Estimating multiple linear regression parameters using term omission method

Jubran A. Labban
University of Al-Qadisiya, Diwaniya

ABSTRACT

In this paper, we introduce a new method to estimate multiple linear regression parameters, namely Multiple Term Omission (MTO). Then, we compare its performance with other three methods: Ordinary Least Square (OLS), Maximum Likelihood (ML) and Bayesian Model using several criteria, such as Mean Average Deviation (MAD), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error standard (RMSE). MTO method has the finest consequences as compared to the other methods for the experimented data.

Keywords: Bayesian Model, Maximum Likelihood, Mean Absolute Percentage Error, Mean Average Deviation, Multiple Linear Regression, Multiple Term Omission, Ordinary Least Square, Root Mean Square Error.

Corresponding Author:
Jubran A. Labban
Department of Mathematics, College of Science,
University of Al-Qadisiya, Diwaniya, Iraq
E-mail: Jubran.Labban@qu.edu.iq

1. Introduction

A prediction has been made by using the modeling technique in linear regression. A bivariate model in simple linear regression is built to predict a dependent (response) variable (y) from an independent (explanatory) variable (x). The expansion that includes more than one explanatory variables (x₁, x₂, ..., xₚ) is called multiple linear regression. Multiple linear regression analysis is interested in studying and analyzing the effect of several quantitative independent variables on a dependent variable and it is used as a predictor of future values by estimating the model parameters that are used in the prediction model for forecasting purposes [1, 2].

The result of the evolution in the methods of estimating the parameters, such as the Ordinary Least Squares (OLS) method, the Minimum Likelihood (ML) method, Bayesian Approach and other methods of estimation have given to researchers to look for ways to estimate the parameters of a multimodal model, so I introduce another method called Multiple Term Omission to estimate the parameters of multiple linear regression that based on Term Omission Method which is used to estimate the parameter of distribution [3].

In this paper, an empirical study was adopted, which includes the most important effects that affected the numbers of human resources employed in electricity distribution in Iraq/Baghdad. The multiple regression model was used as a mean of predicting the numbers of employees in the department and the model parameters that adopted in the predictive model of numbers were estimated by using the proposed methods. The performance of the models were compared using Root Mean Square Error (RMSE), Mean Average Deviation (MAD) and Mean Absolute Percentage Error (MAPE) criteria.

2. Multiple linear regression model

The multiple linear regression model is used to describe the relationship between the dependent variable y consists of (n) of observations and k of independent variables (x₁, x₂, ..., xₖ), which can be expressed by the following equation [4]:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i \]

Or, it can be organized as:
\[ Y_i = \beta_0 + \sum_{k=1}^{j} \beta_k x_{ik} + \epsilon_i \quad i = 1, 2, \ldots, n \]  \hspace{1cm} \text{(1)}

Where \((\beta_0, \beta_1, \beta_2, \ldots, \beta_j)\) stand for the regression coefficients, \(\epsilon_i\) stands for the random error of the \(i\)th observation for \(i=1, 2, \ldots, n\). Since the number of observations is \(n\), then the \(n\) equations can be formulated as matrices as follows:

\[
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n
\end{bmatrix}
= 
\begin{bmatrix}
  1 & x_{i1} & x_{i2} & \cdots & x_{ij} \\
  1 & x_{11} & x_{12} & \cdots & x_{1j} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & x_{n1} & x_{n2} & \cdots & x_{nj}
\end{bmatrix}
\begin{bmatrix}
  \beta_0 \\
  \beta_1 \\
  \vdots \\
  \beta_j
\end{bmatrix}
+ 
\begin{bmatrix}
  \epsilon_1 \\
  \epsilon_2 \\
  \vdots \\
  \epsilon_n
\end{bmatrix}
\hspace{1cm} \text{(2)}
\]

Or, it can be organized as:

\[ Y = X\beta + U, \]
\hspace{1cm} \text{(3)}

Where,

- \(Y\): The vector of dimension \(n\) of the dependent observations.
- \(X\): An independent observation matrix (explanatory), this matrix is of \(n \times (k+1)\) dimension and it is also called the design matrix.
- \(\beta\): The vector of dimension \((j+1)\) of the regression coefficients.
- \(U\): The vector of dimension \(n\) of the random errors.

