INDIANA UNIVERSITY

ECON M524 PROJECT

on

GDP OF MALAYSIA FROM 1960 TO 2021 AND PREDICTION FOR THE NEXT 5 YEARS USING ARIMA MODEL

Submitted by

Aditya Sanjay Mhaske

University ID: 2001059030

Instructor

Ke-Li Xu

Professor of Economics

Econ M524

Financial Econometrics

Master's in Data Science

Indiana University, Bloomington

CONTENTS

ABSTRACT

The Gross Domestic Product (GDP) is an essential index of a nation's economic development. This research aims to anticipate Malaysia's GDP for the next seven years and identify the variables influencing it using data from the World Bank to aid in decision-making. To anticipate the GDP from 2022 to 2026, the Autoregressive Integrated Moving Average (ARIMA) model is utilized, and multiple linear regression has been performed to investigate the components impacting GDP. The ARIMA (1, 1, 1) model is shown to be the most suitable for predicting after several forms of ARIMA (P, I, Q) testing. The model's suitability is examined using the Q-Q plot, residuals plot, PACF, and ACF graphs of the residuals. This analysis shows that Malaysia's GDP trend has been consistently increasing over time and would continue to grow in the future.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION & SIGNIFICANCE OF PROJECT

GDP measures the overall economic output of a nation over a certain period, including all goods and services. One of the most important measures of a nation's monetary operations is its GDP. The development of economic growth objectives can be greatly influenced by a methodical prediction of the indicator. Understanding the nature and trajectory of the link between a country's economic development and its processes becomes crucial as well. It may be highly beneficial in developing new ideas and initiatives for budgeting and economic policy.

1. 1 MALAYSIA GDP

Strategically located in Southeast Asia with a stable economy, Malaysia is well-positioned to be a hub for international business. The country is also home to a growing consumer base with increasing purchasing power and is one of the largest economies in the Association of Southeast Asian Nations (ASEAN). Gross domestic product (GDP) grew at its fastest pace since the second quarter of 2021 in the July-September period. Economists had expected GDP to rise 11.7%. GDP in Malaysia is expected to reach 320.00 USD Billion by the end of 2022, according to Trading Economics global macro models and analysts' expectations.

This statistic displays the contribution of various economic sectors to Malaysia's GDP from 2011 to 2021. Agriculture provided around 9.61 % of Malaysia's gross domestic product in 2021, whilst industry made up around 37.73 % and the services sector made up about 51.55 %. The third quarter of 2022 had outstanding economic growth, which suggests that Malaysia's GDP would surpass pre-pandemic growth in that year. An increase in domestic demand, driven mostly by household spending in accordance with the improvement of the labor market and income expectations, is supportive of this prediction.

Following the 1997–1998 Asian Financial Crisis, Malaysia's economy has been growing at an average rate of 5.4% since 2010, and by 2024 it is anticipated to have made the transition from an upper middle-income economy to a high-income one. According to the Allianz Global Wealth Report 2020: Wealth Immunity, Malaysia was the 35th richest nation by net financial assets per capita and the 31st richest nation by gross financial assets per capita. Gross financial assets and net financial assets both increased by 6.5% and 7.5% over the previous year in the nation.

Therefore, an effort is made in this report to forecast the GDP and identify factors affecting GDP in Malaysia. On the other hand, a significant change in GDP, whether up or down, usually has an impact on the market economy. Therefore, it is necessary to realize the associated factors that affect GDP. Keeping attention to the significance of studying the future trend of GDP and the factors affecting it, a lot of work has been done so far. Many approaches can be applied for macroeconomic estimating like linear regression, AR model, MA model, ARIMA model, VAR model, etc. The specific objectives to be investigated throughout this study are forecasting the GDP of Malaysia and finding the significant factors affecting GDP in Malaysia.

CHAPTER 2: PROBLEM STATEMENT

Malaysia GDP Prediction Using ARIMA (Autoregressive Integrated Moving Average)

1. Import Data and Pre-processing for Econometric Modeler
2. Import Time Series Variables into Econometric App
3. Perform Data Analysis
4. Fit candidate models to the data
5. Conduct goodness-of-fit checks
6. Find model best with best in-sample fit
7. Export session to matlab
8. Forecast GDP

CHAPTER 3: DATASET INFORMATION

Dataset: Gross domestic product of Malaysia in USD

Source: https://fred.stlouisfed.org/series/RGDPNAMYA666NRUG

1. Given Dataset contain data on the GDP of Malaysia from 1960 to 2021
2. Dataset size: 2 columns X 62 rows

Source: University of Groningen, University of California, Davis
Release: Penn World Table 10.0
Units: Millions of 2017 U.S. Dollars, Not Seasonally Adjusted
Frequency: Annual
Source ID: rgdpna

Table 1: Sample Dataset

Year	**GDP in USD**
1960	1916241996.60264
1961	1901868548.28172
1962	2001502678.68810
1963	2510126747.68065
1964 ……	2674441395.53116

CHAPTER 4: ARIMA MODEL

AutoRegressive Integrated Moving Average(ARIMA) is a time series forecasting model that incorporates autocorrelation measures to model temporal structures within the time series data to predict future values. ARIMA models are used because they can reduce a non-stationary series to a stationary series using a sequence of differencing steps.

