SICASH 2023
SLIIT INTERNATIONAL CONFERENCE ON ADVANCEMENTS IN SCIENCES & HUMANITIES

A Model Comparison to Forecast Gross Domestic Product (GDP) in China

*[1] Koswaththa N.B.K.S.M., [2] Gaganathara G.A.G.D., [3] Fernando A.S.M.S., [4] Dissanayake M.D.T.G., [5] Guruge M. L.

[1,2,3,4,5]Department of Mathematics and Statistics, Faculty of Humanities and Sciences,
Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
Corresponding author - hs21826986@my.sliit.lk

ARTICLE INFO

Article History:
Received: 10 September 2023
Accepted: 01 November 2023

Keywords:
GDP; ARIMA; Simple Linear Regression with AR(1) Error structure

Citation:

Koswaththa N.B.K.S.M., Gaganathara G.A.G.D., Fernando A.S.M.S., Dissanayake M.D.T.G., Guruge M. L. (2023). A Model Comparison to Forecast Gross Domestic Product (GDP) in China. Proceedings of SLIIT International Conference on Advancements in Sciences and Humanities, 1-2 December, Colombo, pages 326-331.

ABSTRACT

Gross Domestic Product (GDP) is an accurate indicator to measure the size of the economic performance of a country and its growth rate. This study focuses on finding a suitable model to forecast GDP in China, which is one of the world's largest and most rapidly developing economies. A simple linear regression model with AR(1) error structure and Autoregressive Integrated Moving Average (ARIMA) model were developed and compared for the purpose. A secondary data set which includes GDP in China from 1952 to 2020 was used for this study and the sample size was 69. Residual diagnostics tests were conducted to check the assumptions and model adequacy of each model. It was found that out of the fitted models, ARIMA (1,1,1) is the most appropriate model to forecast GDP in China as it gave lower MAE and RMSE compared to fitted simple linear regression model with AR(1) error structure. Model comparison was done using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The predicted values for 2023, 2024 and 2025 are 1436349, 1447149 and 1457950 respectively. E-views 8.0 and Minitab software were used to analyze the data.

hs21826986@my.sliit.lk

https://doi.org/10.54389/AFOY5964

1. INTRODUCTION

The gross domestic product (GDP), a vital economic indicator, measures the total value of all products and services produced within a nation's borders over a specific time. It serves as an essential measure of a nation's economic development and performance. Policymakers, economists, and academics should first evaluate and understand the factors affecting GDP for the effective understanding of an economy's dynamics. China, is one of the world's largest and fastest-growing economies. Economic growth strategy in China has altered in recent years from being investment-driven to innovation-driven which led to a medium high growth from a high growth rate. Thus, GDP in China might decrease in the future (Ali, 2023). GDP in China is expected to reach 18879.00 USD Billion by the end of 2023, according to Trading Economics global macro models and analysts' expectations. In the long-term, China GDP is projected to be around 19729.00 USD Billion in 2024 and 20755.00 USD Billion in 2025, according to econometric models (Zhang, WenJ, and Yang.2022). Surprisingly, despite the importance of the Chinese economy, there is a paucity of published academic research on forecasting Chinese macroeconomy. Accordingly various authors have developed several forecasting models for assessing GDP in China that ranges from simple smoothing techniques to complicated neutral network models (Ali, 2023). The most common models used to forecast GDP are ARIMA models. However, trend analysis with autoregressive error structure modelling has not been exposed in the past studies. The objective of this study is therefore to develop a linear trend model with AR(1) error structure and the ARIMA model and recommend the most suitable model out of those two to predict GDP in China.

1.1. SIGNIFICANCE OF THE STUDY

The findings of this study are important for decision-makers in government, economists and companies who do businesses inside China or outside China. The accurate and reliable forecasts of the GDP determine the dynamics of China's economic growth, change the economic policy, direct investment tactics and facilitate efficient resource allocation.

2. METHODOLOGY

A secondary data set containing the GDP in China from 1952 to 2020 was used for this study, extracted from the official website of National Bureau of Statistics of China (2023) . The sample size was 69 and the analysis was conducted using E-views 8.0 and Minitab software.