The multiple regression model shown in equation (2) is based on several assumptions [5]:

a- The array of independent variables \(X\) is given and it is measured without errors.

b- There is no complete or semi-completed linear relationship between independent variables.

c- Statistical independence exists between independent observations \((x_{i1}, x_{i2}, \ldots, x_{ij})\), and the random error \(\epsilon_i\), that is the columns of the matrix \(X\) are linearly independent of the vector of random errors \(U\), and this can expressed mathematically as follows:

\[ Cov(X, U) = E(XU) - [E(X)][E(U)] = 0 \]
\hspace{1cm} \text{(4)}

\[ Cov(\epsilon_i, \epsilon_j) = E(\epsilon_i \epsilon_j) = \begin{cases} \sigma^2 & i = j \\ 0 & i \neq j \end{cases} \]
\hspace{1cm} \text{(5)}

That is the error vector \(U\) follows a multivariate normal distribution with zero mean and variance-covariance matrix \(\Sigma\), that is:

\[ U \sim n_n(\mathbf{0}, \Sigma). \]

Where \(\Sigma\) is a symmetric matrix of dimension \((n \times n)\), and is expressed as follows:

\[ \Sigma = \begin{bmatrix}
  \sigma^2 & 0 & \cdots & 0 \\
  0 & \sigma^2 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & \sigma^2
\end{bmatrix} = \sigma^2 I_n \]
\hspace{1cm} \text{(6)}

3. Estimation of the parameters of the multiple linear regression model
3.1. Ordinary least square method (OLS)

The OLS method is one of the important methods of estimating the parameters of the linear regression model. The characteristics of this method distinguish it from other known methods, it is characterized by impartiality, it has the minimum variance and it has the property of BLUE (Best Linear Unbiased Estimate) [6].

The OLS method is used to estimate the parameters of the multiple linear regression model \( \beta_0, \beta_1, \beta_2, \ldots, \beta_j \) of equation (2) and the equation written, as follows, by matrices variable

\[
Y = X\beta + U
\]

Which makes using the matrix from of multiple linear regression model, the sum of the error or residual squares as small as possible:

\[
U = Y - X\hat{\beta}
\]

With squaring the two sides, we get

\[
Q = U'U = (Y - X\beta)'(Y - X\beta)
\]

Alternatively, it can be organized as:

\[
Q = U'U = Y'Y - 2\beta'X'Y + \beta'X'X\beta
\]

To find the value of \( \hat{\beta} \) that makes \( U'U \) minimum as possible, the derivative is taken with respect to \( \beta_j \) and equal to zero. That is:

\[
\frac{\partial Q}{\partial \beta} = 2X'Y + 2X'X\hat{\beta}
\]

Simplifying equation (10) as follows:

\[
X'X\hat{\beta} = X'Y
\]

\[
\hat{\beta} = (X'X)^{-1}X'Y
\]

Where \( \hat{\beta} \) represent the vector of predictor using OLS method.

3.2. Maximum likelihood estimator (MLE)

The deriving of OLS estimator discussed in previous section, which by minimizing the sum of residual. With given additional assumption that errors are normally and independently distributed as also least square estimator. Also the estimator of \( \sigma^2 \) can be computed by using MLE.

Recall the model in eq. (3)

\[
Y = X\beta + U
\]

Where \( U \sim n(0, \sigma^2) \).

