Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 2799–2808
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2023.29395.4148

Robust estimates for a three-parameter exponential regression model

Ahmed Joudah Irshayyid*, Rabab Abdulrida Saleh

Statistics Department, College of Administration and Economics, University of Baghdad, Iraq

(Communicated by Ehsan Kozegar)

Abstract

Exponential regression is one of the most common and widely used models in several fields to estimate the parameters of the exponential regression model using ordinary nonlinear least square but this method is not effective in the presence of outlier values so robust methods were used to treat outlier values in this research exponential regression model are used to estimate the parameters using robust method (Median-of-Means, Forward search, M-Estimation), and the simulation method was used to compare the estimation methods with different sample sizes and assuming four percentages of the outliers of the data (10%, 20%, 30%, 40%). And through the mean square error (MSE) was made to reach the best estimation method for the parameters, where the results obtained using the simulation method showed that the forward search is the best because it gives the lowest mean of error squares.

Keywords: exponential regression model, robust estimates, Median-of-Means, Forward search, Least Absolute Quantile Estimates
2020 MSC: 62K25,62J02, 11L07, 62E15

1 Introduction

The Exponential regression model is one of the nonlinear regression models when the relationship between the variables is exponential, that is, the model parameters appear Exponential or the variables appear Exponential. Classical methods are used to estimate regression parameters, however, in the presence of outliers affecting the data, these methods are not useful for data analysis, Ordinary estimation methods can lead to misleading values for Exponential regression parameters, and the estimates may not be reliable, and the relationship between response and explanatory variables is oblique, then predictions are biased. To address this problem and in order to reduce errors resulting from parameter estimation, new statistical tools have been developed that are not easily affected by outliers, it is the use of robust regression methods. We need robust statistical estimation methods for Exponential regression models due to the presence of influential outliers values to be a more accurate estimation.

For instance The principle of median-of-means (MOM) in (2021), the researcher (Pengfei) [6] and others used the median-of-means method to estimate the coefficients of the exponential regression model and in the presence of outliers, and it was compared with the nonlinear least squares method, and the estimator (MOM) was more efficient. In (2010), the researcher (Atkinson) [3] and others used the forward search method as a general and robust method,

*Corresponding author
Email addresses: Ahmed.jawda1201a@coadec.uobaghdad.edu.iq (Ahmed Joudah Irshayyid), rabab.saleh@coadec.uobaghdad.edu.iq (Rabab Abdulrida Saleh)

Received: November 2022 *Accepted:* January 2023

and this method includes dividing the data into small subtotals to discover the abnormal values affecting the estimation of the parameters, then close observations are added to the model to monitor the estimation of the parameters with Increasing the sub-group by measuring the errors and choosing the sub-group that has the least error as it was applied to the daily ozone concentration data in the extreme hot spot.

2 Meyer7 model (three-parameter exponential regression model)

Suggested by researchers (MEYER and ROTH) [7] in 1971, It was applied to experimental data. This model was also applied to PCB residue concentration data in a series of trout from Cayuga Lake in New York by researcher (Achcar and Lopes) [1], The form is as follows:

$$y_i = \beta_0 + \beta_1 e^{\beta_2 x_i} + \varepsilon \tag{2.1}$$

where $y = [y_1, y_2, ..., y_n]^T$ is response vector $n * 1$.

x_i: They are the explanatory variable in the model.

ε_i: The error is a small value and it is the random part in the Meyer7 model

$\beta_0.\beta_1.\beta_2$ are the model parameters.

the model Meyer7 parameters can be estimated using the nonlinear least squares method (Newton-Raphson method) and the general formula for this method [9]

$$\underline{\beta}_{(i+1)} = \underline{\beta}_{(i)} - \left(J_{(i)}\right)^{-1} * \underline{f}_{(i)} \tag{2.2}$$

where i: is the number of iterations.

$\underline{\beta}_{(i+1)}$: A vector of degree $(2 * 1)$, where 2 represents the number of parameters.

$J_{(i)}$: Matrix of partial derivatives.

$\underline{f}_{(i)}$: Vertical Vector Equations.

$\underline{\beta}_{(i)}$: An initial or hypothetical estimate of parameters is obtained by arithmetic or intelligent guesswork.

When $i = 1$ the vector will be the initial default value, and when $i = 2$ it means that $\underline{\beta}_{(2)}$ will be in the new default value which are the parameter values estimated in the second iteration $\underline{\beta}_{(i+1)}$, In the event that the suspension condition is not met:

$$\left|\underline{\beta}_{(i+1)} - \underline{\beta}_{(i)}\right| \leq 0.01$$

The iteration continues and the default parameter vector values change until the condition is met or we stop so that we cannot continue the iteration because the determinant of the estimated parameter matrix reaches zero [9].

