhou city, and in order to test the accuracy of the model's forecast results, the actual data from 2015-2021 were used as observations to evaluate the forecast accuracy. From Table 3, it is seen that the real value of GDP of Guangzhou city for 2016-2021 is compared with the predicted value data, and comparing the results and calculations, it is found that the predicted value predicted by ARIMA model (2, 2, 1) has a small error compared with the real value, and none of the relative errors exceed 0.05.

From Figure 6, it can be seen that the trend of the predicted values of Guangzhou city from 1990-2015 is basically consistent with the true values, which reflects the good fit of the model and indicates that the prediction results of the ARIMA (2, 2, 1) model on this series of data are credible to some extent.

The prediction results for the GDP of Guangzhou City in 2022-2024 are shown in Table 4, which shows that the GDP of Guangzhou City still maintains its growth in the next three years.

Table 3. Comparison of real and predicted GDP data in Guangzhou (unit: 100 million yuan)

Years	2016	2017	2018	2019	2020	2021
Real value	18559.73	19871.67	21002.44	23844.69	25068.75	28231.97
Forecast value	19107.24	20153.94	21903.91	23457.36	24846.43	26814.83
Relative error	0.03	0.01	0.04	0.02	0.01	0.05

Table 4. Model statistics table (unit: 100 million yuan)

Years	Forecast value	LCL	UCL
2022	28280.92	25202.34	31359.49
2023	30124.50	26494.38	33754.62
2024	3,2048.07	2,7756.96	3,6339.17

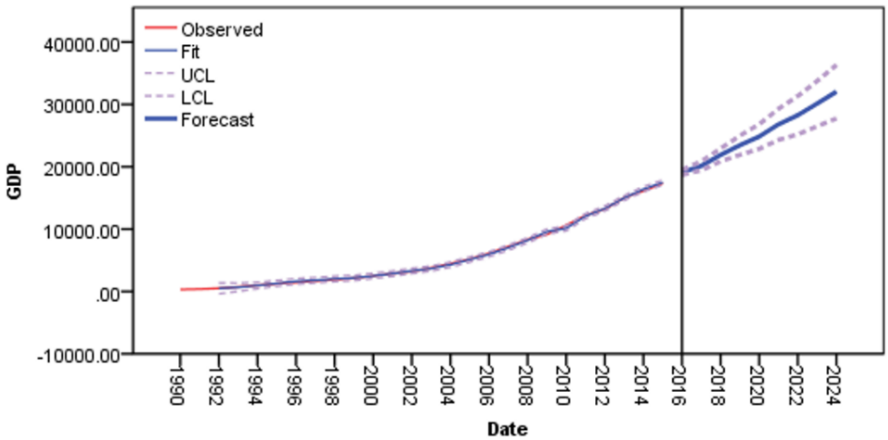

Fig. 6. Model Fitting and Prediction Results

4 Conclusions

This paper selects the annual GDP data of Guangzhou City from 1990 to 2021, conducts a stationary test and processing on the data, and obtains a stationary sequence after the second-order difference. According to the ACF and PACF dia-grams of the sequence after the second-order difference, the ARIMA model is constructed. The three parameters of p, d, and q, by constructing an ARIMA (2, 2, 1) model, predict the annual GDP data of Guangzhou from 2016 to 2024, and use the GDP data of Guangzhou from 2016 to 2021 as observations , to test whether the accuracy of the established model is accurate, whether the model prediction data is credible, and then predict the GDP of Guangzhou City in the future two or three years backwards.

The results indicate that the relative error between the real value and the predicted value is not more than 5% under the prediction of the model, which shows that the future trend of Guangzhou's GDP can be well predicted. Three years later, the GDP of Guangzhou is expected to break through the 3 trillion yuan mark in 2023. The prediction model shows that Guangzhou's GDP will still show a growth trend in the future, which can provide suggestions and scientific references for the future development planning of Guangzhou.

Acknowledgment

This paper here was supported by National Key Research and Development Program (Grant No.2022YFC3303204).

References

1. Hosny, S., Elsaid, E., Hosny, H., "Prediction of construction material prices using ARIMA and multiple regression models," Asian Journal of Civil Engineering, 24, 1697 - 1710 (2023).
2. Zhang, Z., "Sales Prediction Based on ARIMA Time Series and Multifactorial Linear Model," Highlights in Science, Engineering and Technology, 38, 1–8 (2023).
3. Liang, L., Chen, Y., "Quarterly GDP forecast based on deep learning neural network," Statistics and Decision, 39(02), 24-29 (2023).
4. Dong, Z., Gao, X., "The prediction of GDP growth in Shandong Province with gray model," Academic Journal of Business & Management, 3(10), 87-91 (2020).
5. Wabomba, M. S., Mutwiri, M., & Fredrick, M., "Modeling and forecasting Kenyan GDP using autoregressive integrated moving average (ARIMA) models," Science Journal of Applied Mathematics and Statistics, 4(2), 64-73 (2016).
6. Box, G., "Box and Jenkins: Time Series Analysis, Forecasting and Control," A Very British Affair, 161-215 (2013).
7. Ho, S. L., Xie, M., & Goh, T. N., "A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction," Computers & Industrial Engineering, 42(2-4), 371-375 (2002).