The likelihood function of the normal distribution which is defined as follows [7]:

\[
f(U_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left( -\frac{1}{2\sigma^2} U_i^2 \right)
\]

and

\[
L(U, \beta, \sigma^2) = \prod_{i=1}^{n} f(U_i) = \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left( -\frac{1}{2\sigma^2} U'U \right)
\]

Since

\[
U = Y - X\beta
\]

Then, the likelihood function becomes:
\[ L(Y, X, \beta, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp \left( -\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \right) \]

Taking the log of the likelihood function:

\[ \ln(L(Y, X, \beta, \sigma^2)) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \quad \ldots (12) \]

Derive equation (12) w.r.t. \( \beta \) and equality to zero

\[ (Y - X\beta)'(Y - X\beta) = 0 \]

This also can be solved as the OLS in previous section:

\[ \hat{\beta} = (X'X)^{-1}X'Y \]

As it seen, that the ML estimator is equivalent to the OLS estimator under assumption of normal error.(8)

### 3.3. Bayesian linear regression

One of the possible estimation of regression model parameters using Bayesian approach with the parameter estimation. The distribution of prior, likelihood and posterior can be found in Bayesian approach. The posterior distribution is equal to the product of the distribution of prior with likelihood. The normal assumption of linear regression, where the error \( U \sim n(0, \sigma^2) \), leads the variables \((Y | X, \beta, \sigma^2)\) to be also normally distributed. Thus \((Y | X, \beta, \sigma^2) = n(X \beta, \sigma^2)\) and its probability density functions are [9]:

\[ f(Y | X, \beta, \sigma^2) = (2\pi)^{-1/2} \sigma^{-1} \exp \left[ -\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \right] \quad \ldots (13) \]

So, the likelihood function of the probability density function above can be written as follows:

\[ L(Y | X, \beta, \sigma^2) = \prod_{i=1}^{n} (2\pi)^{-1/2} \sigma^{-1} \exp \left[ -\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \right] \quad \ldots (14) \]

and the posterior can be written as follows:

\[ \text{Posterior} \propto \text{Likelihood} \times \text{Prior} \]

That is:

\[ f(\beta, \sigma^2 | Y, X) \propto f(Y | X, \beta, \sigma^2) f(\sigma^2) f(\beta | \sigma^2) \]

\[ f(\beta, \sigma^2 | Y, X) \propto \sigma^{-n/2} \exp \left[ -\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \right] \times (\sigma^2)^{-(\frac{v}{2}+1)} \exp \left[ -\frac{\nu S^2}{2\sigma^2} \right] \times \]

\[ (\sigma^2)^{-k/2} \exp \left[ -\frac{1}{2\sigma^2} (\beta - \mu)'(\beta - \mu) \right] \quad \ldots (15) \]

and it can be obtained the estimation of regression model parameters using Bayesian approach.

The prior distribution used in this study is the Normal distribution for the \( \beta \) parameter and the Gamma inverse distribution for the parameter \( \sigma^2 \).

### 3.4. Multiple term omission method (MTO)

This algorithm was developed by Labban in 2005(10) on simple linear regression. This method is summarized by deleting the terms of the multiple regression model one by one and then finding the values of the parameters using the background steps. This method depends on the number of independent variables \( j \) which is equal to the number of sample data \( n \) minus one, i.e. \( n = j + 1 \). In case that the sample data is greater than the number of variables plus one, i.e. \( n > j + 1 \) then it is resorted to one of the following procedures:

I- Increase the number of variables by adding \( n - j - 1 \) of slack variable, and the condition of these variables is:

a- All values of the each slack variable(s) are different, that is, \( x_{nj} \neq x_{ir} \), where \( j + 2 < r \leq n \) and \( i \neq j \) for all \( j \leq n \) and \( i \geq 1 \).

b- All values of slack variable(s) are different, that is, \( x_{ij} \neq x_{is} \), where \( s \leq n \), \( r \neq s \) and \( 1 \leq i \leq n \).

c- For ease of calculation, it is preferable to use the values of slack variables as a function of rank of the explanatory variables, like:
\[ x_1 = u_1 \text{(rank)} = i \]
\[ x_2 = u_2 \text{(rank)} = i^2 \]
and so on, where \( i \) represent the rank of explanatory variables.