ARIMA model is defined by its three order parameters, p, d, q.

p : number of Autoregressive terms in the model.

d : number of differentiations applied to the time series values.

q : number of Moving Average terms in the model.

A variety of industries employ ARIMA models. It is frequently applied to demand forecasting, such as when predicting future demand for the production of food. This is so that managers have solid rules to follow when making choices on supply chains. On the basis of previous prices, ARIMA models may also be used to forecast the future price of stocks, Growth of GDP and many more.

4.1 Advantages and Disadvantages of ARIMA Model

Advantages of using ARIMA models

1. Only requires the prior data of a time series to generalize the forecast.
2. Performs well on short-term forecasts.
3. Models non-stationary time series.

Disadvantages of using ARIMA models

1. Difficult to predict turning points.
2. There is quite a bit of subjectivity involved in determining (p,d,q) order of the model.
3. Computationally expensive,
4. Poorer performance for long-term forecasts.
5. Cannot be used for seasonal time series.

CHAPTER 5: IMPLEMENTATION

5.1. Time Series: GDP

Time Series Analysis Using ARIMA

A simple autoregressive moving average (ARMA) model is generalized into an autoregressive integrated moving average (ARIMA) model. These two models are employed to anticipate or predict upcoming time-series data items. Regression analysis in the form of ARIMA shows how strong a dependant variable is in comparison to other varying factors.

The model's ultimate goal is to forecast future time series movement by focusing on discrepancies between series values rather than actual values. In situations when the data exhibits signs of non-stationarity, ARIMA models are used. Non-stationary data are always converted into stationary data in time series analysis.

The trend and the seasonal components are the most frequent reasons why time series data are non-stationary. Applying the differencing step is how non-stationary data may be made stationary. To remove the trend component from the data, one or more times of differencing steps may be used. Similar to this, seasonal differencing might be used to eliminate the seasonal components from data.
We may break down the model into smaller parts as follows based on the name:

1. AR: an Autoregressive model that simulates a certain kind of random process. The model's output is linearly related to its prior value, or the number of lagged data points or previous observations.
2. MA: Moving average model is one in which the outcome is linearly dependent on the present and different previous observations of a stochastic variable.
3. I: In this context, the term "integrated" refers to the phase of differencing that produces stationary time series data by eliminating the seasonal and trend components.

