Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 4, December 2020, pp.2290-2299

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 (x_1, x_2, …, x_p) 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 (x_1, x_2, ..., x_j), which can be expressed by the following equation [4]:

$$Y_i = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_j x_{ij} + \varepsilon_i$$

Or, it can be organized as:

$$Y_i = \beta_0 + \sum_{k=1}^{j} \beta_k x_{ik} + \varepsilon_i \quad i = 1,2,\ldots,n \qquad \text{... (1)}$$

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

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1j} \\ 1 & x_{21} & x_{22} & \cdots & x_{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nj} \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \text{... (2)}$$

Or, it can be organized as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}, \qquad \text{... (3)}$$

Where,

$\mathbf{Y}$: The vector of dimension n of the dependent observations.
$\mathbf{X}$: An independent observation matrix (explanatory), this matrix is of n×(k+1) dimension and it is also called the design matrix
$\boldsymbol{\beta}$: The vector of dimension (j+1) of the regression coefficients.
$\mathbf{U}$: 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} , ..., x_{ij}), and the random error ε_i, that is the columns of the matrix $\mathbf{X}$ are linearly independent of the vector of random errors $\mathbf{U}$, and this can expressed mathematically as follows:

$$Cov(\mathbf{X}, \mathbf{U}) = E(\mathbf{X}'\mathbf{U}) - [E(\mathbf{X})]'[E(\mathbf{U})] = 0 \qquad \text{... (4)}$$

d- The random error ε_i, i=1,2,...., n, has a normal distribution of zero mean, and variance σ^2 is constant from one observation to another, that is $\varepsilon_i \sim$ n(0, σ^2). It also assume that the errors ($\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$) are statistically independent, and are expressed mathematically as follows:

$$E(\varepsilon_i) = 0$$

$$Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2 & i = j \\ 0 & i \neq j \end{cases} \qquad \text{... (5)}$$

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

$$\mathbf{U} \sim n_n(0, \textstyle\sum),$$

Where ∑ is a symmetric matrix of dimension (n×n), and is expressed as follows:

$$\sum = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I_{n\times n} \qquad \text{... (6)}$$

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

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{U}$$

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

$$\mathbf{U} = \mathbf{Y} - \mathbf{X\beta} \qquad \text{... (7)}$$

With squaring the two sides, we get

$$\mathbf{Q} = \mathbf{U'U} = (\mathbf{Y} - \mathbf{X\beta})'(\mathbf{Y} - \mathbf{X\beta}) \qquad \text{... (8)}$$

Alternatively, it can be organized as:

$$\mathbf{Q} = \mathbf{U'U} = \mathbf{Y'Y} - 2\mathbf{\beta' X'Y} + \mathbf{\beta' X'X\beta} \qquad \text{… (9)}$$

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

$$\frac{\partial Q}{\partial \mathbf{\beta}} = \begin{bmatrix} \frac{\partial Q}{\partial \beta_o} \\ \frac{\partial Q}{\partial \beta_1} \\ \vdots \\ \frac{\partial Q}{\partial \beta_j} \end{bmatrix} = \mathbf{2X'Y} + \mathbf{2X'X\hat{\beta}} \qquad \text{... (10)}$$

Simplifying equation (10) as follows:

$$\mathbf{X'X\hat{\beta}} = \mathbf{X'Y}$$

$$\hat{\mathbf{\beta}} = (\mathbf{X'X})^{-1}.\mathbf{X'Y} \qquad \text{… (11)}$$

Where $\hat{\mathbf{\beta}}$ 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 σ^2 can be computed by using MLE.
Recall the model in eq. (3)

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{U}$$

Where $\mathbf{U} \sim n(0, \sigma^2)$.
The likelihood function of the normal distribution which is defined as follows [7]:

$$f(\mathbf{U}_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{U}_i^2\right)$$

and

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

Since

$$\mathbf{U} = \mathbf{Y} - \mathbf{X\beta}$$

Then, the likelihood function becomes:

$$L(\mathbf{Y},\mathbf{X},\boldsymbol{\beta},\sigma^2)=\frac{1}{\sigma^n(2\pi)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right)$$

Taking the log of the likelihood function:

$$Ln\left(L(\mathbf{Y},\mathbf{X},\boldsymbol{\beta},\sigma^2)\right)=-\frac{n}{2}\ln(2\pi)-n\ln(\sigma)-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \quad \ldots (12)$$