2- Reducing the number of sample data to \( k+1 \).

This research is dealing with the first procedure. The first procedure has \( n-1 \) of multiple linear regression coefficients with \( k \) of explanatory variables and \( n-k-1 \) of slack variable(s), which can be calculated by:

\[ \hat{\beta}_i = \Delta^{(i)} y_1 - \sum_{j=i+1}^{k} \hat{\beta}_j \Delta^{(i)} x_{j(i-j)} \quad , \quad i=k,...,0 \]  

Where,
\( \hat{\beta}_i \) represents estimated value of \( i \)th multiple linear regression parameter.
\( \Delta^{(i)} y_1 \) represents initial value of definite divided for predicting variable.
\( \Delta^{(i)} x_{j(i-j)} \) represents initial definite divided of explanatory variables.
\( \hat{\beta}_j \) represents estimated calculated value of previous parameter of multiple linear regression.

Moreover, by calculating the values of the estimated parameters of the multiple linear regression equation, the value of the response variable can be calculated by:

\[ \hat{y}_i = \hat{\beta}_i + E_i \]  

where:
\( \hat{y}_i \) represents the final estimated value.
\( \hat{\beta}_i \) represents the initial estimated value.
\( E_i \) represents the error value from the slack variable(s) which can be calculated by:

\[ E_i = \sum_{j=k+2}^{n} x_{ij} \quad , \quad i=1,...,n \]  

Here,
\( x_{ij} \) represents to value of slack variable(s).

For more illustration, we present the following example of two independent variables \( (k=2) \) and four sample data \( (n=4) \).

The multiple linear regression equation can be written as follows:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad , \quad i=1,2,3,4 \]

The data can be written as

\[
\begin{align*}
  x_{11} & \quad x_{12} & \quad y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} \\
  x_{21} & \quad x_{22} & \quad y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} \\
  x_{31} & \quad x_{32} & \quad y_3 = \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} \\
  x_{41} & \quad x_{42} & \quad y_4 = \beta_0 + \beta_1 x_{41} + \beta_2 x_{42} 
\end{align*}
\]

Since the number of variables \( (k=2) \) is less than the number of the sample data minus one \( (n-1) \), i.e. \( k<n-1 \), so we add a one slack variable \( (n-k-1=4-2-1=1) \) to get:
\[ \Delta^{(i)} y_i = \frac{y_i - \hat{y}_i}{x_{i1} - x_{11}} = \beta_1 + \beta_2 \frac{x_{22} - x_{12}}{x_{21} - x_{11}} + \left( \frac{x_{33} - x_{23}}{x_{31} - x_{21}} \right) \]

\[ \Delta^{(i)} y_2 = \frac{y_2 - \hat{y}_2}{x_{21} - x_{11}} = \beta_1 + \beta_2 \frac{x_{22} - x_{12}}{x_{21} - x_{11}} + \left( \frac{x_{33} - x_{23}}{x_{31} - x_{21}} \right) \]

\[ \Delta^{(i)} y_3 = \frac{y_3 - \hat{y}_3}{x_{31} - x_{11}} = \beta_1 + \beta_2 \frac{x_{32} - x_{22}}{x_{31} - x_{21}} + \left( \frac{x_{43} - x_{33}}{x_{41} - x_{31}} \right) \]

\[ \Delta^{(i)} y_4 = \frac{y_4 - \hat{y}_4}{x_{41} - x_{11}} = \beta_1 + \beta_2 \frac{x_{42} - x_{22}}{x_{41} - x_{21}} + \left( \frac{x_{43} - x_{33}}{x_{41} - x_{31}} \right) \]

Here it is noticeable that there is only one slack variable \( x_{i3} \) (i=1,...,4)

The value of \( \Delta^{(i)} y_i \) equal to the coefficient of slack variable \( x_{i3} \).

