Comparison of mistakes criteria using multiple linear regressions applied to cotton data

Azad Adil Shareef[1], Ismat Mousa Ibrahim[2]
[1] College of Administration and Economics, Department of Statistics, University of Dohuk
[2] Duhok Polytechnic University

ABSTRACT

Cotton is the main crop not only among cellulosic crops but also among the various crops that are the raw materials which is used in the yarn and fabrics industries. The data of this crop were obtained from the central organization of Statistics in Iraq including yield, domestic consumption, production, total supply, export and import of this crop for the period (1960-2019).The mistakes criteria (*MSE, MAPE, MSAE, SAZ1 and SAZ2*) were compared using multiple linear regressions. The results showed high-quality ratios in terms of mistakes criteria (*SAZ1 and SAZ2*) as compared to other criteria.

Keywords: *OLS, SAZ1, SAZ2, MSE*

Corresponding Author:
Azad Adil Shareef
College of Administration and Economics, Department of Statistics, University of Dohuk, Iraq
E-mail: Azada@uod.ac

1. Introduction

Cotton is considered to be an important fabric crop in the world and its importance in multiple uses in the clothing and cotton textile industry. The significance of the cotton appears in the economic life in the world through taking the 1st rank in three economic centers. Firstly, it is the agricultural center where it takes an important agricultural center among the various agricultural crops. Secondly, it appears that the importance of the cotton as an International Trade Center. Lastly, its importance is appeared in the industry from being the first main subject where it plays an important role for the number of workers in the industry and the quantity of the merchandise being set and valued for satisfying the economic needs [2].

In this study, “ordinary least squares” has been used so as to compare between mistake criteria like sum of absolute for standardized error “*SAZ1* and *SAZ2*”. These criteria had shown good results from the model used where the value of *SAZ1* was the least among other measurements and its value was 0.055 and the following value was mean absolute percentage error *(MAPE)* criteria with the value of 0.26 and after that *SAZ2* with value of 2.79 and then 52.431 and 6053.1 for mean sum of absolute error *(MSAE) and* mean squared error (*MSE)* respectively. All data had been gained from the Central Statistical Bureau of Iraq and the website of https://www.indexmundi.com/

1.1. Problem of study

To find the best criteria, Ordinary Least Square method has used for estimation of parameters of linear regression model between the following mistake criteria of *MSE, MAPE, MSAE, SAZ1* and *SAZ2*. Ordinary Least Square method is applied to residuals as we study the same problem in our research that can be missing value of the residual during the accounting process.

1.2. Aim of study

The aim of this research is to be compared the Ordinary Least Squares Method by using mistakes criteria (*SAZ1*, *SAZ2*, *MAPE*, *MSE*, and *MSAE*). Some programs are adopted in our research like MATLAB, SPSS and S-plus to find parameters for the best criteria of mistakes. At the same time, it can be used to investigate the role of the cotton to increase the Gross National Product and its effects on the economy within Iraqi national income.

1.3. Hypothesis of study

The study assumes that five factors (Import, Export, Yield, Production, Consumption and Total Supply) impact on the Yield. Minimizing the impact of mistake criteria on the results of statistical analysis of the Yield is examined based on dependent variable of the research studies using multiple linear regressions. We obtained the best results by using ordinary least squares method to obtain the estimated value of the dependent variable closest to its real value.

1.4. Economic importance of cotton

Cotton is regarded to be the most important monetary crop in the world. Due to not being able to consume its cotton fiber and seeds directly before manufacturing, it can bring profits to the farmers as compared with other crops like cotton, barely, and the legumes; and that is why cotton is regarded to be a crucial industrial crop in many countries that produce it [7].

Since, the cotton crop shapes an essential income for the country that produces it. It provides ways of working to a large proportion of the population whether in the industry or agriculture like gins, spinning factories, weaving and other industries.

The area under cultivation of cotton in Iraq in 1974 was 104260 acre and the production 40200 tons of cotton was blossomed with average of 306 kg/acre.

