Analysis and Prediction of Shanghai's GDP Based on ARFIMA and ARIMA Models

Jingyun Xu, Peng Yuan*

School of Applied Science and Civil Engineering, Beijing Institute of Technology, Zhuhai, Guangdong, 519088, China

1401172434@qq.com, *yuan_peng@bitzh.edu.cn

Abstract. This article takes Shanghai's GDP as the research object and predicts and analyzes the GDP of Shanghai for the next ten years by comparing two time series models: ARFIMA and ARIMA. Through a comparison of evaluation metrics, it is revealed that there are certain differences between these two models in estimating future GDP. The ARFIMA model performs better than the ARIMA model in terms of AIC value, while the ARIMA model has certain advantages in RMSE.

Keywords: GDP; ARFIMA Model; ARIMA Model.

1 Introduction

GDP, or Gross Domestic Product, serves as a comprehensive indicator of a country or region's economic performance. It reflects the market value of all final goods and services produced within a certain period of time by utilizing production factors in a country or region. Through a thorough analysis of Shanghai's GDP, we can gain a more comprehensive understanding of its economic dynamics, providing profound insights into the trajectory of China's overall economic growth.

In past studies, scholars have made extensive efforts to delve deeply into GDP research. For instance, Li Zhenliang et al. [1] selected GDP data from Beijing from 1978 to 2020, constructed an optimal ARIMA(2,2,1) model based on the AIC criterion, and predicted that Beijing's GDP would maintain a high growth rate over the next five years, providing a scientific reference for future economic development in Beijing. Yang Ghazo A [2] suggested that modeling and predicting GDP and CPI in Jordan. This study applied the Box- Jenkins methodology for the period 1976-2019, ARIMA (3,1,1) found to be the best model for the GDP, ARIMA (1,1,0) was the best model for forecasting the CPI. The results were supported with the findings of the stationarity and identification rules test of time series under using AIC and SIC criterion.

The ARFIMA model introduces fractional differencing, which differs from the traditional ARIMA model. The Autoregressive Fractionally Integrated Moving Average model is more suitable for describing time series with long-term dependence, making it more powerful in handling non-stationary and nonlinear time series data. Belbute M

Z. Zhan et al. (eds.), *Proceedings of the 2024 10th International Conference on Humanities and Social Science Research (ICHSSR 2024)*, Advances in Social Science, Education and Humanities Research 858,
https://doi.org/10.2991/978-2-38476-277-4_174

J et al. [3] selected worldwide CO2 emissions data from fuel fossil combustion and cement production, ultimately created an ARFIMA model, and used this model to predict CO2 emissions. Finally, suggestions were made to reduce emissions by 57.4% and 97.4% of 2010 emissions by 2030 and 2050, respectively. Safitri D et al. [4] established an ARFIMA gold price prediction model considering long-term memory, and concluded that the ARFIMA model has a different parameter value (d) of integer while the value of d=1,05716 as the best model. Wang Zhenhuan et al. [5] used data on the total output value of agriculture, forestry, animal husbandry, and fisheries in Inner Mongolia from 1947 to 2011. They employed the fractional differencing method instead of the traditional integer differencing method and used the ARFIMA model for model fitting. The results showed that the model's fitting results were relatively accurate, with errors within a controllable range.

Therefore, this article aims to conduct a thorough study of Shanghai's GDP using time series analysis models, namely the Autoregressive Fractionally Integrated Moving Average model (ARFIMA) and the Autoregressive Integrated Moving Average model (ARIMA).

2 Methods

2.1 Data Preprocessing

By referring to the Shanghai Statistical Yearbook, this article selects the GDP data of Shanghai from 1978 to 2023, totaling 46 observations. A time series graph is plotted for the GDP data of Shanghai from 1978 to 2023 (Fig.1).

From the graph, it can be initially identified that the time series data exhibits a non-stationary trend. Due to significant differences in the magnitude of the data within the time range from 1978 to 2023, the time series exhibits an exponential trend. To narrow the magnitude differences between these values, this article adopts the method of taking logarithms to transform the time series, making it exhibit a linear trend. After taking the natural logarithm of the original GDP data, a time series graph of the logarithmic sequence is obtained (Fig.1). Through observation of the graph, it can be found that the time series graph after logarithmic transformation exhibits a linearly increasing trend, indicating that even after the logarithmic transformation, the time series remains non-stationary. To address this non-stationarity issue, a differencing operation needs to be performed on the sequence. Therefore, selecting an appropriate order of differencing to perform the differencing operation is crucial to ensure the stationarity of the sequence, which is essential for subsequent model construction.