Using the background steps to get the remaining of estimated values of multiple linear regression equation as follows:

\[ \hat{\beta}_3 = \Delta^{(i)} y_i \]
\[ \hat{\beta}_2 = \Delta^{(i)} y_1 - [\hat{\beta}_3 \Delta^{(i)} x_{11}] \]
\[ \hat{\beta}_1 = \Delta^{(i)} y_1 - [\hat{\beta}_2 \Delta^{(i)} x_{11} + \hat{\beta}_3 \Delta^{(i)} x_{12}] \]
\[ \hat{\beta}_0 = \Delta^{(i)} y_1 - [\hat{\beta}_1 \Delta^{(i)} x_{11} + \hat{\beta}_2 \Delta^{(i)} x_{12} + \hat{\beta}_3 \Delta^{(i)} x_{13}] \]

To calculate the error that generated from adding slack variable using Eq. (19) as follows:

\[ E_i = \sum_{j=4}^{4} x_{ij}, \quad i = 1,...,4 \]
\[ E_1 = x_{14}, \quad E_2 = x_{24}, \quad E_3 = x_{34}, \quad E_4 = x_{44} \]

Therefore, the values of estimated explanatory variable using MTO are:

\[ \hat{y}_i = \hat{y}_i + E_i \]
4. Applied example

Data adopted by the researcher was the distribution electricity department in Baghdad. We used the data to predict the number of engineering staff the electricity distribution department in Baghdad. The number of workers in the electricity in general is affected by several factors directly or indirectly from the perspective of the researcher, which are:

- $X_1$: Electric power received.
- $X_2$: Electric power sold.
- $X_3$: Number of electric power consumers.
- $X_4$: Number of stations 33/11 KV.
- $X_5$: Number of feeders 33 KV.
- $X_6$: Number of feeders 11 KV.
- $X_7$: Load capacity.

$Y$: numbers of engineering staff.

The data of workers of engineering staff is given in Table (1).

Table 1. The data of workers of Engineering Staff of Electricity Distribution Department in Baghdad From (1985 - 1994)

<table>
<thead>
<tr>
<th>Year</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>5.122</td>
<td>4.627</td>
<td>542</td>
<td>86</td>
<td>167</td>
<td>869</td>
<td>1060</td>
<td>101</td>
</tr>
<tr>
<td>1986</td>
<td>5.659</td>
<td>4.349</td>
<td>580.6</td>
<td>90</td>
<td>179</td>
<td>925</td>
<td>1252</td>
<td>116</td>
</tr>
<tr>
<td>1987</td>
<td>6.330</td>
<td>5.255</td>
<td>609.65</td>
<td>92</td>
<td>184</td>
<td>939</td>
<td>1252</td>
<td>114</td>
</tr>
<tr>
<td>1988</td>
<td>6.545</td>
<td>5.595</td>
<td>638.7</td>
<td>95</td>
<td>191</td>
<td>959</td>
<td>1252</td>
<td>119</td>
</tr>
<tr>
<td>1989</td>
<td>7.312</td>
<td>6.216</td>
<td>668.3</td>
<td>92</td>
<td>184</td>
<td>998</td>
<td>1354</td>
<td>134</td>
</tr>
<tr>
<td>1990</td>
<td>6.046</td>
<td>7.001</td>
<td>706.5</td>
<td>97</td>
<td>206</td>
<td>1008</td>
<td>1435</td>
<td>142</td>
</tr>
<tr>
<td>1991</td>
<td>5.114</td>
<td>4.490</td>
<td>720.5</td>
<td>98</td>
<td>208</td>
<td>1050</td>
<td>1534</td>
<td>143</td>
</tr>
<tr>
<td>1992</td>
<td>8.050</td>
<td>7.245</td>
<td>741.2</td>
<td>100</td>
<td>211</td>
<td>1077</td>
<td>1453</td>
<td>145</td>
</tr>
<tr>
<td>1993</td>
<td>7.911</td>
<td>6.934</td>
<td>759.2</td>
<td>101</td>
<td>211</td>
<td>1088</td>
<td>1534</td>
<td>128</td>
</tr>
<tr>
<td>1994</td>
<td>8.768</td>
<td>7.253</td>
<td>759.2</td>
<td>101</td>
<td>211</td>
<td>1088</td>
<td>1534</td>
<td>128</td>
</tr>
</tbody>
</table>