5.1.1 Time Series Plot

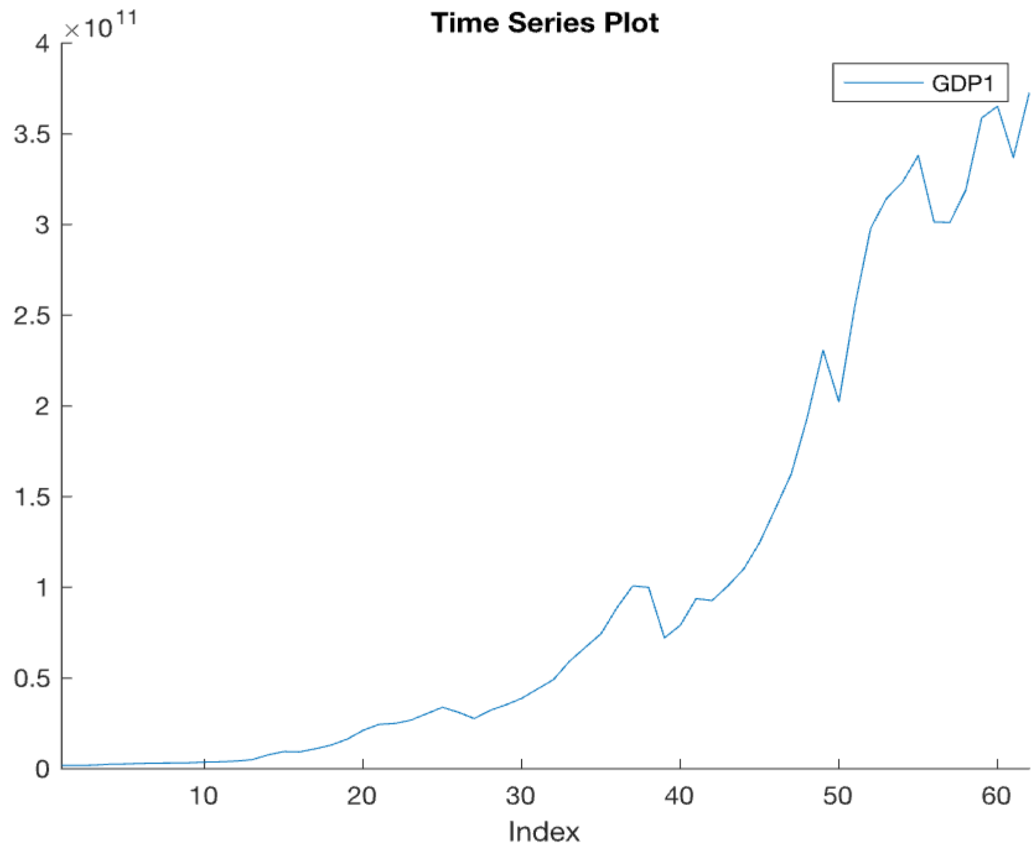

Figure 1. Time Series Plot of GDP

5.1.2 Time Series: GDPLog

Time Series Plot

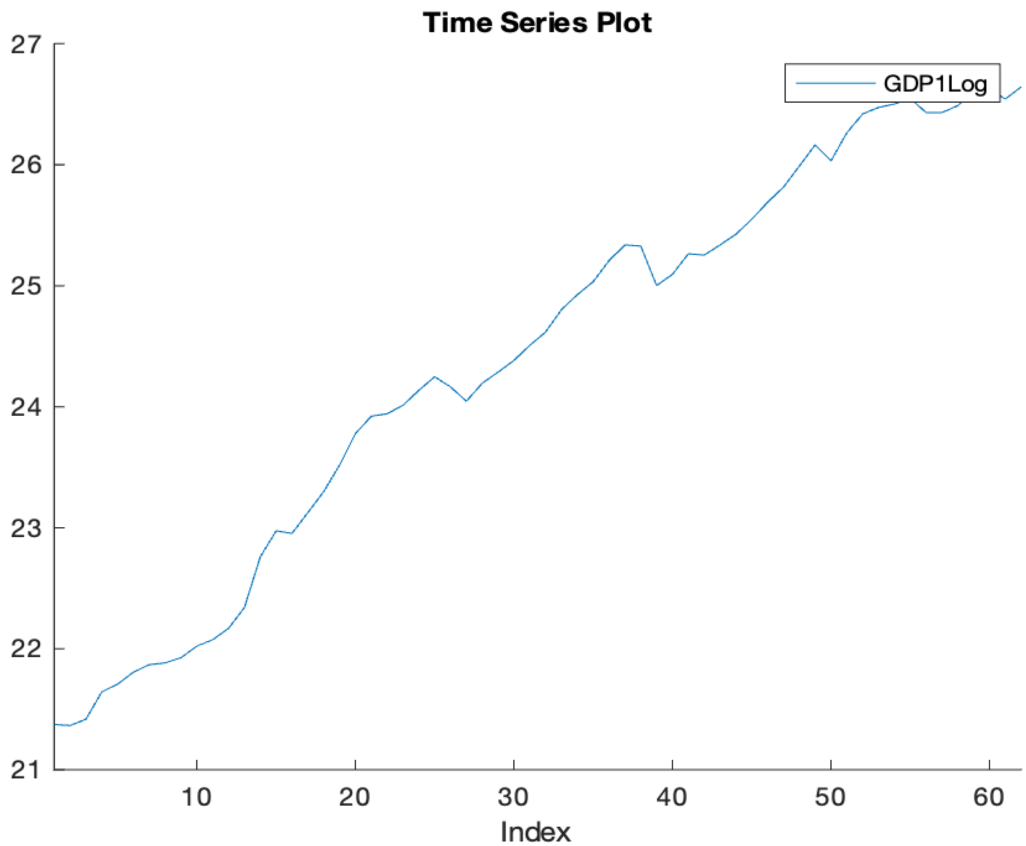

Figure 2. Time Series Plot of GDPLog

5.1.3 Time Series: GDPLogDiff

Time series GDPLogDiff is the first-order difference of time series GDPLog.

Time Series Plot

Figure 3. Time series GDPLogDiff is the first-order difference of time series GDPLog.

5.2. Sample Autocorrelation Function

In contrast to the ACF, autocorrelation measures how linked a time series is with its prior values. A time series' average relationship between data points and the data points before it is determined by the autocorrelation function (ACF).

5.2.1 Analysis of a sample ACF and PACF plot

- ACF: The autocorrelation coefficient function, describes the relationships between the data points in a time series and their forerunners.
- PACF: Partial Autocorrelation Coefficient Function provides crucial details on the dependency structure of a stationary process.

Matching the sample ACF and sample PACF of the data with the ACF and PACF of the model, respectively, is one method for fitting a model to the data given observations from a time series. The sample ACF plot may be used to determine whether or not the given time series data is stationary.