A simple linear regression model with AR(1) error structure (Joseph and Georage, 1988) and Box-Jenkins ARIMA modelling techniques were applied to find an appropriate model to forecast GDP data in China. A simple linear regression model with AR(1) error structure was applied to predict GDP in China based on time in years as errors in the simple linear regression model were correlated with each other. Scatter plot and Pearson's correlation coefficient were also used to identify the relationship between GDP and time in years. This model estimates the coefficient of the predictor variable and provides details on how they impact changes in GDP. Durbin Watson test statistic was used to check the significance of correlations among error terms in the fitted model. Then, model adequacy and assumptions were checked using residual diagnostics to ensure the validity of the model.

The ARIMA model is also fitted for data which combines autoregressive (AR), differencing (I), and moving average (MA) components. It accurately depicts the temporal and autocorrelational trends in the GDP data. The model parameters were estimated using maximum likelihood estimation. Diagnostic tests along with inverse roots of AR/MA polynomials were conducted to evaluate residual features and the goodness of fit. Before fitting ARIMA model, it was tested for stationary conditions using time series plot, ACF (Auto Correlation Function) plots and Augmented Dicker-Fuller (ADF) test.

The above two models were compared to identify the best model for forecasting GDP in China. To determine how effectively the built-in models work, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were utilized to measure the performance of the included models.

3. RESULTS AND DISCUSSION

3.1. DEVELOPING A SIMPLE LINEAR REGRESSION MODEL WITH AR(1) ERROR STRUCTURE

The original GDP data in China showed a non linear relationship with time in years. Thus, a natural log transformation has been applied for GDP to obtain a linear relationship in between GDP and time in years. Then, the updated data set of LGDP (Logs of Gross Domestic Product) in China showed a linear relationship between LGDP and year.

Figure 1 depicted that the relationship between LGDP in China and time in years. As in figure 1, there was a positive linear relationship between LGDP in China and time in years as the correlation between the two variables was significant (r= 0.984, p = 0.00). A significant positive linear relationship was depicted between LGDP in China and time in years (95%). Autocorrelation parameter for residuals in the fitted simple linear regression model was 0.9623. Thus, Cochrane-Orcutt procedure was applied to fit a simple linear regression model between LGDP and Year. According to the this procedure, a transformation was applied for both LGDP and Year. Both ANOVA and the parameters were significant at 5% level of significance. The fitted model was as follows:

$LGDP = -276.45 + 0.1436 * Year$ *(R^2 =67.1% , $AdjR^2$ = 65.5% , p = 0.00)*

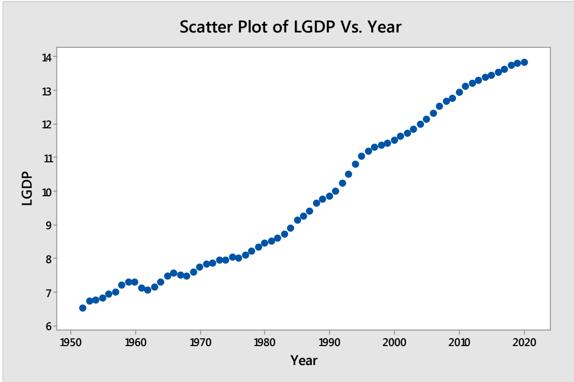


Figure1: Scatter Plot for LGDP Vs Time in Years

The significance of autocorrelations among error terms in the model was checked using the Durbin Watson test statistic (1.58) and as it was close to 2, it can be assumed that the errors are random. The Anderson- Darling (AD) test confirmed that the errors were not signifiacntly devaited from the normality (AD test statistic = 0.736, p = .736). As there was no systemetic pattern in the scatter plot between standardized residuals and fitted values (Fig. 2), it can be concluded the variance of the error terms were constant.

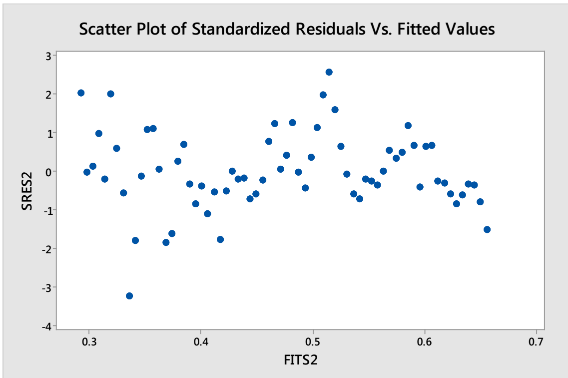


Figure 2: Scatter Plot between Standardized Residuals and Fitted Values

According to the above results, it can be concluded that all the assumptions under simple linear regression with AR(1) error structure were satisfied. Thus, fitted simple linear regression model with AR(1) error structure was appropriate to predict

GDP in China according to the year.