Derive equation (12) w.r.t. $\boldsymbol{\beta}$ and equality to zero

$$(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})=0$$

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

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}.\mathbf{X}'\mathbf{Y}$$

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 regression modeling is Bayesian linear regression 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 ~ n(0,σ^2), leads the variables ($\mathbf{Y}\,|\,\mathbf{X}$, $\boldsymbol{\beta}$, σ^2) to be also normally distributed. Thus ($\mathbf{Y}\,|\,\mathbf{X}$, $\boldsymbol{\beta}$, σ^2) ~ n($\mathbf{X}\boldsymbol{\beta}$,σ^2) and its probability density functions are [9]:

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

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

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

and the posterior can be written as follows:

Posterior $\propto$ Likelihood $\times$ Prior

That is:

$$f(\boldsymbol{\beta},\sigma^2\,|\,\mathbf{Y},\mathbf{X})\propto f(\mathbf{Y}\,|\,\mathbf{X},\boldsymbol{\beta},\sigma^2)\,f(\sigma^2)\,f(\boldsymbol{\beta}\,|\,\sigma^2)$$

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

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

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 β parameter and the Gamma inverse distribution for the parameter σ^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:

1- 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, $x_{ri}\neq x_{rj}$, where j + 2 < r ≤ n and i ≠ j for all j ≤ n and i ≥ 1.
   b- All values of slack variable(s) are different, that is, $x_{ri}\neq x_{si}$, where s ≤ n , r ≠ s and 1 ≤ i ≤ 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:

$$x_1 = u_1(rank) = i$$
$$x_2 = u_2(rank) = i^2 \qquad \dots (16)$$

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:

$$\hat{\beta}_i = \Delta^{(i)} y_1 - \sum_{j=i+1}^{k} \hat{\beta}_j \Delta^{(i)} x_{1(j-i)} \qquad , \text{ i=k,..,0} \qquad \dots (17)$$

Where,
$\hat{\beta}_i$ represents estimated value of ith multiple linear regression parameter.
$\Delta^{(i)} y_1$ represents initial value of definite divided for predicting variable.
$\Delta^{(i)} x_{1(j-i)}$ 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:

$$\hat{y}_i = \hat{\hat{y}}_i + E_i , \qquad \dots (18)$$

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

$$E_i = \sum_{j=k+2}^{n} x_{ij} \qquad , \qquad i = 1, \dots, n \qquad \dots (19)$$

Here,
x_{ij} represent 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:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i , \qquad \text{i=1,2,3,4}$$

The data can be written as

$$\begin{array}{lll} x_{11} & x_{12} & y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} \\ x_{21} & x_{22} & y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} \\ x_{31} & x_{32} & y_3 = \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} \\ x_{41} & x_{42} & y_4 = \beta_0 + \beta_1 x_{41} + \beta_2 x_{42} \end{array}$$

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:

$$x_{11} \quad x_{12} \quad x_{13} \quad y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + x_{13}$$

$$\Delta^{(1)} y_1 = \frac{y_2 - y_1}{x_{21} - x_{11}} = \beta_1 + \beta_2 \frac{x_{22} - x_{12}}{x_{21} - x_{11}} + \left(\frac{x_{23} - x_{13}}{x_{21} - x_{11}} \right)$$

$$x_{21} \quad x_{22} \quad x_{23} \quad y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + x_{23}$$

$$\Delta^{(1)} y_2 = \frac{y_3 - y_2}{x_{31} - x_{21}} = \beta_1 + \beta_2 \frac{x_{32} - x_{22}}{x_{31} - x_{21}} + \left(\frac{x_{33} - x_{23}}{x_{31} - x_{21}} \right)$$

$$x_{31} \quad x_{32} \quad x_{33} \quad y_3 = \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + x_{33}$$

$$\Delta^{(1)} y_3 = \frac{y_4 - y_3}{x_{41} - x_{31}} = \beta_1 + \beta_2 \frac{x_{42} - x_{32}}{x_{41} - x_{31}} + \left(\frac{x_{43} - x_{33}}{x_{41} - x_{31}} \right)$$

$$x_{41} \quad x_{42} \quad x_{43} \quad y_4 = \beta_0 + \beta_1 x_{41} + \beta_2 x_{42} + x_{43}$$