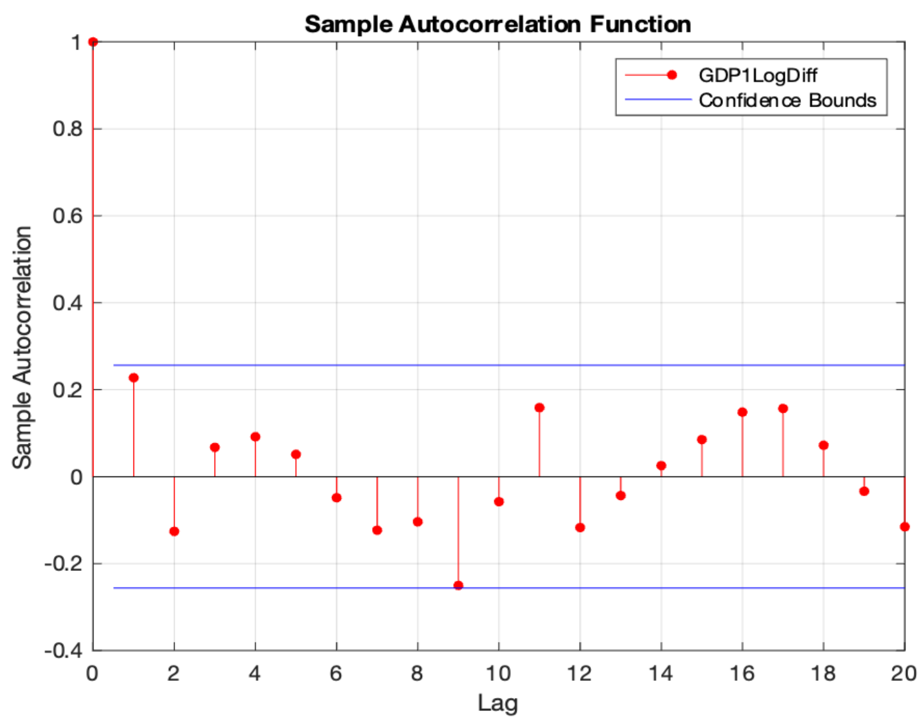

Figure 4. Sample autocorrelation function of GDPLogDiff

5.3. Augmented Dickey-Fuller Test

The ADF test expands the Dickey-Fuller test equation to include high order regressive process in the_model.

$$y_t = c + \beta t + \alpha y_{t-1} + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} .. + \phi_p \Delta Y_{t-p} + e_t$$

The p-value should be smaller than the significance threshold (let's say 0.05) in order to reject the null hypothesis since the null hypothesis implies the presence of unit root, which is =1. resulting in the conclusion that the series is stationary.

5.3.1 Null Hypothesis: GDPLogDiff contains a unit root

With just time series data, ARIMA models may be employed quickly and reliably for short-term forecasting, but it might require some trial and error to discover the best combination of parameters fr each application.

Table 2. Augmented Dickey-Fuller Test Parameters

	Lags	Model	Test Statistic	Significance Level
1	0	AR	t1	0.05

Table 3. Augmented Dickey-Fuller Test Results

	Null Rejected	P-Value	Test Statistic	Critical Value
1	true	0.001	-4.2638	-1.9458

5.4. ARIMA(1,1,1) Model (Gaussian Distribution) (ARIMA_GDP1)

ARIMA (1,1,1), a model with one AR (autoregression) term and one MA(egression error) term is being applied to the variable $Z\ t = X\ t - X\ t - 1$. A first difference might be used to account for a linear trend in the data. The differencing order refers to successive first differences.

Autoregressive integrated moving average model of time series GDP1 with the following equation:

AIC and BIC

As for other regression processes, (AIC) and or (BIC), can be used for this purpose. Generally, the process with the lower AIC or BIC value should be selected.

we simply use AIC as a way to compare regression models. The model with the lowest AIC offers the best fit.

5.4.1 Model Estimation

Maximum likelihood estimation is one common parameter estimation method for the time series ARMA(p,q) model. This method is relatively complex with large estimation errors in high-order cases. Graupe et.al proposed to fit the ARMA model into a higher-order AR model by the least square method.

Table 4. ARIMA Estimation Results

Parameter	Value	Standard Error	t Statistic	P-Value
Constant	6078444538.0928	8.4143e-11	7.223917981610452e+19	0
AR{1}	-0.036876	0.32106	-0.11486	0.90856
MA{1}	0.13443	0.31717	0.42386	0.67167
Variance	2.343466697326245e+20	6.5992e-22	3.5511425364656e+41	0

5.4.2 Model Fit

Arima Model Fit:

1. Make data stationarity by differencing the data (if required)
2. Determine AR and MA lags via model selection.
3. Estimate the parameters (fit the model)
4. Assess the residuals for problems.

The coefficient is statistically significant if the p-value is less than or equal to the significance threshold. You cannot draw the conclusion that the coefficient is statistically significant if the p-value is higher than the significance level. You may wish to refit the model without the word.