3.2. DEVELOPING A TIME SERIES MODEL

3.2.1. STATIONARY OF THE GDP SERIES

The time series plot of the original series of the GDP (Fig. 3) had an increasing trend after 1990 which confirms the non-stationary of the series. The non significance of the ADF test (t = 1.74, p = 0.996) and significance of many autocorrelations at different lags (Fig 4) confirmed that original series was not stationary.

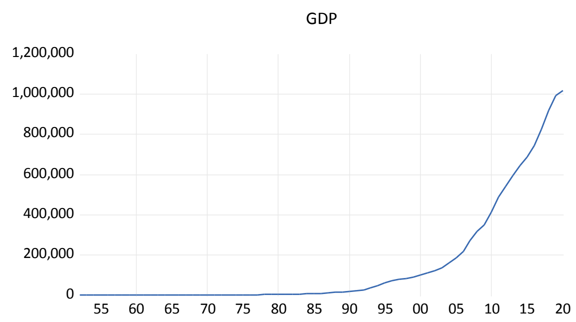
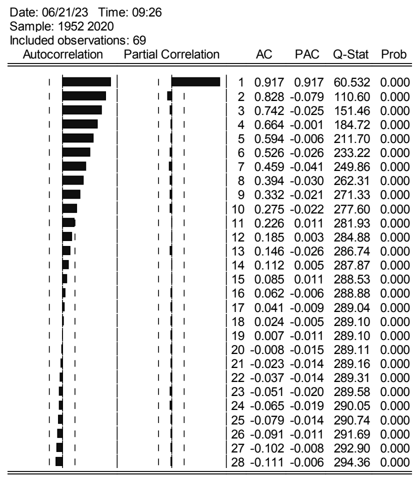


Figure 3: Time Series Plot of GDP Data from 1952 –2020 Period

As the original series was exponentially increasing to reduce the heteroscedasticity, the log transformation was considered (Fig 5). Since the log series was linearly increasing to make the series stationary, the first difference of the log series was considered.

The ADF test for the LGDP was significant (t= -4.47, p = 0.006). Thus, it can be confirmed that the LGDP series was stationary. The ACF and PACF of the stationary series were shown in Figure 6. Since only a few lags lie beyond the confidence interval in ACF (Fig. 6) , it was assumed that DLGDP series was stationary.



	AC	PAC	Q-Stat	Prob
1	0.917	0.917	60.532	0.000
2	0.828	-0.079	110.60	0.000
3	0.742	-0.025	151.46	0.000
4	0.664	-0.001	184.72	0.000
5	0.594	-0.006	211.70	0.000
6	0.526	-0.026	233.22	0.000
7	0.459	-0.041	249.86	0.000
8	0.394	-0.030	262.31	0.000
9	0.332	-0.021	271.33	0.000
10	0.275	-0.022	277.60	0.000
11	0.226	0.011	281.93	0.000
12	0.185	0.003	284.88	0.000
13	0.146	-0.026	286.74	0.000
14	0.112	0.005	287.87	0.000
15	0.085	0.011	288.53	0.000
16	0.062	-0.006	288.88	0.000
17	0.041	-0.009	289.04	0.000
18	0.024	-0.005	289.10	0.000
19	0.007	-0.011	289.10	0.000
20	-0.008	-0.015	289.11	0.000
21	-0.023	-0.014	289.16	0.000
22	-0.037	-0.014	289.31	0.000
23	-0.051	-0.020	289.58	0.000
24	-0.065	-0.019	290.05	0.000
25	-0.079	-0.014	290.74	0.000
26	-0.091	-0.011	291.69	0.000
27	-0.102	-0.008	292.90	0.000
28	-0.111	-0.006	294.36	0.000