The proposed model to predict the number of engineering staff can be written as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \epsilon_i \text{ i=1,...,10 \hspace{1cm}} (20)$$

Table 2 showing the parameters estimation of the model by OLS method that can be getting from the data in Table 1 and the model in (20), which is equivalent to the MLE.

Table 2. Parameters estimation using OLS and MLE methods

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimation Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-479.76254</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-13.09306</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>13.99935</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.92476</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>4.05582</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.66876</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.76924</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-0.04194</td>
</tr>
</tbody>
</table>

So, the forecasting equation using the methods OLS and MLE is:

$$\hat{y}_{OLS,MLE} = -479.76254-13.09306x_1 +13.99935x_2 -0.92476x_3 + 4.05582x_4 + 0.66876x_5 + 0.76924x_6 - 0.04194x_7$$

The estimated explanatory values for Table (1) that shows the error according to the values of the estimated parameters by OLS and MLE methods shown in Table (3).

Table 3. Estimated explanatory values for Table (1) and the error according to the values of the estimated parameters using OLS and MLE methods

<table>
<thead>
<tr>
<th>i</th>
<th>Years</th>
<th>$y$</th>
<th>$\hat{y}_{OLS,MLE}$</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1985</td>
<td>101</td>
<td>101.2281</td>
<td>0.228095</td>
</tr>
<tr>
<td>2</td>
<td>1986</td>
<td>116</td>
<td>116.4836</td>
<td>0.483609</td>
</tr>
<tr>
<td>3</td>
<td>1987</td>
<td>114</td>
<td>113.142</td>
<td>0.857977</td>
</tr>
<tr>
<td>4</td>
<td>1988</td>
<td>119</td>
<td>119.0722</td>
<td>0.072197</td>
</tr>
</tbody>
</table>
Also the model of linear regression analysis with Bayesian approach shown in Table (4)

Table 4. Linear regression model using Bayesian approach

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Value</td>
<td>-481.4514</td>
<td>-12.2967</td>
<td>13.2804</td>
<td>-0.9237</td>
<td>4.0905</td>
<td>0.6658</td>
<td>0.7708</td>
<td>-0.04534</td>
</tr>
</tbody>
</table>

So, the forecasting equation using the Bayesian approach is:

$$\hat{y}_{\text{Bayes}} = -481.4514 - 12.2967x_1 + 13.2804x_2 - 0.9237x_3 + 4.0905x_4 + 0.6658x_5 + 0.7708x_6 - 0.04534x_7$$

And the estimated explanatory values for Table (1) that shows the error according to the values of the estimated parameters by Bayesian approach shown in Table (5).

Table 5. Estimated explanatory values for Table 1 and the error according to the values of the estimated parameters using Bayesian Approach