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

$$\Delta^{(1)} x_{11} \quad \Delta^{(1)} x_{12} \quad \Delta^{(1)} y_1 = \beta_1 + \beta_2 \Delta^{(1)} x_{11} + \Delta^{(1)} x_{12}$$

$$\Delta^{(2)} y_1 = \frac{\Delta^{(1)} y_2 - \Delta^{(1)} y_1}{\Delta^{(1)} x_{21} - \Delta^{(1)} x_{11}} = \beta_2 + \left(\frac{\Delta^{(1)} x_{22} - \Delta^{(1)} x_{12}}{\Delta^{(1)} x_{21} - \Delta^{(1)} x_{11}} \right)$$

$$\Delta^{(1)} x_{21} \quad \Delta^{(1)} x_{22} \quad \Delta^{(1)} y_2 = \beta_1 + \beta_2 \Delta^{(1)} x_{21} + \Delta^{(1)} x_{22}$$

$$\Delta^{(2)} y_2 = \frac{\Delta^{(1)} y_3 - \Delta^{(1)} y_2}{\Delta^{(1)} x_{31} - \Delta^{(1)} x_{21}} = \beta_2 + \left(\frac{\Delta^{(1)} x_{32} - \Delta^{(1)} x_{22}}{\Delta^{(1)} x_{31} - \Delta^{(1)} x_{21}} \right)$$

$$\Delta^{(1)} x_{31} \quad \Delta^{(1)} x_{32} \quad \Delta^{(1)} y_3 = \beta_1 + \beta_2 \Delta^{(1)} x_{31} + \Delta^{(1)} x_{32}$$

$$\Delta^{(2)} x_{11} \quad \Delta^{(2)} y_1 = \beta_2 + \Delta^{(2)} x_{11}$$

$$\Delta^{(3)} y_1 = \frac{\Delta^{(2)} x_{21} - \Delta^{(2)} x_{11}}{\Delta^{(2)} x_{21} - \Delta^{(2)} x_{11}}$$

$$\Delta^{(2)} x_{21} \quad \Delta^{(2)} y_2 = \beta_2 + \Delta^{(2)} x_{21}$$

The value of $\Delta^{(3)} y_1$ 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:

$$\hat{\beta}_3 = \Delta^{(3)} y_1$$
$$\hat{\beta}_2 = \Delta^{(2)} y_1 - [\hat{\beta}_3 \Delta^{(2)} x_{11}]$$
$$\hat{\beta}_1 = \Delta^{(1)} y_1 - [\hat{\beta}_2 \Delta^{(1)} x_{11} + \hat{\beta}_3 \Delta^{(1)} x_{12}]$$
$$\hat{\beta}_0 = \Delta^{(0)} y_1 - [\hat{\beta}_1 \Delta^{(0)} x_{11} + \hat{\beta}_2 \Delta^{(0)} x_{12} + \hat{\beta}_3 \Delta^{(0)} x_{13}]$$

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

$$E_i = \sum_{j=4}^{4} x_{ij} \quad , \quad i = 1, \ldots, 4$$

$$E_1 = x_{14} \ , \ E_2 = x_{24} \ , \ E_3 = x_{34} \ , \ E_4 = x_{44}$$

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

$$\hat{y}_i = \hat{\hat{y}}_i + E_i$$

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)

1985	**5.122**	**4.627**	**542**	**86**	**167**	**869**	**1060**	**101**
1986	5.659	4.349	580.6	90	179	925	1190	116
1987	6.330	5.255	609.65	92	184	939	1252	114
1988	6.545	5.595	638.7	95	191	959	1285	119
1989	7.312	6.216	668.3	96	202	998	1277	133
1990	8.046	7.001	678.64	97	206	1008	1354	134
1991	7.911	6.934	706.5	98	208	1046	1435	142
1992	5.114	4.490	720.5	99	208	1050	1312	143
1993	8.050	7.245	741.2	100	211	1077	1453	145
1994	8.768	7.253	759.2	101	211	1088	1534	128

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i \quad i=1,..,10 \qquad \dots (20)$$

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

Parameters	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Estimation Value	-479.76254	-13.09306	13.99935	-0.92476	4.05582	0.66876	0.76924	-0.04194

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

i	Years	y	$\hat{y}_{OLS,MLE}$	Error $\lvert \hat{y}_{OLS,MLE} - y \rvert$
1	1985	101	101.2281	0.228095
2	1986	116	116.4836	0.483609
3	1987	114	113.142	0.857977
4	1988	119	119.0722	0.072197

i	Years	y	$\hat{y}_{OLS,MLE}$	Error $\lvert \hat{y}_{OLS,MLE} - y \rvert$
5	1989	133	132.0984	0.901557
6	1990	134	135.1097	1.109711
7	1991	142	141.4031	0.596923
8	1992	143	143.1545	0.154507
9	1993	145	145.0572	0.05724
10	1994	128	128.2433	0.243347

