Mathematical Approach for Predicting the Gross Domestic Product of Malaysia

Maryam Rukayyah Al-Munirah Ayob¹, Azizah Mohd Rohni¹
¹School of Quantitative Sciences, Universiti Utara Malaysia (UUM)

Correspondence: Maryam Rukayyah Al-Munirah Ayob (yamira.ayob@gmail.com)

Abstract

Gross domestic product (GDP) is a monetary measure of the market value of overall final goods and services produced in a given year, and serves as a gauge of the economy’s overall health and size. The GDP prediction is significant, as it can capture and understand the future developments of a country’s economy. In this paper, three different mathematical models have been used to predict Malaysia’s gross domestic product using regressions. The models discussed in this paper are linear, exponential and parabolic regressions. In developing the models, data from year 1970 to 2014 has been employed and data from year 2015 to 2019 has been used to examine the models' accuracy. The models are then observed to identify the most appropriate to express the relationship between the years and Malaysia’s gross domestic product. In this study, it is found that the parabolic regression model is more accurate compared to the linear and exponential regression models. The parabolic regression model is also the most appropriate since it is adjusted to the real conditions of Malaysia's gross domestic product which is the main subject of this paper. Finally, it is obtained that the prediction values of GDP in Malaysia will increase for the next ten years (2020 - 2029).

Keywords: gross domestic product, linear regression, exponential regression, parabolic regression

Introduction

Gross domestic product (GDP) is a monetary measure of the market value of overall final goods and services produced in a period of quarterly or annually (Montemayor et al., 2018). According to Fagan (2019), economists and policymakers depend on statistics of gross domestic product to appraise the well-being of economy and to decide an informed decision. Delay of the monetary policy decisions can affect the economy, thus monetary policy authorities need to be looking forward earlier, as they should predict what is likely to be happened in the future. The GDP prediction is a crucial issue, as it can capture and understand the future developments of an economy. Hence, it is significant to assess the economic trends and predict the current economic conditions on the future GDP trend.

In literature, there are many existing methods that have been used under the forecasting techniques for the prediction, such as moving average, trend and seasonal decomposition, exponential smoothing and simple regression (Armstrong & Green, 2018). There are a lot of studies were carried out by researchers to predict the GDP of a country using times series forecasting techniques. However, there were several difficulties arose when performing the time series forecasting techniques, and the accuracy of forecasting models is an uncertainty (Makridakis & Bakas, 2016). This indicates that the accuracy and uncertainty in predictions using forecasting techniques varies considerably depending on
the time horizon of predictions, such as immediate, short, medium and long term (Christensen, Gillingham & Nordhaus, 2018).

It is also still ambiguous to choose the forecasting techniques that can provide a satisfactory accuracy for the gross domestic product. Therefore, this study is conducted to examine mathematically the models of economic movement on gross domestic product (GDP) in Malaysia, since it can provide information and impact in the changing conditions that represent the real problem situations and helps the economist to take decision faster and more accurate. This study is conducted to fulfil the objectives; 1) To construct mathematically the models of economic movement on gross domestic product (GDP) in Malaysia using linear, exponential and parabolic regression models, 2) To examine the accuracy of the three different models, 3) To determine the best mathematical model for gross domestic product in Malaysia, and 4) To compute the prediction values for gross domestic product in Malaysia.

According to literature, Colin Clark was one of the most important modern pioneers of gross domestic product (Lepenies & Gaines, 2016). He formulated many of what are still today’s standard elements in the compilation of GDP. Recently, Malaysia has recorded a remarkable economic growth, as the real GDP grew at an average rate of 6.1% per year over 1970 to 2018 period, higher than the average growth of the advanced countries (OECD Economic Survey, 2019). However, further progress is needed so that reforms aimed at enhancing the investment climate and ease of doing business can lead to enhance economic growth and increase its GDP. A significant change in GDP, whether up or down, usually has a significant effect on market, and investors always worry about negative GDP growth, because it is one of the factors economists use to determine, whether an economy is in a recession.

Mathematical modelling in gross domestic product has been an important feature that often used in applied situation such as predicting the future outcomes (Kumar & Rai, 2015). Among of the pioneers in the field of applications and modelling of mathematics is Henry Otto Pollak (Mudaly & Dowlath, 2016). Mathematics is the backbone of modern economics, as it plays a basic role in creating ideas, constructing new theories, and empirically testing ideas and theories. Stundziene (2013) mentioned prediction of the state’s economic situation is very important, as it shows the competitiveness of the state, production, sales, profit of the firm and human welfare. In fact, the prediction of economic outcomes in the contemporary literatures are difficult to explain without mathematics.