<table>
<thead>
<tr>
<th>i</th>
<th>Years</th>
<th>$y$</th>
<th>$\hat{y}_{\text{Bayes}}$</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1985</td>
<td>101</td>
<td>101.0848</td>
<td>0.0848</td>
</tr>
<tr>
<td>2</td>
<td>1986</td>
<td>116</td>
<td>116.7551</td>
<td>0.7551</td>
</tr>
<tr>
<td>3</td>
<td>1987</td>
<td>114</td>
<td>113.1913</td>
<td>0.8087</td>
</tr>
<tr>
<td>4</td>
<td>1988</td>
<td>119</td>
<td>119.0801</td>
<td>0.0801</td>
</tr>
<tr>
<td>5</td>
<td>1989</td>
<td>133</td>
<td>132.3918</td>
<td>0.6082</td>
</tr>
<tr>
<td>6</td>
<td>1990</td>
<td>134</td>
<td>135.2099</td>
<td>1.2099</td>
</tr>
<tr>
<td>7</td>
<td>1991</td>
<td>142</td>
<td>141.2841</td>
<td>0.7159</td>
</tr>
<tr>
<td>8</td>
<td>1992</td>
<td>143</td>
<td>143.0395</td>
<td>0.0395</td>
</tr>
<tr>
<td>9</td>
<td>1993</td>
<td>145</td>
<td>144.9082</td>
<td>0.0918</td>
</tr>
<tr>
<td>10</td>
<td>1994</td>
<td>128</td>
<td>128.4542</td>
<td>0.4542</td>
</tr>
</tbody>
</table>

The slack variable values for MTO method that can be used are as follows:

$$x_{i8} = i$$

$$x_{i9} = i^2$$

and so the Table 1 can be written as follows:

Table (6) shows the data of Table (1) after adding slack variables.

Table 6. Data of Table 1 after adding two slack variables

<table>
<thead>
<tr>
<th>i</th>
<th>Years</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
<th>$x_7$</th>
<th>$x_8=i$</th>
<th>$x_9=i^2$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1985</td>
<td>5.122</td>
<td>4.627</td>
<td>542</td>
<td>86</td>
<td>167</td>
<td>869</td>
<td>1060</td>
<td>1</td>
<td>1</td>
<td>103</td>
</tr>
<tr>
<td>2</td>
<td>1986</td>
<td>5.659</td>
<td>4.349</td>
<td>580.6</td>
<td>90</td>
<td>179</td>
<td>925</td>
<td>1190</td>
<td>2</td>
<td>4</td>
<td>122</td>
</tr>
<tr>
<td>3</td>
<td>1987</td>
<td>6.330</td>
<td>5.255</td>
<td>609.65</td>
<td>92</td>
<td>184</td>
<td>939</td>
<td>1252</td>
<td>3</td>
<td>9</td>
<td>126</td>
</tr>
<tr>
<td>4</td>
<td>1988</td>
<td>6.545</td>
<td>5.595</td>
<td>638.7</td>
<td>95</td>
<td>191</td>
<td>959</td>
<td>1285</td>
<td>4</td>
<td>16</td>
<td>139</td>
</tr>
<tr>
<td>5</td>
<td>1989</td>
<td>7.312</td>
<td>6.216</td>
<td>668.3</td>
<td>96</td>
<td>202</td>
<td>998</td>
<td>1277</td>
<td>5</td>
<td>25</td>
<td>163</td>
</tr>
<tr>
<td>6</td>
<td>1990</td>
<td>8.046</td>
<td>7.001</td>
<td>678.64</td>
<td>97</td>
<td>206</td>
<td>1008</td>
<td>1354</td>
<td>6</td>
<td>36</td>
<td>176</td>
</tr>
<tr>
<td>7</td>
<td>1991</td>
<td>7.911</td>
<td>6.934</td>
<td>706.5</td>
<td>98</td>
<td>208</td>
<td>1046</td>
<td>1435</td>
<td>7</td>
<td>49</td>
<td>198</td>
</tr>
<tr>
<td>8</td>
<td>1992</td>
<td>5.114</td>
<td>4.490</td>
<td>720.5</td>
<td>99</td>
<td>208</td>
<td>1050</td>
<td>1312</td>
<td>8</td>
<td>64</td>
<td>215</td>
</tr>
</tbody>
</table>
Where \( x_i \) and \( x_j \) are slack variables.

Also, Table 7 depicts the estimated parameters from the data in Table (1) and by using MTO method.