Table 5. ARIMA Goodness of Fit

AIC	3091.4495
BIC	3099.8268

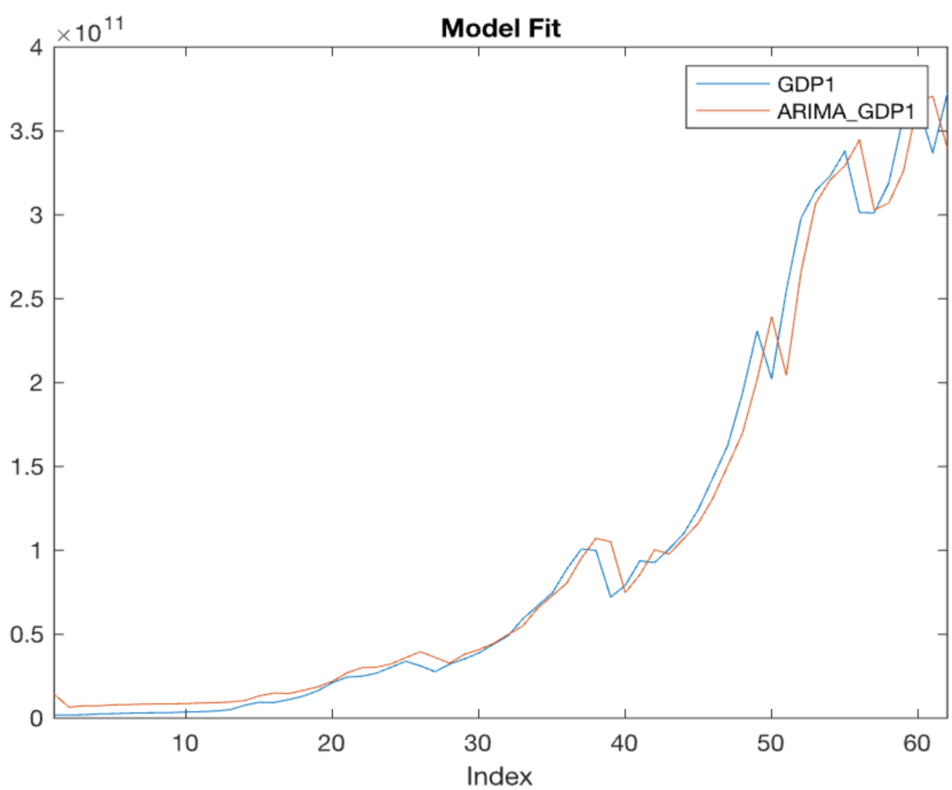

Figure 5. Plot the fit of model ARIMA_GDP time series GDP

4.5.3 Residual Plot:

For regression models with ARIMA errors, there exist regression residuals, which are the original data less the influence of the regression variables. The errors will be the same as the original series if there are no regression variables (possibly adjusted to have zero mean).