The economy of Iraq was not in its best conditions from the beginning of the establishment of the Iraqi State in 1921.The agriculture was the traditional and basic activity in Iraq, but it suffered from major problems because of the system of feudal Kingdom as well as environmental obstacles with respect to the salinity deposition. Nevertheless, the agriculture was the major resource for increasing the national incomes. About 70% of the population was living in the village and the agricultural activity to absorb nearly half of the workforce in the country. The agricultural production was satisfactory with the local consumption to a great extent and Iraq issued lots of agricultural products primarily cotton, seed dates and some of the products of the animal revolution.

Iraq was able to make an agreement with the foreign companies which were getting benefits from the extracting and exporting of Iraqi oil and gave a very small amounts to the government in the 1950s of the last decade. Based on the partnership principle in exporting the raw oil, this contributed the increasing in State revenue and increased the reliance on that revenue which led to dwindle the contribution of other economic activities in gross domestic product. Rather, it distorted the economical frame, so the agricultural activity declined and massive rural to urban migration launched. In that time, job opportunities became greater and the government became the only big operator for people and consumerism had emerged with people due to increased governmental revenue and the increased population of Iraq from about seven millions in 1997 up to approximately 35 million people in 2012 according to the estimates of the Central Statistical Bureau at an average of growth over 3 % and workforce formed nearly a quarter of the population in 1997.

2. Linear regression models

2.1. Simple linear regression of Y on X

In linear regression, the values of *Y* are obtained from several populations, each populations universe has determined by a corresponding *X* value. Randomness of *Y* is fundamental for probability theory to be applied

[5]. Also, it can be assumed that Y populations are normal and have a joint variance. The variable Y is defined as the response variable or dependent variable. The X variable is called the explanatory variable or independent variable [9]. The simple regression model is:

$$Y_i=\beta_0+\beta_1X_i+e_i \quad (1)$$

$i=1, 2, \ldots, n$ and n is sample size.

Where,

Y_i: Dependent variable (Response variable)

Xi: Independent variable (Explanatory variable)

$\beta 0, \beta_1$: Regression coefficient or regression parameters (β_0- Intercept; β_1 - slope coefficient)

ei: Random error

Equation above implies:

$$\left.\begin{aligned} Y_1 &= \beta_0+\beta_1X_1+e_1 \\ Y_2 &= \beta_0+\beta_1X_2+e_2 \\ Y_3 &= \beta_0+\beta_1X_3+e_3 \\ &. \\ &. \\ &. \\ Y_n &= \beta_0+\beta_1X_n+e_n \end{aligned}\right\} \quad (2)$$

The observations of variable Y defined as the following vector: $\boldsymbol{Y}=\begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \boldsymbol{Y}_3 \\ . \\ . \\ . \\ \boldsymbol{Y}_n \end{bmatrix}_{nx1}$ (3)

And the values of variable X are put in the following matrix: $X=\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ . & . \\ 1 & X_n \end{bmatrix}_{nx2}$ (4)

The random term is as the vector: $\boldsymbol{e}=\begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ . \\ . \\ . \\ \boldsymbol{e}_n \end{bmatrix}_{nx1}$ (5)

The regression coefficients are determined by: $\beta=\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_{2x1}$ (6)

Now, the equation above can be written in the matrix compactly as follows:

$$Y_{nx1} = X_{nx2}\beta_{2x1} + e_{nx1}\text{, or} \tag{7}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ . & . \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ . \\ . \\ . \\ \beta_0 + \beta_1 X_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 + e_1 \\ \beta_0 + \beta_1 X_2 + e_2 \\ . \\ . \\ . \\ \beta_0 + \beta_1 X_n + e_n \end{bmatrix} \tag{8}$$

Note that $X\beta$ is the vector of the expected values of the Y_i observation since

$$E\,(Y_i) = \beta_0 + \beta_1 X_i\text{, hence } E(Y) = X\beta \tag{9}$$

where, $E(Y)$ is defined as:

$$E(Y) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ . \\ . \\ . \\ E(Y_n) \end{bmatrix} \tag{10}$$