Figure 4: Correlogram of the GDP Series

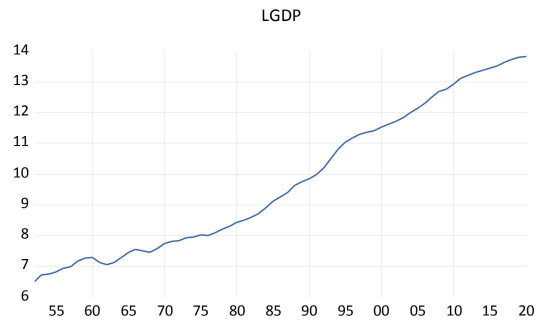
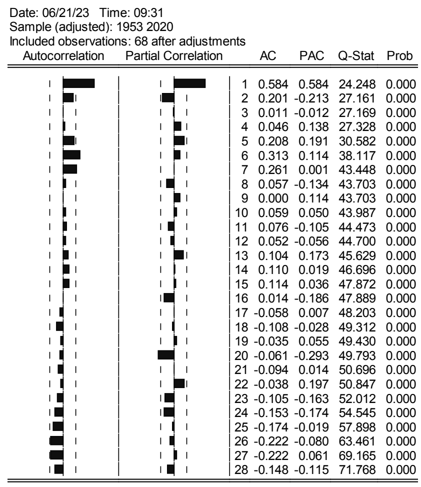


Figure 5: Time Series Plot of LGDP Data from 1952 – 2020



	AC	PAC	Q-Stat	Prob
1	0.584	0.584	24.248	0.000
2	0.201	-0.213	27.161	0.000
3	0.011	-0.012	27.169	0.000
4	0.046	0.138	27.328	0.000
5	0.208	0.191	30.582	0.000
6	0.313	0.114	38.117	0.000
7	0.261	0.001	43.448	0.000
8	0.057	-0.134	43.703	0.000
9	0.000	0.114	43.703	0.000
10	0.059	0.050	43.987	0.000
11	0.076	-0.105	44.473	0.000
12	0.052	-0.056	44.700	0.000
13	0.104	0.173	45.629	0.000
14	0.110	0.019	46.696	0.000
15	0.114	0.036	47.872	0.000
16	0.014	-0.186	47.889	0.000
17	-0.058	0.007	48.203	0.000
18	-0.108	-0.028	49.312	0.000
19	-0.035	0.055	49.430	0.000
20	-0.061	-0.293	49.793	0.000
21	-0.094	0.014	50.696	0.000
22	-0.038	0.197	50.847	0.000
23	-0.105	-0.163	52.012	0.000
24	-0.153	-0.174	54.545	0.000
25	-0.174	-0.019	57.898	0.000
26	-0.222	-0.080	63.461	0.000
27	-0.222	0.061	69.165	0.000
28	-0.148	-0.115	71.768	0.000

Figure 6: Correlogram of Residuals for DLGDP Series

3.3. MODEL IDENTIFICATION

Based on ACF and PACF plots for DLGDP data, three possible ARIMA models were identified.

They were compared with respect to Aitken Information Criteria (AIC), Schwarz Criteria (SC) and standard error on regression to select the best forecasting model (Table 1).

The lowest AIC, lowest SC, and lowest standard error was seen for the ARIMA (1, 1, 1) model and thus, ARIMA (1,1,1) model was selected as the best-fitted model to forecast LGDP in China. The properties of the ARIMA(1,1,1) were shown in Table 2.

Table 1: Comparison of three parsimonious ARIMA Models

ARIMA Model	AIC	SC	S.E on regression
ARIMA (0,1,1)	-2.583656	-2.485737	0.064877
ARIMA (1,1,0)	-2.609615	-2.511696	0.064022
ARIMA (1,1,1)	-2.631582	-2.521023	0.062836

Table 2: Properties of the coefficients for ARIMA (1, 1, 1) Model

Variable	Coef.	SE Coef.	P-Value
C	0.108007	0.017540	0.0000
AR(1)	0.415204	0.208344	0.0505
MA(1)	0.316389	0.218762	0.1530
SIGMASQ	0.003716	0.000617	0.0000

Thus, the fitted ARIMA (1,1,1) model can be written as

$$Y'_t - Y'_{t-1} = 0.1080 + 0.415204(Y'_{t-1} - Y'_{t-2}) + e_t + 0.316389e_{t-1}, \text{ where, } Y'_t = log\ (Y_t)$$

3.4. CHECKING MODEL ADEQUACY FOR ARIMA (1,1,1)

The plot of ACF of residuals of the ARIMA (1, 1, 1) model confirmed that residuals are random as all the autocorrelations are within the 95% confidence interval. The p-value of the Jarque-Bera test was less than 0.05 indicating that null hypothesis of residuals was normally distributed can be rejected at 5% level of significance. This can be due to an observation which was far away from the rest of the observations in the beginning. The plot of residuals vs predicted also confirmed that errors were having constant variance.