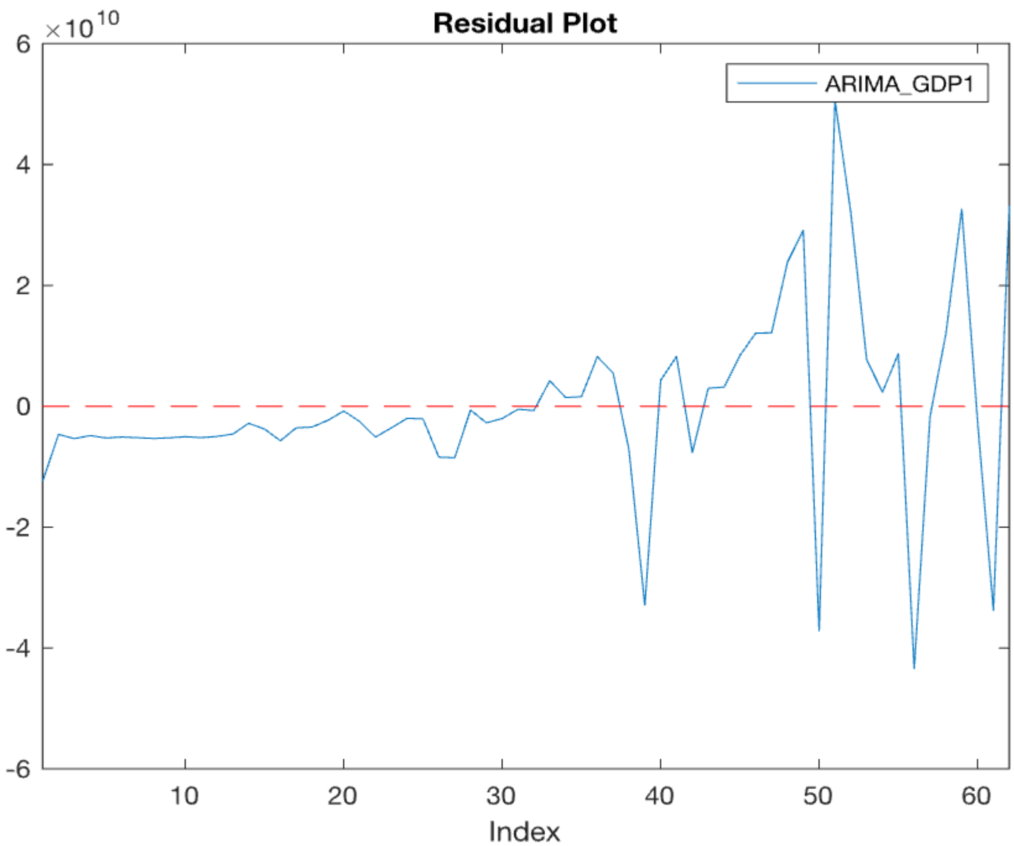

Figure 6. Plot of the residuals of model ARIMA_GDP

5.4.2 Residual Histogram

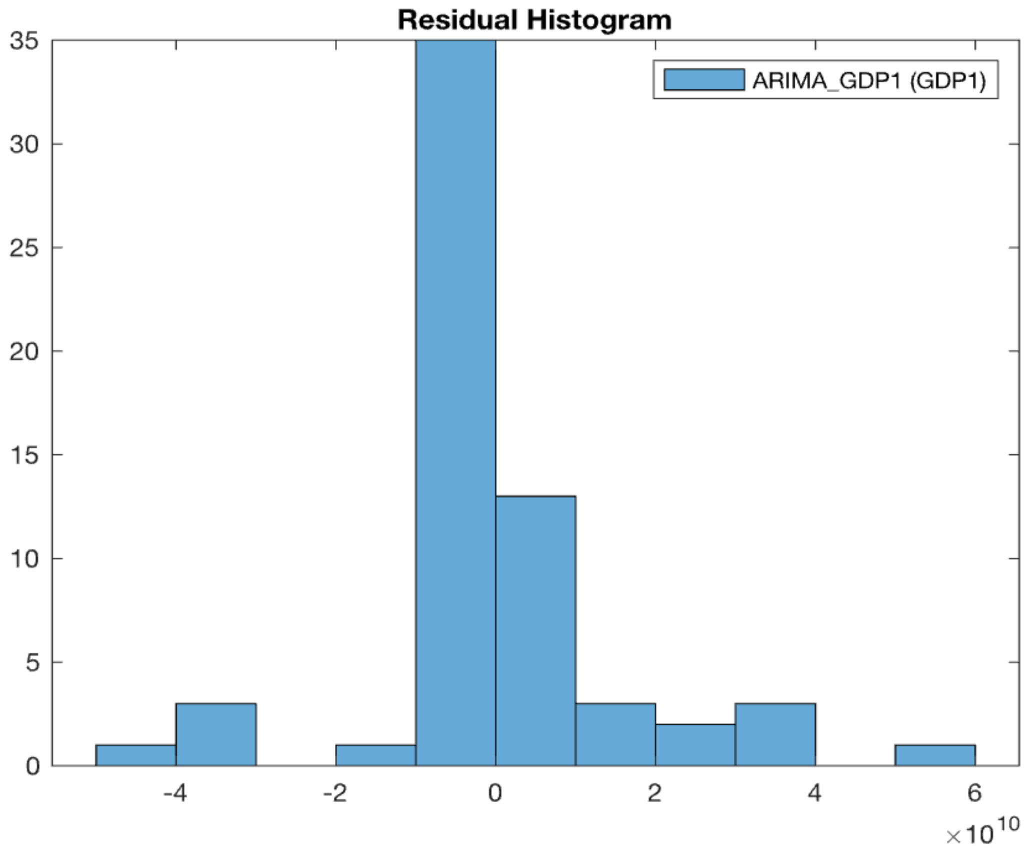

Figure 7. Histogram of the residuals of model ARIMA_GDP

6.4.3 Residual Quantile-Quantile plot

Residual plots and Q-Q plots are used to visually check that your data meets the homoscedasticity and normality assumptions of linear regression. A residual plot lets you see if your data appears homoscedastic.

The kind of distribution for a random variable may be determined using Q-Q plots, including Gaussian, Uniform, Exponential, Pareto, and other distributions. Simply by glancing at the plot, you may determine the kind of distribution using the Q-Q plot's efficiency.

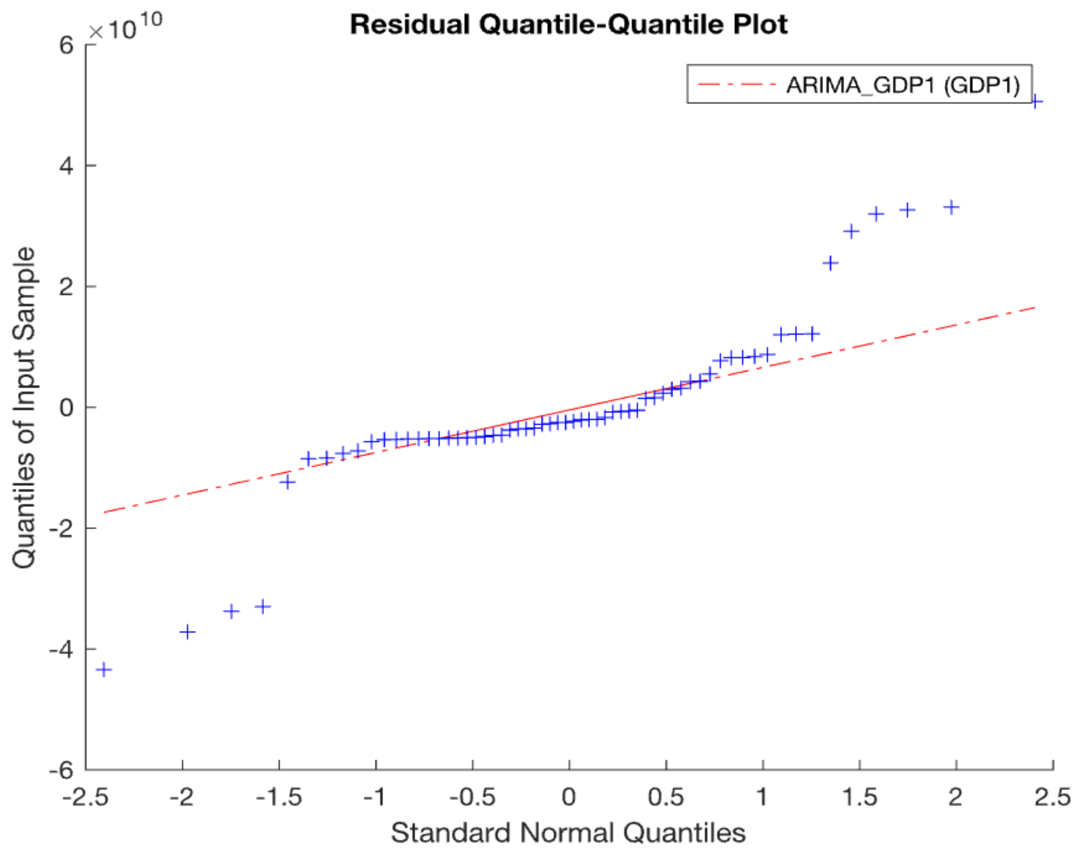

Figure 8. Quantile-quantile plot of the residuals of model ARIMA_GDP.