The column of 1's in the matrix $\boldsymbol{X}$ could be seen as consisting of the dummy variable in the alternate regression model: $Y_i = \beta_0 + \beta_1 X_1 + e_i$ where $X_0 \equiv 1$. Thus, the matrix $\boldsymbol{X}$ may be considered to include column vector of the dummy variable X_0 and another column vector containing predictor variable observations X_i, based on the terms of error of the regression model. Therefore, the normal error regression model in matrix term is: $\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{e}$ is called as a vector of independent normal variables with $E(\boldsymbol{e})=\boldsymbol{0}$ and $Var(\boldsymbol{e}) = \sigma^2 I$, or $e \sim N\,(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ [4].

2.2. Multiple linear regression models:

The multiple linear regression models suppose a linear relationship between the dependent variable and a number of independent variables [6]. In this study, we have five basic independent variables within the model as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} \dots + e_i \quad , i = 1,2, \dots, n \tag{11}$$

Where:

Y_i: The value i-th observation of the dependent variable (Yield).

X_{i1}: The value i-th observation of the first independent variable (Production).

X_{i2}: The value i-th observation of the second independent variable is (Imports).

X_{i3}: The value i-th observation of the third independent variable (Exports).

X_{i4}: The value i-th observation of the forth independent variable (Consumption).

X_{i5}: The value i-th observation of the fifth independent variable (Total supply).

e_i: Random Error

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$: Linear Regression Parameters.

The equation of multiple linear regression can be written as the following [8]:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + e_i \qquad i = 1, \ldots\ldots, n. \tag{12}$$

Where Y_i: Dependent variable

X_i: Independent variables

$\beta_0, \beta_1, \ldots, \beta_p$: Regression parameters (β_0- Intercept; $\beta_1, \ldots, \beta_p$- slope coefficients)

e_i: Random error

p: Number of explanatory variables

Also, we can write model above by matrix form as: $\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{e}$, where

$\boldsymbol{Y}$: vector (nx1) of dependent variables

X: matrix (nxp) of independent variables

$\boldsymbol{\beta}$: vector (px1) of regression parameters model

$\boldsymbol{e}$: vector (nx1) of random errors

The true regression of Y on X consists of the means of population of Y values, where a population is determined by the X values [9] as in:

$$E(Y|(X_1, X_2, \ldots X_p) = \boldsymbol{X}\boldsymbol{\beta}. \tag{13}$$

With respect to the error terms, regression model assumes that $E(e_i)=0$, $\sigma^2(e_i)=\sigma^2$, and $\boldsymbol{e_i}$ are independent normal random variables, the condition $E(e_i)=0$ in matrix terms is: $E(e_i)_{nx1}=0_{nx1}$ and can expressed as:

$$\begin{bmatrix} E(e_1) \\ E(e_2) \\ . \\ . \\ . \\ E(e_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} \tag{14}$$

The condition that the error terms have constant variance σ^2 and that all covariance's $\sigma(e_i, e_j)$ for $i \neq j$ are zero (since e_i are independent), is expressed in matrix terms:

$$\sigma^2(e)_{nxn} = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & 0\ldots & 0 \\ \ldots & & & . \\ \ldots & & & . \\ \ldots & & & . \\ & & & \\ 0 & 0 & 0\ldots & \sigma^2 \end{bmatrix} \tag{15}$$

Since this is scalar matrix, it can be considered in the following simple fashion: $\sigma^2(e)_{nxn} = \sigma^2{}_{nxn}I$. Thus, the normal error regression model in matrix form is: $Y=X\beta+e$ where [6]: e is a vector of independent normal variables with $E(e)=0$ and $\sigma^2(e) = \sigma^2 I$ (16)