Many studies in economic field of macro economy, linear regression model has been widely used as prediction methods in predicting the GDP and inflation (Priambodo, Rahayu, Hazidar et al., 2019). When looking for a mathematical representation of the GDP data, exponential function can be implied, as it discovers the simplest description of the data and well known for mathematical-defines distributions (Nielsen, 2015). In GDP prediction, many researchers have conducted in-depth discussion on exponential model for different situations, which makes the model more efficient and can be used more widely (Dong & Zhu, 2014). Lega and Brown (2016) have stated that the parabolic model can provide good predictions of the future value for the economic growth, as this approach is easily applicable, due to its mathematical simplicity. Significantly, linear, exponential and parabolic regression are the mathematical approach that can be applied to predict the gross domestic product. Therefore, this study is conducted to apply the mathematical approach in predicting gross domestic product (GDP) in Malaysia. It is hope that this study will contribute towards better understanding of the GDP behaviour in future.

**Methodology**

There are three mathematical models that were applied and compared in this gross domestic product in Malaysia, which are linear, exponential and parabolic regression. The type of data used to conduct this study is a yearly secondary data for the period from year 1970 to 2019 and the data is collected from the reliable and trusted source, known as Department of Statistics Malaysia official portal. Based on the yearly secondary data from Department of Statistics Malaysia, the data partition between modeling and validation part are in ratio of 90:10, which indicates 90% for modeling part and 10% for validation of the model accuracy (Ram & Davim, 2018). The division of data partition is shown below:

**Linear Regression Model**

The linear regression models that describes the relationship between independent variable, \( x \) and dependent variable, \( y \) can be expressed by the following equation:

\[
Y = a + bX + \varepsilon
\]

(1)

where, \( Y \) is the dependent (response) variable, \( a \) and \( b \) are parameters of the model called regression coefficients, \( X \) is the independent (exploratory) variable, and \( \varepsilon \) is the random residual (error) variable, and random variable is a normally distributed with mean equal to zero (Montemayor et al., 2018).

Based on the above algebraic formula with linear regression for making prediction is:

\[
Y = a + bX
\]

(2)

The residuals data of the linear regression model is the difference between the observed data of the dependent variable, \( y \) and the fitted value of estimated, \( \hat{y} \).

The parameters \( a \) and \( b \) are obtained from the following equations:

\[
a = \frac{\sum x^2 \sum y - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}
\]

(3)

\[
b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}
\]

(4)

where,

\( n \) = Number of data
\( x \) = Year
\( y \) = Gross domestic product

**Exponential Regression Model**

The exponential regression model that describes the relationship between two variables, \( x \) and \( y \) can be expressed by the following equation:

\[
Y = ab^x \varepsilon
\]

(5)

The algebraic formula with exponential regression for making prediction is:

\[
Y = ab^x
\]

(6)

By applying logarithms on both sides for \( Y = ab^x \), it is presented as follows:

\[
\log Y = \log a + \log b^x
\]

\[
\log Y = \log a + X \log b
\]

(7)

For this model, the parameters of \( \log a \) and \( \log b \) are obtained from the following equations:

\[
\log a = \frac{\sum x^2 \sum \log Y - \sum X \sum x \log Y}{n \sum x^2 - (\sum X)^2}
\]

(8)

\[
\log b = \frac{n \sum X \log Y - \sum \log Y \sum X}{n \sum x^2 - (\sum x)^2}
\]

(9)
where,
\( n \) = Number of data
\( x \) = Year
\( y \) = Gross domestic product

**Parabolic Regression Model**

The parabolic regression model that describes the relationship between two variables \( x \) and \( y \) can be expressed by the following equation:

\[
Y = a + bX + cX^2 + \epsilon
\]  

(10)

The algebraic formula with parabolic regression for making predictions is:

\[
Y = a + bX + cX^2
\]

(11)

where \( Y \) is the dependent variable, \( a \) is the constant (free term of equation), \( b \) and \( c \) are the coefficients of independent variable, and \( X \) is the independent variable.