CHAPTER 6: RESULT AND ANALYSIS

Prediction for next five years:

- 2021 - 372701358820.26
- 2022 - 381915204837.51
- 2023 - 387653879327.06
- 2024 - 393520784340.99
- 2025 - 399382883672.53
- 2026 - 485245877268.46

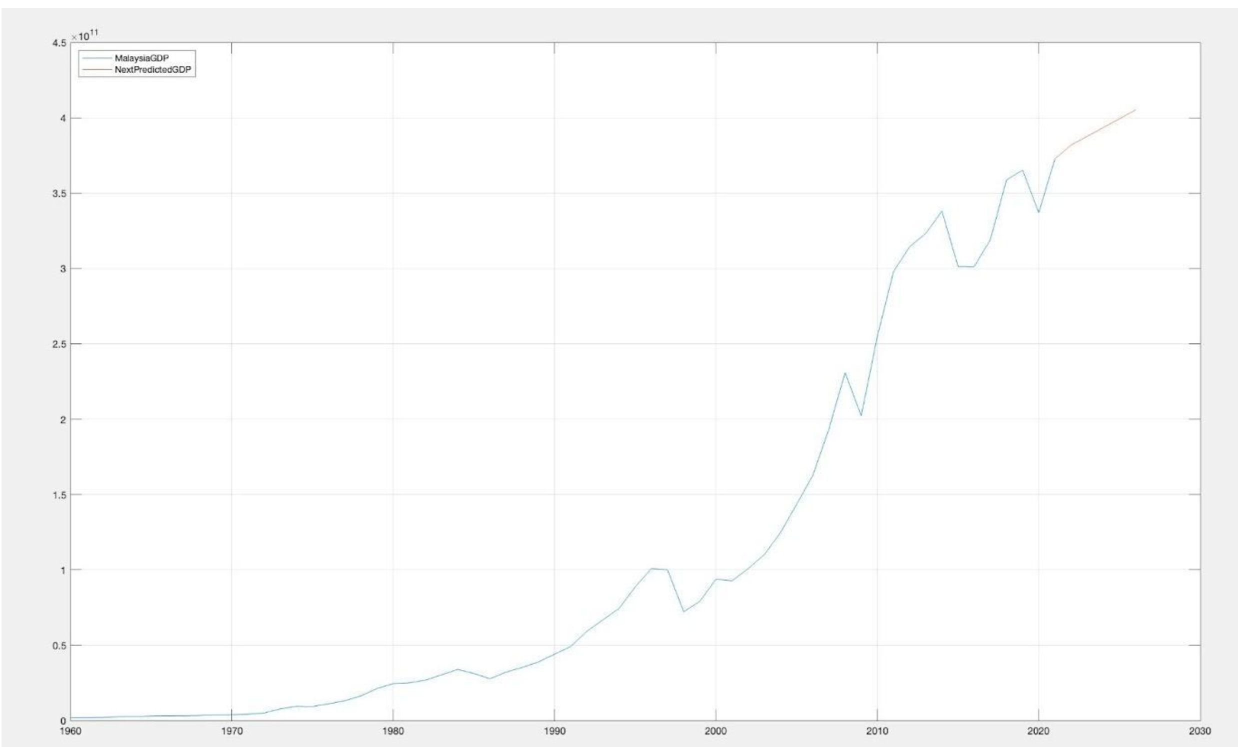

Figure 9. Prediction of GDP for the next 5 Years.

CHAPTER 7: CONCLUSION

- The ARIMA methodology is a statistical method for analyzing and building a forecasting model which best represents a time series by modeling the correlations in the data.
- ARIMA models only need the historical data of a time series to generalize the forecast and manage to increase prediction accuracy.
- Using ARIMA models has a number of possible drawbacks despite being frugal. Most essential of these derives from the subjectivity involved in determining p and q factors. Although autocorrelation and partial autocorrelations are employed, the model developer's expertise and experience will choose which p and q to use. Additionally, ARIMA models are more sophisticated and have poorer explanatory power than straightforward exponential smoothing and the Holt Winters technique.
- With just time series data, ARIMA models may be employed quickly and reliably for short-term forecasting, but it might require some trial and error to discover the best combination of parameters for each application.