The stationarity of AR roots and invertible MA terms was assessed using inverse roots for the fitted ARIMA model. The stability of the model was confirmed as the roots lie inside the unit circle (Fig.7). Thus, best fitted ARIMA model is stationary and invertible and the model is appropriate to forecast GDP in China.

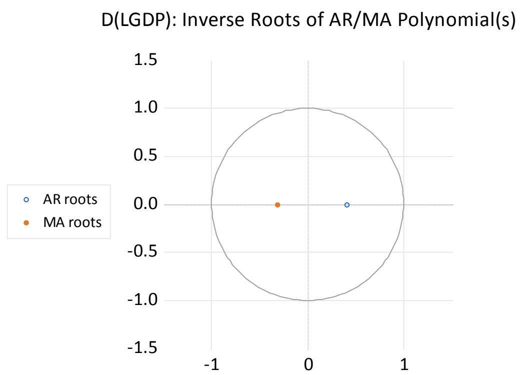


Figure 7: Inverse Roots of AR/MA Polynomials

3.5. COMPARISON OF TWO MODELS

A fitted simple linear regression model with AR(1) error structure was compared with the fitted ARIMA (1,1,1) model to identify the best model to forecast GDP in China. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were the lowest (Table 3) for ARIMA (1,1,1). Thus, the ARIMA (1,1,1) was more suitable than other one to predict GDP in China.

Table 3: Comparison of Fitted Models

Fitted Model	MAE	RMSE
SLR with AR(1)	0.0553	0.07354
ARIMA (1,1,1)	0.0445	0.0609

3.6. FORECASTING GDP USING THE ARIMA (1,1,1) MODEL

Figure 8: Actual Series, Fitted Series and Residual Series of the DLGDP Data

According to the figure 8, it was seen that the fitted values were close to the actual values of GDP, indicating that the model had a good fitting effect.

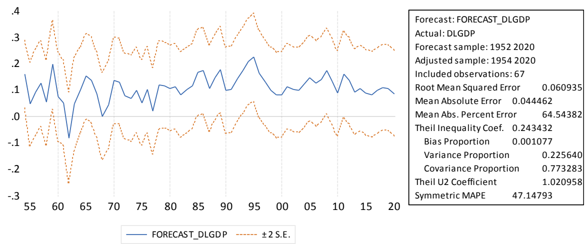


Figure 9: Forecasted DLGDP for ARIMA (1, 1, 1) Model

As in figure 9, it was seen that forecasted DLGDP lie within the confidence interval indicating that no significant deviations from the actual DLGDP. Then, the fitted ARIMA (1,1,1) model was used to forecast GDP in China for next five years (2021 – 2025) and results were given in table 4.

Table 4: Actual GDP values and Forecasted GDP values from 2021 to 2025

Year	Actual GDP value	Forecasted GDP value
2021	1775931	1414747
2022	1810004	1425548
2023	-	1436349
2024	-	1447149
2025	-	1457950

4. CONCLUSIONS

This study was conducted to identify a suitable model to forecast GDP in China. A simple linear regression model with AR(1) error structure and ARIMA model were developed and compared to identify the best fitted model to forecast GDP in China. Based on the results, it is identified that the fitted ARIMA (1,1,1) model is suitable to forecast GDP in China as it gives lower MAE and RMSE compared to fitted simple linear regression with AR(1) error structure. Goodness of fit for the model was evaluated using residual diagnostics tests and no violations were found. Therefore, the forecast values up to 2025 can be effectively used by the policy makers.

REFERENCES

Ali, F. H. A. A. (2023)., *Chinese GDP Forecast Using ARIMA Model* https://ssrn.com/abstract=4377685

China GDP (2023). *Trading Economy.* https://tradingeconomics.com/china/gdp

Economy of china (2023) *Wikipedia*. Available at: https://en.m.wikipedia.org/wiki/Economy_of_China

Joseph, A, and George, S. T. (1988). Econometric modelling of partial adjustment: the Cochrane-Orcutt procedure, flaws and remedies. *Economic Modelling, 5, (1)*, 2-8. https://www.worldbank.org/en/country/china/overview

National Bureau of Statistics of China (2023). http://www.stats.gov.cn/english/PressRelease/

Zhang J, Wen J, Yang Z (2022) *China's GDP forecasting using Long Short Term Memory Recurrent Neural Network and Hidden Markov Model. PLoS ONE 17*(6): e0269529. https://doi. org/10.1371/journal.pone.0269529