The parameters \( a \), \( b \) and \( c \) are obtained from the following equations:

\[
a = \frac{\sum y((\Sigma x^3)^2 - \Sigma x^2 \Sigma x^4) + \sum xy[(\Sigma x^2)^2 - n \Sigma x^4] + \sum x^2 y[(\Sigma x^2)^2 - \Sigma x \Sigma x^3]}{\sum x^2 \Sigma x^4 - 2 \Sigma x \Sigma x^2 \Sigma x^3 + (\Sigma x^2)^3 - n \Sigma x^2 \Sigma x^4 + n(\Sigma x^3)^2}
\]  

(12)

\[
b = \frac{\sum y[(\Sigma x^2)^2 - n \Sigma x \Sigma x^3] + \sum xy[(\Sigma x^2)^2 - n \Sigma x^4] + \sum x^2 y[(\Sigma x^2)^2 - n \Sigma x^3]}{\sum x^2 \Sigma x^4 - 2 \Sigma x \Sigma x^2 \Sigma x^3 + (\Sigma x^2)^3 - n \Sigma x^2 \Sigma x^4 + n(\Sigma x^3)^2}
\]  

(13)

\[
c = \frac{\sum y[(\Sigma x^2)^2 - n \Sigma x \Sigma x^3] + \sum xy[n \Sigma x^3 - n \Sigma x^2] + \sum x^2 y[(\Sigma x^2)^2 - n \Sigma x^2]}{\sum x^2 \Sigma x^4 - 2 \Sigma x \Sigma x^2 \Sigma x^3 + (\Sigma x^2)^3 - n \Sigma x^2 \Sigma x^4 + n(\Sigma x^3)^2}
\]  

(14)

where,
\( n \) = Number of data
\( x \) = Year
\( y \) = Gross domestic product

**Model Accuracy**

The estimation error such as mean absolute error (MAE), mean square error (MSE) and mean absolute percentage error (MAPE) can be calculated to compare the model accuracy for three of the models, as these estimation errors are top most common errors used in numerous studies and most popular in the independent surveys that were performed over a timeline of 25 years (Botchkarev, 2018).

The formula of estimation error, \( \epsilon \) can be done by using the following formula:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|
\]

(15)

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

(16)

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \times 100
\]

(17)

Where MAE, MSE and MAPE indicates the estimation errors, in which the calculation only involves the actual, \( y_i \) and estimated, \( \hat{y}_i \) value of gross domestic product in Malaysia, and these estimation errors are used to check the validation consistency of the best fitted model. Meanwhile, \( n \) is the number of observations in measuring the model accuracy.
Prediction of Malaysia GDP for years 2020 to 2029

After the best-fitted model in predicting the GDP data is determined either linear, exponential or parabolic regression, the GDP prediction can be made in which the model with the least value of error will be chosen as the best-fitted model in this study. Hence, the best-fitted model will be used to predict the Malaysian GDP that can give a long-term outlook for the period of 2020 to 2029.

Result and Discussion

Linear Regression Model

The linear regression model is constructed based on the following values:

\[ n = 45, \sum x = 89,640, \sum y = 11,888,184, \sum x^2 = 178,570,470, \sum xy = 23,834,667,606 \]

where, \( n \) is the number of data, whilst \( x \) and \( y \) are year and gross domestic product value respectively.

Substituting these values into Equations (3) and (4), the values of \( a \) and \( b \) are obtained:

\[ a = -39,997,072, \quad b = 20,211 \]

Next, substituting the values \( a, b \) and year, \( X \) into Equation (2), the linear regression model is obtained as below:

\[ Y = -39,997,072 + 20,211X \]  

(18)

The accuracy of the model is calculated by substituting the year, \( x \) into Equation (18) and the result is shown below:

| Year | \( Y \) | \( \hat{Y} \) | \( |Y - \hat{Y}| \) | \( (Y - \hat{Y})^2 \) | \( \frac{|Y - \hat{Y}|}{Y} \times 100 \) |
|------|--------|--------|----------------|----------------|------------------|
| 2015 | 1,063,355 | 729,046 | 334,309 | 111,762,681,727 | 31.44 |
| 2017 | 1,173,600 | 769,469 | 404,131 | 163,322,119,886 | 34.44 |
| 2018 | 1,230,000 | 789,680 | 440,320 | 193,881,563,633 | 35.80 |
| 2019 | 1,420,000 | 809,892 | 610,108 | 372,232,222,774 | 42.97 |

Referring to the table above, \( Y \) is the gross domestic product and \( \hat{Y} \) is the estimation of gross domestic product value using linear regression model.