2.3. Ordinary least squares method (OLS)

The ordinary least squares is also called classical least squares. This method regarded as the most extremely utilized in statistics applications. It is relied on relationship between two or more than two variables. The principle of this method depends on finding straight line that implement spread shape points by the graph that makes sum of squares to split up points from it less than possible [4]. Determining coefficient (β) value makes this sum less than possible in term of multiple linear regression and estimation linear relation between various variables. One of them is dependent variable and the others are independent variables that impact on dependent variable. In general, the estimation parameters of regression in case of (p) in independent variables is based on [1]. The linear regression model in matrix form is given for the following form:

$$\boldsymbol{Y}=\boldsymbol{X}\,\boldsymbol{\beta}+\boldsymbol{e}. \tag{17}$$

The vector of error $\boldsymbol{e}$ is defined:

$$\boldsymbol{e}=\boldsymbol{Y}-\boldsymbol{X}\,\boldsymbol{\beta}. \tag{18}$$

The sum of squares of coordinates of $\boldsymbol{e}$ as in $e_1^2+\cdots+e_n^2$ can be expressed as $\boldsymbol{e'e}$.

Then for $\boldsymbol{e'e}$ we have:

$$\boldsymbol{e'e}=(\boldsymbol{Y}-\boldsymbol{X}\,\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\,\boldsymbol{\beta})=(\boldsymbol{Y'}-\boldsymbol{\beta'}\,\boldsymbol{X'})(\boldsymbol{Y}-\boldsymbol{X}\,\boldsymbol{\beta}) \tag{19}$$

$$=\boldsymbol{Y'Y}-2\,\boldsymbol{\beta'}\,\boldsymbol{X'Y}+\boldsymbol{\beta'}\,\boldsymbol{X'}\,\boldsymbol{X}\,\boldsymbol{\beta} \tag{20}$$

Based on matrix rules for differentiation, we can obtain the following:

$$\frac{\partial(e'e)}{\partial\boldsymbol{\beta}}=-2\boldsymbol{X'Y}+2\boldsymbol{X'X\beta} \tag{21}$$

By setting the preceding equation to zero, we will have normal equations:

$$(\boldsymbol{X'X})\boldsymbol{\beta}=\boldsymbol{X'Y} \tag{22}$$

The solutions of the normal equations we represented by $\widehat{\boldsymbol{\beta}}$ and the coordinates of this vector $\widehat{\beta_1},\widehat{\beta_2},\ldots,\widehat{\beta_p}$ are estimations of unknown parameters $\beta_1,\ldots,\beta_p$, obtained by the Ordinary Least Squares method. From the last equation, we obtain only one solution if the determinant of the matrix $(\boldsymbol{X'X})$ is not zero which is

$$\widehat{\boldsymbol{\beta}}=(\boldsymbol{X'X})^{-1}\,\boldsymbol{X'Y} \tag{23}$$

Substituting $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{e}$ into the preceding expression we get:

$$\widehat{\boldsymbol{\beta}}=(\boldsymbol{X'X})^{-1}\,\boldsymbol{X}'(\boldsymbol{X\beta}+\boldsymbol{e}) \tag{24}$$

$$=(\boldsymbol{X'X})^{-1}\,\boldsymbol{X'X\beta}+(\boldsymbol{X'X})^{-1}\,\boldsymbol{X'e} \tag{25}$$

$$=\boldsymbol{I\beta}+(\boldsymbol{X'X})^{-1}\,\boldsymbol{X'e} \tag{26}$$

$$=\boldsymbol{\beta}+(\boldsymbol{X'X})^{-1}\,\boldsymbol{X'e} \tag{27}$$

where the matrix $\boldsymbol{I}$ is identity matrix.