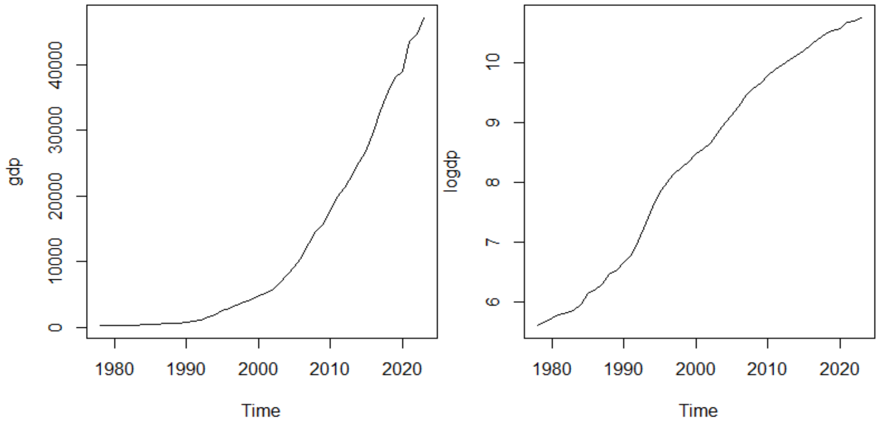

Fig. 1. GDP Time Series Chart & Logarithmic GDP Time Series Chart

2.2 Analysis and Prediction of Shanghai's GDP Based on the ARFIMA Model

The modeling steps of the ARFIMA model include:

(1) Long-memory test: The rescaled range method is used to test the long-memory property of the sequence and obtain the Hurst exponent value. The Hurst exponent is then utilized to judge the long-memory property of the time series: when ($0 \le H < 0.5$), the time series exhibits anti-persistent Brownian motion, also known as "mean reversion"; when (H = 0.5), the time series is a standard random walk; when (0.5 < H < 1), the time series possesses long-term memory; and when (H = 1), the time series is a straight line, and the future can be fully predicted using the present.

(2) Stationarization of the time series: If the sequence exhibits long-memory properties, fractional differencing is applied for stationarization. The stationarity of the time series after differencing is judged based on the differenced time series plot and the ADF test.

(3) Model identification and order determination: The process of model order determination involves finding the relatively optimal combination of orders. There may be multiple significantly effective models, and the optimal values of parameters p and q can be selected using criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

(4) Parameter estimation: Maximum likelihood estimation is used to estimate and conduct significance tests on the model parameters.

(5) Model validation: This includes parameter significance tests and model significance tests. The former is addressed in the parameter estimation process, while the latter involves residual white noise testing after model fitting. If the tests pass, the model exhibits good fitting performance.

(6) Model evaluation: Multiple metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) can be selected to analyze the model, and a prediction effect chart can be plotted.

In the ARFIMA model, there is a relationship between the fractional differencing order d and the Hurst exponent. Using Python, we calculate the Hurst exponent to be 0.841. Since this value falls within the range of ($0.5 < H < 1$), the time series possesses long-term memory. When it is determined that the sequence exhibits long-memory properties, there exists a relationship between Hurst and the fractional differencing order d.

$$d=H-0.5 \tag{1}$$

Furthermore, the fractional differencing order d is calculated to be 0.341. Through model identification and order determination, the optimal ARFIMA model selected is ARFIMA(1, 0.341, 0).

At this point, the AIC value is -241.96. AIC is an information criterion used to compare the relative fitting effects of different models. A smaller AIC value indicates a better fitting effect of the model. In this case, the negative AIC value suggests that this ARFIMA model performs relatively well in fitting the data.

Using R language programming to predict Shanghai's GDP for the next five years, the results are presented in the following Table 1:

Table 1. GDP Prediction Based on ARFIMA Model

Year	GDP prediction
2024	49746.22
2025	52334.86
2026	54980.01
2027	57677.24
2028	60423.42

2.3 Analysis and Prediction of Shanghai's GDP Based on the ARIMA Model

In the ARIMA model, the non-stationary sequence obtained from the logarithmic transformation in the data preprocessing section is still used for further research. Since the GDP time series after logarithmic transformation has already been obtained in the data preprocessing section, it visually exhibits a linear trend. Therefore, a first-order difference is applied. The ADF test results for the data after the first-order difference show a P-value < 0.05, indicating that the sequence is stationary. As a result, the order d of the ARIMA model can be determined as 1. A white noise test conducted on the sequence after the first-order difference reveals a P-value much smaller than 0.05, indicating that the sequence is a stationary and non-white noise sequence. Therefore, it can be used for modeling in the ARIMA framework.

Model identification primarily relies on the analysis of the autocorrelation plot and the partial autocorrelation plot. Using R language, we draw the autocorrelation plot and the partial autocorrelation plot separately. The autocorrelation plot exhibits a trailing-off phenomenon, while the partial autocorrelation plot shows a first-order cutoff. Subsequently, through model parameter testing, the AIC value is obtained as -134.37. Based on the above, the ARIMA(1, 1, 0) model is used to fit the time series data of

Shanghai's GDP. In addition, the P-value of the white noise test is much greater than 0.05, indicating that the residuals of the model are a white noise sequence. This suggests that the fitted ARIMA(1, 1, 0) model is effective.