Thus,

\[ \sum |Y - \hat{Y}| = 2,147,839, \quad \sum (Y - \hat{Y})^2 = 970,057,896,630, \quad \sum \frac{|Y - \hat{Y}|}{Y} \times 100 = 177.03 \]

By substituting these values and \( n = 5 \) into Equation (15), (16) and (17), the value for MAE, MSE and MAPE are obtained as follows:

\[ \text{MAE} = 429,568 \quad \text{MSE} = 194,011,579,326 \quad \text{MAPE} = 35.41 \]

Exponential Regression Model

For this model, the values of \( \log Y \) are calculated and the values are shown in the Table 2.
Table 2: Values of the log Y.

<table>
<thead>
<tr>
<th>Year</th>
<th>log Y</th>
<th>Year</th>
<th>log Y</th>
<th>Year</th>
<th>log Y</th>
<th>Year</th>
<th>log Y</th>
</tr>
</thead>
</table>

From the table above, the following values are obtained:

\[ n = 45, \ \Sigma x = 89,640, \ \Sigma \log Y = 229.14395, \ \Sigma x^2 = 178,570,470, \ \Sigma X \log Y = 456795.00379 \]

where, \( n \) is the number of data, whilst \( x \) and \( y \) are year and gross domestic product value respectively.

Substituting these values into Equations (8) and (9), the values \( a \) and \( b \) are found:

\[
\log a = -84.20988, \text{ thus } a = 6.16764(10^{-85})
\]

\[
\log b = 0.04483, \text{ thus } b = 1.108742
\]

Next, substituting the values \( a \) and \( b \) into Equation (6), the exponential regression model is constructed as follows:

\[ Y = 6.16764(10^{-85})(1.108742)^x \]  

(19)

The accuracy of the model is calculated by substituting the year, \( x \) into Equation (19) and the result is shown below:

Table 3: Calculation for the accuracy of exponential regression model.

| Year | \( Y \) | \( \hat{Y} \) | \(|Y - \hat{Y}|\) | \((Y - \hat{Y})^2\) | \(|Y - \hat{Y}| \times 100\) |
|------|--------|--------|-----------------|-----------------|-----------------|
| 2015 | 1,063,355 | 1,327,959 | 264,604 | 70,015,376,700 | 24.88 |
| 2016 | 1,108,227 | 1,472,363 | 364,136 | 132,595,368,564 | 32.86 |
| 2017 | 1,173,600 | 1,632,470 | 458,870 | 210,562,126,145 | 39.10 |
| 2018 | 1,230,000 | 1,809,988 | 579,988 | 336,385,833,818 | 47.15 |
| 2019 | 1,420,000 | 2,006,809 | 586,809 | 344,344,314,218 | 41.32 |

Referring to the table above, \( Y \) is the gross domestic product and \( \hat{Y} \) is the estimation of gross domestic product value using exponential regression model.

Thus,

\[
\Sigma |Y - \hat{Y}| = 2,254,408, \ \Sigma (Y - \hat{Y})^2 = 1,093,903,019,445, \ \Sigma \frac{|Y - \hat{Y}|}{Y} \times 100 = 185.32
\]

By substituting these values and \( n = 5 \) into Equation (15), (16) and (17), the value for MAE, MSE and MAPE are obtained as follows:

\[
\text{MAE} = 450,882 \quad \text{MSE} = 218,780,603,889 \quad \text{MAPE} = 37.06
\]
Parabolic Regression Model

The following values are obtained to construct the parabolic regression model.

\[ n = 45, \sum x = 89,640, \sum y = 11,888,184, \sum x^2 = 178,570,470, \sum xy = 23,834,667,606, \]
\[ \sum x^3 = 3.55743 \times 10^{11}, \sum x^4 = 7.08730 \times 10^{14}, \sum x^2y = 4.77871 \times 10^{13} \]

where, \( n \) is the number of data, whilst \( x \) and \( y \) are year and gross domestic product value respectively.