CHAPTER 8: CODE

ARIMA Model

1. Data Preparation

Malaysia GDP Extraction

```
format bank
ConnectionFred = fred;
MalaysiaGDPInformation =
fetch(ConnectionFred,'MKTGDPMYA646NWDB');
[year,~,~,~,~,~]=datevec(MalaysiaGDPInformation.Data(:,1));
GDP = MalaysiaGDPInformation.Data(:,2);
MalaysiaGDP = table(year,GDP);
```

Checking for missing values

```
Check1 = any(any(ismissing(GDP)));
if Check1
   fillmissing(GDP,'previous')
end
```

Econometrics MATLAB

2. Perform Exploratory Data Analysis

   To transform Data to Stationary Data:

   A. Perform Log of GDP
   B. Perform First Order of Differences
   C. Perform Augmented Dickey-Fuller Test
   D. Auto-Correlogram

3. Fitting Model to the Data
4. Analyze the model: Residul Plot, Residual Histogram, and Residual QQ Plot

5. AIC & BIC of Different Settings of Models

```
load ModelComparison.mat
[~,iter]=min(ModelComparison.AIC);
disp(ModelComparison(iter,1));
```

6. Forecasting Data

```
load ARIMAModel.mat
ARIMA_GDP_Prediction_Next5Years =
forecast(ARIMA_GDP,5,'Y0',GDP);
Add5Years = year(end):1:year(end)+5
```

Output:

```
Add5Years =

       2021.00        2022.00        2023.00        2024.00        2025.00        2026.00
```

Figure 10: Output of addition

```
ARIMA_GDP_Prediction_Next5Years =
[GDP(end);ARIMA_GDP_Prediction_Next5Years]
```

Output:

```
ARIMA_GDP_Prediction_Next5Years =

 372701358820.26
 381915204837.51
 387653879327.06
 393520704340.99
 399382803672.53
 405245077268.46
```

Figure 11. Output of Prediction function.

7. Graph plot - Prediction for next 5 years

```
figure
plot(year,GDP);
hold on
grid on
plot(Add5Years,ARIMA_GDP_Prediction_Next5Years);
legend('MalaysiaGDP','NextPredictedGDP','Location','northwest')
```

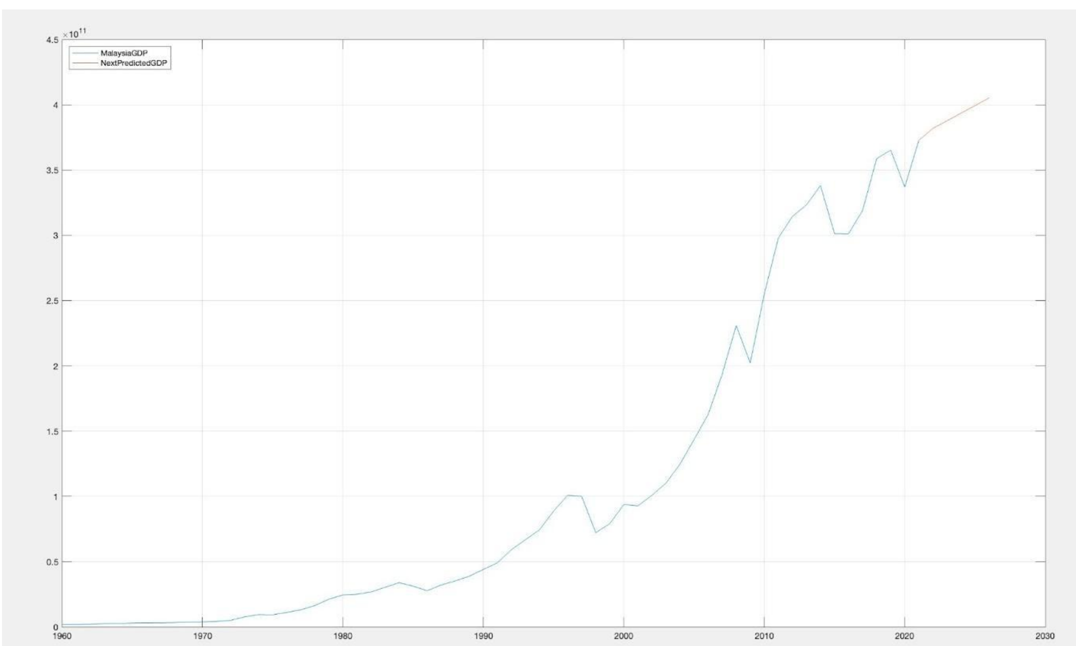

Figure 12. The output of graph GDP 1960 to 2026

REFERENCES

[1] University of Groningen and University of California, Davis, Real GDP at Constant National Prices for Malaysia [RGDPNAMYA666NRUG], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/RGDPNAMYA666NRUG, December 11, 2022.

[2] Ho, S.L. and Xie, M., 1998. The use of ARIMA models for reliability forecasting and analysis. Computers & industrial engineering, 35(1-2), pp.213-216..

[3]https://www.freemalaysiatoday.com/category/leisure/money/2022/02/17/malaysias-best-and-worst-performing-sectors-in-2021/

[4] Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, pp.159-175.

[5] Wabomba, M.S., Mutwiri, M. and Fredrick, M., 2016. Modeling and forecasting Kenyan GDP using autoregressive integrated moving average (ARIMA) models. Science Journal of Applied Mathematics and Statistics, 4(2), pp.64-73.

[6] https://www.britannica.com/place/Malaysia/Economy

[7] Economy of Malaysia - Wikipediahttps://en.wikipedia.org › wiki › Economy_of_Malaysia (https://en.wikipedia.org/wiki/Economy_of_Malaysia)