The sum of the squared errors can be calculated using the least squares method, according to the following equation:

$$h(\beta_0.\beta_1.\beta_2) = \sum_{i=1}^{n} \varepsilon i^2 = \sum_{i=1}^{n} (yi - \beta_0 - \beta_1 e^{\beta_2 x_i})^2 \tag{2.3}$$

We differentiate equation (2.3) for $\beta_0.\beta_1.\beta_2$ and then set the derivatives to zero to obtain the estimations of $\beta_0.\beta_1.\beta_2$.

$$a = \frac{\partial h(\beta_0.\beta_1.\beta_2)}{\partial \beta_0} = -2\sum_{i=1}^{n}(yi - \beta_0 - \beta_1 e^{\beta_2 x_i}) = 0 \tag{2.4}$$

$$\begin{aligned} b = \frac{\partial h(\beta_0.\beta_1.\beta_2)}{\partial \beta_1} &= 2\sum_{i=1}^{n}(yi - \beta_0 - \beta_1 e^{\beta_2 x_i}) * -e^{\beta_2 x_i} = 0 \\ &= -2\sum_{i=1}^{n} yie^{\beta_2 x_i} + 2\sum_{i=1}^{n} \beta_0 e^{\beta_2 x_i} + 2\sum_{i=1}^{n} \beta_1 e^{2\beta_2 x_i} \end{aligned} \tag{2.5}$$

$$c = \frac{\partial h(\beta_0.\beta_1.\beta_2)}{\partial \beta_2} = 2\sum_{i=1}^{n}(yi - \beta_0 - \beta_1 e^{\beta_2 x_i}) * -\beta_1 x_i e^{\beta_2 x_i} = 0$$

$$= -2\sum_{i=1}^{n} yix_i\beta_1 e^{\beta_2 x_i} + 2\sum_{i=1}^{n}\beta_0\beta_1 x_i e^{\beta_2 x_i} + 2\sum_{i=1}^{n} x_i\beta_1^2 e^{2\beta_2 x_i} \tag{2.6}$$

We derive an equation (2.4) for $\beta_0.\beta_1.\beta_2$

$$\frac{\partial a}{\partial \beta_0} = 2n \tag{2.7}$$

$$\frac{\partial a}{\partial \beta_1} = 2\sum_{i=1}^{n} e^{\beta_2 x_i} \tag{2.8}$$

$$\frac{\partial a}{\partial \beta_2} = 2\sum_{i=1}^{n} \beta_1 x_i e^{\beta_2 x_i} \tag{2.9}$$

We derive an equation (2.5) for $\beta_0.\beta_1.\beta_2$

$$\frac{\partial b}{\partial \beta_0} = 2\sum_{i=1}^{n} e^{\beta_2 x_i} \tag{2.10}$$

$$\frac{\partial b}{\partial \beta_1} = 2\sum_{i=1}^{n} e^{2\beta_2 x_i} \tag{2.11}$$

$$\frac{\partial b}{\partial \beta_2} = -2\sum_{i=1}^{n} yix_i e^{\beta_2 x_i} + 2\sum_{i=1}^{n}\beta_0 x_i e^{\beta_2 x_i} + 4\sum_{i=1}^{n}\beta_1 x_i e^{2\beta_2 x_i} \tag{2.12}$$

We derive an equation (2.6) for $\beta_0.\beta_1.\beta_2$

$$\frac{\partial c}{\partial \beta_0} = 2\sum_{i=1}^{n} \beta_1 x_i e^{\beta_2 x_i} \tag{2.13}$$

$$\frac{\partial c}{\partial \beta_1} = -2\sum_{i=1}^{n} yix_i e^{\beta_2 x_i} + 2\sum_{i=1}^{n}\beta_0 x_i e^{\beta_2 x_i} + 4\sum_{i=1}^{n} x_i\beta_1 e^{2\beta_2 x_i} \tag{2.14}$$

$$\frac{\partial c}{\partial \beta_2} = -2\sum_{i=1}^{n} yix_i^2\beta_1 e^{\beta_2 x_i} + 2\sum_{i=1}^{n}\beta_0\beta_1 x_i^2 e^{\beta_2 x_i} + 4\sum_{i=1}^{n} x_i^2\beta_1^2 e^{2\beta_2 x_i} \tag{2.15}$$