Substituting these values into Equations (12), (13) and (14), the values of \( a, b \) and \( c \) are obtained as follow:

\[ a = 3,279,644,865, \quad b = -3,312,922, \quad c = 836.63435 \]

Next, substituting the values of \( a, b \) and \( c \) into Equation (11), the parabolic regression model is constructed as below:

\[ Y = 3,279,644,865 - 3,312,922X + 836.63435X^2 \] (20)

The accuracy of the model is calculated by substituting the year, \( x \) into Equation (20) and the result is shown below:

| Year | \( Y \) | \( \hat{Y} \) | \( |Y - \hat{Y}| \) | \( (Y - \hat{Y})^2 \) | \( \frac{Y - \hat{Y}}{Y} \times 100 \) |
|------|--------|--------|--------|----------------|-----------------|
| 2015 | 1,063,355 | 1,031,378.44 | 31,977 | 1,022,500,348 | 3.01 |
| 2016 | 1,108,227 | 1,090,929.83 | 17,297 | 299,192,122 | 1.56 |
| 2017 | 1,173,600 | 1,152,154.49 | 21,446 | 459,910,062 | 1.83 |
| 2018 | 1,230,000 | 1,215,052.41 | 14,948 | 223,430,386 | 1.22 |
| 2019 | 1,420,000 | 1,279,623.61 | 140,376 | 19,705,531,833 | 9.89 |

Referring to the table above, \( Y \) is the gross domestic product and \( \hat{Y} \) is the estimation of gross domestic product value using parabolic regression model.

Thus,

\[ \Sigma |Y - \hat{Y}| = 226,043, \quad \Sigma (Y - \hat{Y})^2 = 21,710,564,751, \quad \Sigma \frac{|Y - \hat{Y}|}{Y} \times 100 = 17.50 \]

By substituting these values and \( n = 5 \) into Equation (15), (16) and (17), the value for MAE, MSE and MAPE are obtained as follows:

\[ \text{MAE} = 45,209 \quad \text{MSE} = 4,342,112,950 \quad \text{MAPE} = 3.50 \]