Finally, using the ARIMA model, the GDP of Shanghai for the five-year period from 2024 to 2028 is predicted. The prediction results are presented in the following Table 2:

Table 2. GDP Prediction Based on ARIMA Model

Year	GDP prediction
2024	49671.67
2025	52005.25
2026	54215.38
2027	56300.01
2028	58258.77

2.4 Model Evaluation and Comparison

To further compare the two models, their evaluation metrics are presented in the following Table 3:

Table 3. Evaluation of the model

Model	AIC	RMSE
ARFIMA	-241.96	0.05893
ARIMA	-134.37	0.05047

The Table 3 presents the evaluation metrics for the two models, ARFIMA and ARIMA, including the Akaike Information Criterion (AIC) and Root Mean Squared Error (RMSE), which are crucial for assessing the fitting ability and prediction accuracy of the models.

AIC is a widely used statistical indicator for model comparison, which considers the trade-off between the model's fitting effect and the number of parameters. As can be seen from the table, the AIC value for the ARIMA model is -134.37, while the AIC value for the ARFIMA model is -241.96. A smaller AIC value indicates better performance in fitting the data. Therefore, based on the AIC metric, the ARFIMA model outperforms the ARIMA model in describing the GDP time series of Shanghai.

RMSE is used to measure the prediction accuracy of the model, which is the square root of the mean squared difference between the actual observed values and the model's predicted values. The table shows that the RMSE for the ARIMA model is 0.05047, while the RMSE for the ARFIMA model is 0.05893. A smaller RMSE value indicates more accurate predictions. Therefore, based on the RMSE metric, the ARIMA model performs better in predicting the future GDP of Shanghai.

Taking into account both the AIC and RMSE metrics, although the RMSE of the ARFIMA model is larger, the AIC value of the ARFIMA model is much smaller than that of the ARIMA model. Therefore, studying the ARFIMA model holds profound significance.

3 Conclusions

Through the application and comparative analysis of the ARFIMA and ARIMA models in predicting Shanghai's GDP, this paper draws a series of key conclusions.

In terms of AIC evaluation, the ARFIMA model outperforms the ARIMA model, with a significantly lower AIC value, indicating that ARFIMA is more effective in fitting the GDP time series of Shanghai. On the other hand, the RMSE results show that the ARIMA model has an advantage in the accuracy of future GDP predictions. This suggests that each model has its own strengths and weaknesses in different aspects, and the choice of model should be made based on specific needs and objectives.

Specifically, Shanghai's GDP is expected to gradually increase in the next ten years, which is consistent with the general law of urban economic development. This trend may be influenced by various factors, including government policies, investment environment, technological innovation, and so on. Therefore, we can discuss the possibilities and potential driving factors of Shanghai's future economic development in combination with this trend. Policymakers can draw on the results of our models, gain a deeper understanding of the performance of ARFIMA and ARIMA in different aspects, and formulate economic policies more targeted. For example, if the government places more emphasis on accurate GDP predictions in the short term, it may prefer the ARIMA model with a smaller RMSE. Conversely, if the government focuses more on grasping long-term trends and structural changes, it may consider the ARFIMA model.

Overall, this paper provides useful predictions and analysis for the future development trend of Shanghai's GDP, serving as a reference for decision-makers. However, in practical applications, it is necessary to combine more actual situations and domain expertise to formulate reasonable development strategies from a more comprehensive perspective.

Acknowledgements

This work was supported by Research and Teaching Project of Beijing Institute of Technology, Zhuhai [Project Number: 2023025ZLGC].

References

1. Li Zhenliang, Le Xinyu. Analysis and Prediction of Beijing's GDP Based on ARIMA Model [J]. Small and Medium-sized Enterprise Management and Technology, 2023(01): 153-155.
2. Ghazo A. Applying the ARIMA Model to the Process of Forecasting GDP and CPI in the Jordanian Economy[J]. International Journal of Financial Research,2021,12(3).
3. Belbute M J, Pereira M A. ARFIMA Reference Forecasts for Worldwide CO2 Emissions and the Need for Large and Frontloaded Decarbonization Policies[J].Journal of Economics and Public Finance,2020,6(4).
4. Safitri D, Mustafid, Ispriyanti D, et al. Gold price modeling in Indonesia using ARFIMA method[J].Journal of Physics: Conference Series,2019,1217(1):012087 (11pp).

5. Wang Zhenhuan, Sun Pengzhe, Tao Weiqing, et al. Research on the Total Output Value of Agriculture, Forestry, Animal Husbandry, and Fisheries in Inner Mongolia Based on ARFIMA Model [J]. Journal of Inner Mongolia Normal University (Natural Science Edition in Chinese), 2014, 43(06): 696-698.