Therefore,

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}=(\boldsymbol{X'X})^{-1}\,\boldsymbol{X'e} \tag{28}$$

By definition

$$\text{Var}-\text{Cov}\,(\widehat{\boldsymbol{\beta}})=E[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'] \tag{29}$$

$$=E\{[(\boldsymbol{X'X})^{-1}\,\boldsymbol{X'e}]\left[(\boldsymbol{X'X})^{-1}\,\boldsymbol{X'e}\right]'\} \tag{30}$$

$$=E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'ee'X}(\boldsymbol{X'X})^{-1}] \tag{31}$$

In the last step, we use the fact that for arbitrarily $\boldsymbol{A}$ and $\boldsymbol{B}$ matrices as:

$$(\boldsymbol{AB})'=A'B' \tag{32}$$

Noting that the X's are no stochastic and E ($\boldsymbol{ee'}$) = $\sigma^2\boldsymbol{I}$ and we get

$$\text{Var}-\text{Cov}\,(\widehat{\boldsymbol{\beta}})=E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'ee'X}(\boldsymbol{X'X})^{-1}] \tag{33}$$

$$=(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\sigma^2\boldsymbol{IX}(\boldsymbol{X'X})^{-1} \tag{34}$$

$$=\sigma^2(\boldsymbol{X'X})^{-1} \tag{35}$$

2.4. Investigation of the accuracy of estimation OLS

Criteria always utilize in the practical statistic in term of judge on the models to obtain the best model. There are several kinds of mistake criteria. Some of them are information criteria and the others are mistakes criteria that have been used in this research. Exactness is the most widely used criterion to evaluate the achievement of the estimation methods. The mistakes criteria used in this study are:

2.4.1. SAZ1 and SAZ2 criteria

In this research, we adopted sum of absolute for standardized error (*SAZ1)* and (*SAZ2)* criteria besides the mean square error (*MSE*) as measures for finding the amount of error by using specific models in this work. These two functions give us an idea about the data error behaviour. It can be seen from the application that it is behaving as *MSE*. When the value of *MSE* is growing in size and becoming bigger, the value of *SAZ1* and *SAZ2* are in the same direction. They are both studying the error observations ratio divided by its number, resulting in a criterion that measures the average of the error ratio. We can conclude that they can be used besides the *MSE* measure. The functions are [3]:

$$SAZ1 = \frac{\left|\sum_{i=1}^{n-1} \frac{e_i}{e_{i+1}}\right|}{n-1} \quad (36)$$

$$SAZ2 = \frac{\left|\sum_{i=1}^{n-1} \frac{e_{i+1}}{e_i}\right|}{n-1} \quad (37)$$

Where:

e_i: Previous random error

e_{i+1}: Subsequent random error

n: Number of sample

7.2 Mean Squares Error (MSE)

The mean squared error (*MSE*) is a measure of accuracy calculated by squaring the individual error for each item in a data set and then finding the arithmetic mean or average value of the sum of those squares. The mean squared error gives larger weight to large residuals than to small errors because the residuals are squared before being summed, it takes the following equation:

$$MSE = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-p-1} = \frac{\sum_{i=1}^{n}(e_i)^2}{n-p-1} \quad (38)$$

where,

$$e_i = Y_i - \hat{Y}_i \quad (39)$$

e_i: Random error

n: Number of samples

p: Number of parameters

2.4.2. Mean absolute percentage error (MAPE)

The mean absolute percentage error (*MAPE*) is also known as mean absolute percentage deviation (*MAPD*). The mean absolute percentage error is the average or arithmetic mean of the sum of all percentage errors for a given data set taken without regard to sign. Explicitly, their absolute value is summed and the average computed. It is defined as the following formula:

$$MAPE = \frac{\sum_{i=1}^{n}|PE_i|}{n} \quad (40)$$

where,

$$PE_i = \frac{(Y_i - \hat{Y}_i)}{Y_i} \quad (41)$$

2.4.3. Mean sum of absolute error (MSAE):

Mean sum of absolute error is a mathematical optimization technique similar to the common least squares technique that attempts to obtain a function which closely convergent to a set of data. It is a common standard to differentiate between the method of less absolute error and least squares method that takes the following formula:

$$MSAE = \frac{\sum_{i=1}^{n} \left| Y_i - \hat{Y}_i \right|}{n} = \frac{\sum_{i=1}^{n} \left| e_i \right|}{n} \quad (42)$$

It should be mentioned that minimizing the average absolute errors can be used to minimize the total absolute errors and this is called the least absolute value (*LAV*).