Estimation Values for the Three Different Models

<table>
<thead>
<tr>
<th>Year</th>
<th>( \hat{Y}_{\text{Linear}} )</th>
<th>( \hat{Y}_{\text{Exponential}} )</th>
<th>( \hat{Y}_{\text{Parabolic}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>-180,471</td>
<td>12,759</td>
<td>83,399</td>
</tr>
<tr>
<td>1971</td>
<td>-160,259</td>
<td>14,147</td>
<td>67,653</td>
</tr>
<tr>
<td>1972</td>
<td>-140,048</td>
<td>15,685</td>
<td>53,581</td>
</tr>
<tr>
<td>1973</td>
<td>-119,836</td>
<td>17,390</td>
<td>41,182</td>
</tr>
<tr>
<td>1974</td>
<td>-99,625</td>
<td>19,282</td>
<td>30,456</td>
</tr>
<tr>
<td>1975</td>
<td>-79,413</td>
<td>21,376</td>
<td>21,403</td>
</tr>
<tr>
<td>1976</td>
<td>-59,202</td>
<td>23,703</td>
<td>14,024</td>
</tr>
<tr>
<td>Year</td>
<td>Linear Value</td>
<td>Exponential Value</td>
<td>Parabolic Value</td>
</tr>
<tr>
<td>------</td>
<td>--------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>1977</td>
<td>-38,990</td>
<td>26,280</td>
<td>8,318</td>
</tr>
<tr>
<td>1978</td>
<td>-18,779</td>
<td>29,138</td>
<td>4,285</td>
</tr>
<tr>
<td>1979</td>
<td>1,433</td>
<td>32,307</td>
<td>1,925</td>
</tr>
<tr>
<td>1980</td>
<td>21,644</td>
<td>35,820</td>
<td>1,239</td>
</tr>
<tr>
<td>1981</td>
<td>41,856</td>
<td>39,715</td>
<td>2,226</td>
</tr>
<tr>
<td>1982</td>
<td>62,067</td>
<td>44,034</td>
<td>4,886</td>
</tr>
<tr>
<td>1983</td>
<td>82,279</td>
<td>48,822</td>
<td>9,220</td>
</tr>
<tr>
<td>1984</td>
<td>102,490</td>
<td>54,131</td>
<td>15,227</td>
</tr>
<tr>
<td>1985</td>
<td>122,702</td>
<td>60,017</td>
<td>22,907</td>
</tr>
<tr>
<td>1986</td>
<td>142,913</td>
<td>66,544</td>
<td>32,260</td>
</tr>
<tr>
<td>1987</td>
<td>163,125</td>
<td>73,780</td>
<td>43,287</td>
</tr>
<tr>
<td>1988</td>
<td>183,336</td>
<td>81,802</td>
<td>55,987</td>
</tr>
<tr>
<td>1989</td>
<td>203,547</td>
<td>90,698</td>
<td>70,360</td>
</tr>
<tr>
<td>1990</td>
<td>223,759</td>
<td>100,560</td>
<td>86,406</td>
</tr>
<tr>
<td>1991</td>
<td>243,970</td>
<td>111,496</td>
<td>104,126</td>
</tr>
<tr>
<td>1992</td>
<td>264,182</td>
<td>123,620</td>
<td>123,519</td>
</tr>
<tr>
<td>1993</td>
<td>284,393</td>
<td>137,062</td>
<td>144,585</td>
</tr>
<tr>
<td>1994</td>
<td>304,605</td>
<td>151,967</td>
<td>167,324</td>
</tr>
<tr>
<td>1995</td>
<td>324,816</td>
<td>168,492</td>
<td>191,737</td>
</tr>
<tr>
<td>1996</td>
<td>345,028</td>
<td>186,814</td>
<td>217,823</td>
</tr>
<tr>
<td>1997</td>
<td>365,239</td>
<td>207,128</td>
<td>245,582</td>
</tr>
<tr>
<td>1998</td>
<td>385,451</td>
<td>229,652</td>
<td>275,015</td>
</tr>
<tr>
<td>1999</td>
<td>405,662</td>
<td>254,624</td>
<td>306,121</td>
</tr>
<tr>
<td>2000</td>
<td>425,874</td>
<td>282,313</td>
<td>338,900</td>
</tr>
<tr>
<td>2001</td>
<td>446,085</td>
<td>313,012</td>
<td>373,352</td>
</tr>
<tr>
<td>2002</td>
<td>466,297</td>
<td>347,049</td>
<td>409,478</td>
</tr>
<tr>
<td>2003</td>
<td>486,508</td>
<td>384,788</td>
<td>447,277</td>
</tr>
<tr>
<td>2004</td>
<td>506,720</td>
<td>426,630</td>
<td>486,749</td>
</tr>
<tr>
<td>2005</td>
<td>526,931</td>
<td>473,022</td>
<td>527,894</td>
</tr>
<tr>
<td>2006</td>
<td>547,142</td>
<td>524,460</td>
<td>570,713</td>
</tr>
<tr>
<td>2007</td>
<td>567,354</td>
<td>581,490</td>
<td>615,205</td>
</tr>
<tr>
<td>2008</td>
<td>587,565</td>
<td>644,722</td>
<td>661,370</td>
</tr>
<tr>
<td>2009</td>
<td>607,777</td>
<td>714,830</td>
<td>709,209</td>
</tr>
<tr>
<td>2010</td>
<td>627,988</td>
<td>792,562</td>
<td>758,721</td>
</tr>
<tr>
<td>2011</td>
<td>648,200</td>
<td>878,746</td>
<td>809,906</td>
</tr>
<tr>
<td>2012</td>
<td>668,411</td>
<td>974,302</td>
<td>862,764</td>
</tr>
<tr>
<td>2013</td>
<td>688,623</td>
<td>1,080,250</td>
<td>917,295</td>
</tr>
<tr>
<td>2014</td>
<td>708,834</td>
<td>1,197,718</td>
<td>973,500</td>
</tr>
</tbody>
</table>

Table above present the data obtained by the equation of linear, exponential and parabolic regression model, whilst Figure 1 illustrates the graphical design of the historical GDP data and the estimation values using three different mathematical models.
Figure 1: Comparison between historical data and three regression models.

Figure above shows the graphics design of the historical data and three mathematical models, which are linear, exponential and parabolic regression models. Comparing each of the models with respect to the real historical values, linear regression model is smaller with respect to historical values from year 1970 until 1981, even there are negative values that are found from 1970 to 1978. Starting from 1982 to 2004, the values of the linear regression model are larger than the historical data. Then, in 2005 to 2014, the values of the linear regression model are smaller with respect to historical data.