Also the model of linear regression analysis with Bayesian approach shown in Table (4)

Table 4. Linear regression model using Bayesian approach

Parameters	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Estimation Value	-481.4514	-12.2967	13.2804	-0.9237	4.0905	0.6658	0.7708	-0.04534

So, the forecasting equation using the Bayesian approach is:

$$\hat{y}_{Bayes} = -481.4514 - 12.2967x_1 + 13.2804x_2 - 0.92376x_3 + 4.0905x_4 + 0.6658x_5 + 0.7708x_6 - 0.04534x_7$$

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

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

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

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

i	Years	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8=i	x_9=i^2	y
9	1993	8.050	7.245	741.2	100	211	1077	1453	9	81	235
10	1994	8.768	7.253	759.2	101	211	1088	1534	10	100	238

Where x_8 and x_9 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

Parameters	**β_0**	**β_1**	**β_2**	**β_3**	**β_4**	**β_5**	**β_6**	**β_7**	**Coeff. Slack Var. 1**	**Coeff. Slack Var. 2**
Estimation Value	-409.978	-10.749	12.5265	-0.773	2.9846	0.1022	0.8212	-0.06	4.7611	0.6645

So, the forecasting equation using the MTO method is:

$$\hat{y} = -409.9782 - 10.7498 x_1 + 12.5265 x_2 - 0.7739 x_3 + 2.9846 x_4 + 0.1022 x_5 + 0.8212 x_6 - 0.0600 x_7 + 4.7611 x_8 + 0.6645 x_9$$

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	$\hat{\hat{y}}_{MTO}$	$E_i = x_8 + x_9$	$\hat{y}_{MTO} = \hat{\hat{y}}_{MTO} - E_i$	Error $\lvert \hat{y}_{Bayes} - y \rvert$
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:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \quad , \qquad MAD = \frac{1}{n}\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert \, , \qquad MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{\lvert y_i - \hat{y}_i \rvert}{y_i}$$

Table 9. Comparison between methods in multiple linear regression model

	OLS/MLE	**Bayesian**	**MTO**
RMSE	4.705163	4.8482	0.215942
MAD	0.47051	0.48482	0.21316
MAPE	0.371999	0.38217	0.001282

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

[1]. S. Aarthi, M. Sarvathanayan, B. Kumar, R. Rakesh, Post-Graduate College Admission Recommender Using Data Analytics, *International Journal of Innovative Technology and Exploring Engineering*, vol.8, no.6, pp.840-842, 2019.

[2]. B. Shyti, D. Valera, The Regression Model for the Statistical Analysis of Albanian Economy, *International Journal of Mathematics Trends and Technology*, vol.62, no.2, pp.90-96, 2018.

[3]. J.A. Labban, Modified and Simplified Method for Finite Difference (Divided), *Al-Qadisiyah Journal for Science*, vol.10, no.2, pp. 244-251, 2005.

[4]. H. Chen, Y. Chang, T. Tung, Comparison of Two Quantitative Analysis Techniques to Predict the Evaluation of Product Form Design, *Mathematical Problems in Engineering*, vol.2014, pp.1-9, 2014.

[5]. A. Rencher, G. Schaalje, Linear models in statistics. Hoboken, N.J.: Wiley-Interscience; 2008.

[6]. B. Geremew, K. Rao, Multiple Linear Regression Model with Two Parameter Doubly Truncated New Symmetric Distributed Errors, *The International Journal of Engineering and Science*, vol.6, no.3, pp.1-8, 2017.

[7]. L. Jäntschi, D. Bálint, S. Bolboacă , Multiple Linear Regressions by Maximizing the Likelihood under Assumption of Generalized Gauss-Laplace Distribution of the Error, *Computational and Mathematical Methods in Medicine*, vol.2016, pp.1-8, 2016.

[8]. D.C. Montgomery, E.A. Peck, G.G. Vining, Introduction to Linear Regression Analysis, 5th Edition. Wiley.com. 2012.

[9]. S. Permai, H. Tanti, Linear regression model using Bayesian approach for energy performance of residential building, *Procedia Computer Science*, vol.135, pp.671-677, 2018.

[10]. J.A. Laban, On 2-Parameter Estimation of Lomax Distribution, *Journal of Physics: Conference Series*, 2019.