2.5. Data collection

Data were obtained from the Iraqi Central Office of Statistics that is shown in the following link (website), www.index mundi.com. The study was depended on 60 observations within years of (1960-2019). It is applied to the statistical analysis of the data in order to find the best way to model appropriate regression for cotton, and data include five main variables, namely, (production, consumption, total supply, import and export) it affects directly Yield, which would be the response variable

2.5.1. Cotton line model results

$$\hat{Y}_i = 355.35 - 0.488 X_{i1} - 2.84 X_{i2} + 7.6 X_{i3} - 0.83 X_{i4} + 0.97 X_{i5} \quad (43)$$

The results of regression analysis are represented in the Tables (1) and (2) as follows

Table 1. Regression coefficients for Cotton parameters by using Ordinary Least Squares

Predictor	Coefficients	Standard error of coefficients	t-value	P-value
(Constant)	355.354	32.845	10.819	.000
onsumption (X1)	-.488	.973	-.502	.618
Production (X2)	-2.848	1.054	-2.702	.009
Export (X3)	7.607	2.735	2.781	.007
Import (X4)	-.830	.604	-1.373	.175
Supply (X5)	.977	.749	1.306	.197

Table 2. Analysis of variance (ANOVA) for Cotton by using Ordinary Least Squares Method

Source	DF	SS	MS	F	P
Regression	5	115534.310	23106.862	3.817	.005a
Residuals	54	326867.624	6053.104		
Total	59	442401.933			

a. Predictors: (Constant), supply, export, production, import, consumption

b. Dependent Variable: yield

From Table (1), it can be seen that the two parameters are production and exports in the cotton line that are significantly affected to the yield and we can see from Table (2) which is ANOVA Table. The F-value is a significant in cotton line because the value of F-calculated (3.817) is more than the F-tabulated value under significant level ($\alpha = 0.05$) with degrees of freedoms $df_1 = 5$, $df_2 = 54$ which is $F(0.05,5,54) = 2.52$ or the p-value is less than 0.05.

Therefore, the fitted model in the Table (2) is an important for the data in the cotton line

Table 3. Estimation Of coefficient by *OLS* method

Variables	Coefficients	Estimation of coefficient by using *OLS*
Constant	B_0	355.35
X1	B_1	-0.488
X2	B_2	-2.84
X3	B_3	7.6
X4	B_4	-0.83
X5	B_5	0.97
SAZ1		0.055
SAZ2		2.792
MSE		6053.1
MAPE		0.264
MSAE		52.431

3. Comparison of results

After assessing multiple linear regression model coefficients of the *OLS* method, a comparison between the results of analysis can be done. Several measures were used to judge the estimated accuracy of regression equation for the purpose of discussing the results obtained using the mistakes criteria and compared with results of the ordinary least square method. This is to show the purpose of highlighting the advantages of each mistake criteria in finding parameters and logical consequences.

Table (3) shows illustrations variables involved in the study as well as the estimated coefficients for all mistake's criteria values and the value of the comparison criteria.

From observed results in Table (3), we notice that the estimated parameter $\boldsymbol{\beta_1}$ is signalled left and negative, this is logical because there is an indirect correlation between yield and consumption. By increasing the consumption, the income decreases and by increasing the income yield decreases for the consumption that is function in income. Parameter $\boldsymbol{\beta_2}$ is also estimated in left and negative signal and it is logical because the relationship between yield and production in an indirect correlation. This direct correlation is based on reduction of value of production goods. When the values are reduced, the production of individual raises, and so the yield decreases. As we see, the estimated $\boldsymbol{\beta_3}$ is reasonable. It is right and illogical according to classical methods, the positive signal for exported parameter refers to indirect relationship between export and yield. Although the exports are one of the resources to increase income, in Iraq this may contradict the economic theory for Iraq that is a country which is depending in exporting oil and oil exports contribute to Iraq's budget with about 95%.