Additionally, the values of exponential regression model in year 1970 to 1977 are larger than the historical data. While, in year 1978 to 1993, the values of historical data are larger than the exponential regression model, with exception of 1985 and 1986 as these values are presented to the contrary. From 1994 to 1999, the values of historical data are smaller compared to the historical data. Then, in years 2000 until 2007, the values of historical data are larger than the exponential regression model values, with exception in year 2004 since the value is presented to the contrary. From 2008 until 2014, the values of exponential regression model are larger than the historical data, with the exception in 2010, as this value illustrate that the historical value is larger than the exponential regression model.

Furthermore, starting year 1970 to 1975, the values of parabolic regression model are larger than the historical data. However, from 1976 to 1992 the historical data values are larger with respect to the parabolic regression model. Then, in year 1993 to 2009, the parabolic regression values are larger compared to the historical data, with the exception of 2000, 2005 and 2006, as these values are presented to the contrary. From 2010 to 2014, the values of parabolic regression model are smaller with respect to the historical data.

**Comparison for Models Validation**

The comparison of estimation error between the model that stated in Table 6. Parabolic regression model have the smallest error measurement as compared to other models. It signifies that the estimation values of parabolic regression model are nearly or closed to the historical data of gross domestic product. Therefore, parabolic regression model is the most suitable mathematical model to predict the future value of gross domestic product in Malaysia, since this model has the smallest error from all the other analysed models.
Table 6: Accuracy validation for each model.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>429,568</td>
<td>194,011,579,326</td>
<td>35.41</td>
</tr>
<tr>
<td>Exponential Regression</td>
<td>450,882</td>
<td>218,780,603,889</td>
<td>37.06</td>
</tr>
<tr>
<td>Parabolic Regression</td>
<td><strong>45,209</strong></td>
<td><strong>4,342,112,950</strong></td>
<td><strong>3.50</strong></td>
</tr>
</tbody>
</table>

**Malaysia GDP Prediction using Parabolic Regression Model**

Table 6 clearly shows that the parabolic regression model fits the historical data of the Malaysia’s gross domestic product. Therefore, the parabolic regression model is used to predict the Malaysia’s gross domestic product for years of 2020 to 2029, and the calculation is shown below:

\[
Y_{2020} = 3,279,644,485 - 3,312,922(2020) + 836.63435(2020)^2 = 1,345,227
\]

\[
Y_{2021} = 3,279,644,485 - 3,312,922(2021) + 836.63435(2021)^2 = 1,413,144
\]

\[
Y_{2022} = 3,279,644,485 - 3,312,922(2022) + 836.63435(2022)^2 = 1,482,735
\]

\[
Y_{2023} = 3,279,644,485 - 3,312,922(2023) + 836.63435(2023)^2 = 1,553,999
\]

\[
Y_{2024} = 3,279,644,485 - 3,312,922(2024) + 836.63435(2024)^2 = 1,626,936
\]

\[
Y_{2025} = 3,279,644,485 - 3,312,922(2025) + 836.63435(2025)^2 = 1,701,546
\]

\[
Y_{2026} = 3,279,644,485 - 3,312,922(2026) + 836.63435(2026)^2 = 1,777,830
\]

\[
Y_{2027} = 3,279,644,485 - 3,312,922(2027) + 836.63435(2028)^2 = 1,855,787
\]

\[
Y_{2028} = 3,279,644,485 - 3,312,922(2028) + 836.63435(2028)^2 = 1,935,418
\]

\[
Y_{2029} = 3,279,644,485 - 3,312,922(2029) + 836.63435(2029)^2 = 2,016,721
\]

The figure above demonstrates the prediction result for gross domestic product in Malaysia for the next ten years from period 2020 to 2029. Next, prediction of the gross domestic product values has manifested an increasing figure for the ten years prediction period. Hence, it can be concluded that based on the predicted values above, Malaysian gross domestic product has a good economic growth in the next ten years.
Conclusion

In this study, three regression models have been constructed based on the Malaysia’s gross domestic product data estimated from year 1970 until 2019. This study clearly manifests the differences among the linear regression model, exponential regression model and parabolic regression model. It is observed that parabolic regression model for predicting the Malaysia’s gross domestic product is the most appropriate model in this study. This is because parabolic regression model is the best fitted model, since it is adjusted to the real conditions of the Malaysia’s gross domestic product. Also, it is obtained that the prediction values of GDP in Malaysia will increase for the next ten years (2020 - 2029).

References