Now we apply the equation (2.2) of Newton-Raphson's method and it becomes as follows:

$$(\hat{\beta}_0\hat{\square}_1\hat{\square}_2) = (\beta_0\square_1\square_2) - \left(\frac{\partial a}{\partial \beta_0}\frac{\partial a}{\partial \beta_1}\frac{\partial a}{\partial \beta_2}\frac{\partial b}{\partial \beta_0}\frac{\partial b}{\partial \beta_1}\frac{\partial b}{\partial \beta_2}\frac{\partial c}{\partial \beta_0}\frac{\partial c}{\partial \beta_1}\frac{\partial c}{\partial \beta_2}\right)^{-1} * (a\ b\ c) \tag{2.16}$$

3 Robust estimation methods

We use classical methods to estimate parameters in a Meyer7 exponential model but in the presence of outliers they are unsafe, therefore, we use robust methods because they are not affected by outliers, and can be used even if the basic assumptions are violated or the evaluation and test conditions are not met, and there are many Powerful methods for estimating parameters of the Meyer7 model exponential. The goal of estimation by powerful methods is to find new statistical methods that are less sensitive and prone to abnormal values and get the best estimation results.

3.1 The Median-of-Means method

This estimator is used systematically to build robust estimates. The MOM methods estimate parameters for β is using the following steps:

The first step: We separate $(y_i.x_i)$. $i = 1, 2, ..., n$ into g of groups, the number of observations in each group is equal to T as in the equation below [5]:

$$T = \frac{n}{g} \tag{3.1}$$

For ease of calculation, let's say n is always divisible by g, Then we extract the number of groups g from the following equation [10]:

$$g = 8 * \log\log\left(\frac{1}{\zeta}\right) \tag{3.2}$$

where $\{\zeta \in (0, 1)\}$ and g are rounded to the nearest number in the positive direction, The structure of observations is always undefined, and diagnosing outliers is complicated, so ζ is determined from the following equation [5]:

$$\zeta = \frac{C}{\sqrt{n}} \tag{3.3}$$

C: A positive integer regardless of outliers and equal to.

$$c = tr(\sum) \tag{3.4}$$

$tr(\sum)$: The order of the covariance and covariance matrix [6].

Step two: We estimate the parameters β in each group j by a nonlinear least squares method so that $j = 1, 2, ..., g$.

$$\hat{\beta}^{(j)} = \left(\beta_0^{(j)}.\beta_1^{(j)}, \beta_2^{(j)}\right)^T \tag{3.5}$$

The third step: We find the estimator MOM

$$\hat{\beta}^{MOM} = \left(\hat{\beta}_0^{MOM}.\hat{\beta}_1^{MOM}.\hat{\beta}_2^{MOM}\right)^T \tag{3.6}$$

So that the median of each parameter is found in the following way:

$$\hat{\beta}_0^{MOM} = median\left(\hat{\beta}_0^1.\hat{\beta}_0^2. \dots .\hat{\beta}_0^g\right)^T \tag{3.7}$$

$$\hat{\beta}_1^{MOM} = median\left(\hat{\beta}_1^1.\hat{\beta}_1^2. \dots .\hat{\beta}_1^g\right)^T \tag{3.8}$$

$$\hat{\beta}_2^{MOM} = median\left(\hat{\beta}_2^1.\hat{\beta}_2^2. \dots .\hat{\beta}_2^g\right)^T \tag{3.9}$$

3.2 Forward search method

It is a robust data analysis method used for parameter estimation, which starts from a small selected subset of data and monitors the effect of adding observations to the subset until all data is added at the end. Thus, we avoid the influence of outliers, and estimating parameters is a very reliable method, suppose we have $Z_i = (y_i.x_i)$, Which model contains p parameters, the forward search algorithm starts by choosing a subset of size m where $(m = p)$, the subset of size m to $m + 1$ is increased at each step to form a new subset of observations, where the subgroups are $p \leq m \leq n$ and Z_m is a vector of the observations and the sizes of the groups are as follows [2]:

$$S_{i1...im}^{(m)} = \{Z_m.Z_{m+1}.Z_{m+2}. \dots .Z_n\} \tag{3.10}$$

The parameters are estimated by the least squares method for each sub-group so that we have a set of forward search parameters $\hat{\beta}_{Fc}$ where [3]:

$$\hat{\beta}_{Fc} = \left(\hat{\beta}_m.\hat{\beta}_{m+1}. \dots .\hat{\beta}_n\right) \tag{3.11}$$

We find the random errors by replacing all the parameters of the subset with the main observations [8]

$$\begin{aligned} \grave{y}_{im} &= \hat{\beta}_{0mi*} + \hat{\beta}_{1mi*}e^{\hat{\beta}2_{mi*}x_i} \\ e_{i_m} &= y_i - \hat{y}_{im} \qquad i = 1.2. \dots n \end{aligned} \tag{3.12}$$

We choose the parameters that have the least random errors.