Table 7. Estimated parameters using MTO method

<table>
<thead>
<tr>
<th>Parameters</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
<th>( \beta_6 )</th>
<th>( \beta_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Value</td>
<td>-409.978</td>
<td>-10.749</td>
<td>7.245</td>
<td>12.5265</td>
<td>-0.773</td>
<td>2.9846</td>
<td>0.1022</td>
<td>0.8212</td>
</tr>
</tbody>
</table>

So, the forecasting equation using the MTO method is:

\[
\hat{y} = -409.978 -10.749x_1 + 12.5265x_2 - 0.7739x_3 + 2.9846x_4 + 0.1022x_5 + 0.8212x_6 - 0.0600x_7 + 4.7611x_8 + 0.6645x_9
\]

Table 8 shows the initial and final estimated explanatory variable values and the error according to Table (1) using estimated parameters according to MTO method.

Table 8. The initial and final estimated explanatory variable values and the error according to Table (1) using estimated parameters according to MTO method

| i | Years | \( y \) | \( \hat{y}_{MTO} \) | \( \hat{E}_i \) = \( x_8 + x_9 \) | \( \hat{y}_{MTO} = \hat{y}_{MTO} - E_i \) | Error \( |\hat{y}_{Bayes} - y| \) |
|---|-------|--------|-----------------|-----------------|-----------------|----------------|
| 1 | 1985  | 101    | 103.1511        | 2               | 101.1511        | 0.1511         |
| 2 | 1986  | 116    | 122.1654        | 6               | 116.1654        | 0.1654         |
| 3 | 1987  | 114    | 126.1867        | 12              | 114.1867        | 0.1867         |
| 4 | 1988  | 119    | 139.2049        | 20              | 119.2049        | 0.2049         |
| 5 | 1989  | 133    | 163.2157        | 30              | 133.2157        | 0.2157         |
| 6 | 1990  | 134    | 176.2225        | 42              | 134.2225        | 0.2225         |
| 7 | 1991  | 142    | 198.2327        | 56              | 142.2327        | 0.2327         |
| 8 | 1992  | 143    | 215.2389        | 72              | 143.2389        | 0.2389         |
| 9 | 1993  | 145    | 235.2504        | 90              | 145.2504        | 0.2504         |
| 10| 1994  | 128    | 238.2633        | 110             | 128.2633        | 0.2633         |

5. Results and discussion

The estimates of the multiple linear regression model obtained using the OLS, MLE shown in Table (2), Bayesian, and MTO shown in Tables (4) and (7), respectively, can be compared using several criteria such as Root Mean Square Error (RMSE), Mean Average Deviation (MAD) and Mean Absolute Percentage Error (MAPE) which can be calculated for the data as:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2}, \quad MAD = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \quad MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|
\]

Table 9. Comparison between methods in multiple linear regression model

<table>
<thead>
<tr>
<th>OLS/MLE</th>
<th>Bayesian</th>
<th>MTO</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>4.705163</td>
<td>4.8482</td>
</tr>
<tr>
<td>MAD</td>
<td>0.47051</td>
<td>0.48482</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.371999</td>
<td>0.38217</td>
</tr>
</tbody>
</table>
From the comparison of criteria in Table 9, the order of the methods based on the minimum of RMSE as follows: MTO, OLS/MLE and finally Bayesian with small differences between OLS/MLE and Bayesian. While the order of methods based on MAD criteria is the same order in RMSE criteria. MAPE criteria also have the same order.

6. Conclusions

1. According to RMSE criterion, it is obvious that the MTO method have preference compared to the other methods.
2. The OLS, MLE and Bayesian methods are the second, third and fourth in order according to the RMSE criterion respectively.
3. Also, in MAD criterion, it is clear that the MTO method has the preference compared to the other methods, while the methods OLS, MLE is better than the Bayesian approach.
4. MTO method has the preference over all other methods, according to the MAPE criterion, followed by the OLS, MLE methods and Bayesian approach in order.
5. It can be concluded that the estimate parameters of the multiple linear regression model using MTO Method is better than OLS, MLE methods and Bayesian approach.

References