Parameter $\boldsymbol{\beta_4}$ is also estimated in left and negative signal and it is logical because the relationship between yield and import is indirect correlation. When the values are decreased, the import of individual rises, and so the yield decreases. We can conclude that the final parameter β5 is reasonable. It is right and illogical according to classical methods. The positive signal for supply parameter refers to direct relationship between supply and yield. This enormous contributed ratio in funding country incomes are affected by international oil price fluctuations and hence the whole income fluctuates. On the other hand, the negative signal means lack of justice and equality in distributing the income due to corruption in Iraq, also it could be because of the inflation phenomena in Iraq's economy which reduces the real income of individuals and thus reduces yield.

The comparison is made between mistakes criteria that is used and for the comparison to be accurate. The criteria used for comparison is supposed to be very accurate, and our observation is in general in Table (3). So, we can show that the mistakes criteria is the best criteria to estimate by adopting the value (*SAZ1*) with much less error. The higher value of *SAZ1* led to a small rise in the proportion of average coefficient of determination.

MAPE comes in second place for the trade-off. *SAZ2* comes in third place. Then, *MSAE* has been in the fourth place and finally the *MSE* criteria has been come by using *OLS* in the multiple regression models to estimate the best mistake criteria in this work that was applied in the cotton data.

4. Conclusion

According to the results of the practical side of our study the following conclusions are yielded:

1- Throughout the study, it became clear that the mistakes criteria “*SAZ1* and *SAZ2*’’ gained good results from the model used where the value of *SAZ1* was the least among other measurements and its value was 0.055. Secondly, the *MAPE* measurement came by the value of 0.26 and after that *SAZ2* with value of 2.79 and then 52.431 and 6053.1 for *MSAE* and *MSE* respectively.

2- During the observing data, the cotton crop data were contributing to the increasing of the economy of Iraq since 1920s of last decade where that crop has produced and issued up to 1970. After that, Iraq stopped the exporting and began to import and the overall provision became on oil which now it forms 95% from the total Iraqi budget.

References

[1] D. N. Gujarati, “Basic Econometrics”, *McGraw-Hill, Inc. NewYork,1995, Journal of Finance and Economics,* Vol. 2 No. 6, 210-214

[2] H. A. Mosly, “Cotton Cultivation and death and manufacture of oil and vegetable butter from seed”, T1, Aladdin Publishing, Damascus, 2000.

[3] Ismat mousa Ibrahim, “Using New Criteria to Compare between Some Robust Method and Ordinary Least Squares in Multiple Regression with Application on Wheat Data in Iraq”, Sofia University "ST. KLIMENT OHRIDSKI", 2016.

[4] J. Neter, W. Wasserman and M. Kutner, Applied Linear Statistical Models, 2nd edition, IRWIN, 1985. McGraw-Hill, "a business unit of The McGraw-Hill Companies.

[5] P. J. Huber, “Robust Statistical Procedures”. *Der Universitat Bayreuth, Germany,1996, 2nd ed. — Society for Industrial and Applied Mathematics, 1996*. — 79 p. — ISBN: 089871379X, 9780898713794.

[6] R.G. D. Steel and H. T. James, Principles and Procedures of Statistic, 2nd edition, New York : McGraw-Hill, c1997.

[7] I. FAO, and W.W UNICEF. "The state of food security and nutrition in the world", *Food and Agriculture Organization of the United Nations*, Rome, Italy, 2017.

[8] T.F. Moberg, J. S. Ramberg, and R. H. Randles, “An Adaptive Multiple Regression Procedure Based on M-estimators”, *Tech*. Vol. 22, No. 2, PP. 213- 224, 1980. Inc., 1221 Avenue of the Americas, New York

[9] W. Sanford, Applied Linear Regression, 3rd edition, John Wiley & Sons, Inc. Hoboken, New Jersey, Canada, 2005.