$$e_i = (e_{i_m}) \tag{3.13}$$

3.3 Least Absolute Quantile Estimates

It is one of the estimation methods based on minimizing the sum of absolute errors. It is much better than the least squares method when the error is not distributed normally. This method replaces the error squares with absolute values, the absolute least estimated value (LA) can be obtained as follows [4]:

$$\hat{\beta}_{LA} = \arg\arg \sum_{i=1}^{n} |r_i(\beta)| \tag{3.14}$$

The error is calculated from the following equation:

$$r_i(\beta) = y_i - \grave{y}_i.i = 1 \dots n \tag{3.15}$$

$$\frac{\partial |r_i(\beta)|}{\partial \beta} = |\dot{\eta}(\beta)| \tag{3.16}$$

where $\dot{\eta}(\beta)$: Derivative a three-parameter exponential regression function,

$$\dot{\eta}(\beta) = [1x_ie^{\beta}2^{x_i}x_i\beta_1e^{\beta}2^{x_i}]_{n*3}$$

To get the estimator $\hat{\beta}$ to solve equation (3.14) It is a reformulation of the minimization problem as a linear programming problem through a linear approximation of the three-parameter exponential regression. Necessary condition that there is a vector $d \in [-1, 1]^n$, to help solve a series of problems to reduce errors:

$$d_i = sgn(r_i) \tag{3.17}$$

where $sgn(r_i)$ error signal

When applying the vector d, we will face a problem of finding the estimation of the parameters and to solve this problem we apply the Meketon algorithm, we find through it the amount of $\hat{d}$ through the following equation

$$\hat{d} = \left[I_n - \dot{\eta}(\beta)(\dot{\eta}'(\beta)_{p*n}\dot{\eta}(\beta)_{p*n})^{-1}_{p*p}\dot{\eta}'(\beta)\right]_{n*n} d_{n*1} \tag{3.18}$$

where I_n is a ones matrix of size $n * n$

We find the D, and it is a diagonal matrix through the following equation:

$$D = diag(\min\{1 - \hat{d}_i.1 + \hat{d}_i\}) \tag{3.19}$$

We find the parameters of the three-parameter exponential regression model according to the following equation:

$$\hat{\beta}_{LA} = \left(\dot{\eta}'(\beta)_{p*n} D^2_{n*n} \dot{\eta}(\beta)_{n*p}\right)^{-1} \dot{\eta}'(\beta)_{p*n} D^2_{n*n} y_{n*1} \tag{3.20}$$

We calculate s from the following equation:

$$s = D^2 \left[I_n - \dot{\eta}(\beta)(\dot{\eta}'(\beta) D^2 \dot{\eta}(\beta))^{-1} \dot{\eta}'(\beta) D^2\right] * y \tag{3.21}$$

We calculate α from the following equation:

$$\alpha = \max\left\{\max\left\{\frac{s_i}{(1-d_i)} \cdot \frac{-s_i}{(1+d_i)}\right\}\right\} \tag{3.22}$$

We calculate the new d from the following equation:

$$d = d + \left(\frac{c}{\alpha}\right) s \tag{3.23}$$

Where $c = 0.97$ and we stop in the solution when

$$|B_i - B_{i-1}| \cong 0.01$$

4 Simulation

The simulation experiments for this study included writing a number of programs in the language (MATLAB) to generate simulated data for the purpose of comparing methods with different sample sizes. ($n1 = 25, n2 = 50, n3 = 100, n4 = 150$), in generating data for random variables and different outliers ratio, each experiment was repeated 1000 times, Taking into account the exponential model that has been relied upon. The stages of the simulation experiment are described through the following steps:

1. Set default values for parameters, and it is one of the important and basic steps on which exponential models depend, where ($\beta_0 = 15.5, \beta_1 = 1.2, \beta_2 = 0.02$) was chosen.
2. Generating the values of the independent variable from the uniform distribution based on the inverse method based on the (CDF) function of the exponential distribution, where the function is derived to obtain the independent variable. According to the following steps:

$$f_{(x)} = \lambda e^{-\lambda x},\ x \geq 0 \tag{4.1}$$

$$F_{(x)} = 1 - e^{-\lambda x},\ x \geq 0 \tag{4.2}$$

where the simulation works, in general, to generate random numbers with a uniform distribution within a period $u(0,1)$, and depending on these numbers, we generate random variables, and thus the distribution of random numbers approaches the standard uniform distribution, and it is called the inverse transformation method, as follows:

$$\begin{aligned} u &= 1 - e^{-\lambda x} \\ 1 - u &= e^{-\lambda x} \\ x &= \frac{-\log\log(1-u)}{\square} \end{aligned} \tag{4.3}$$

where we assume that $\lambda = 0.03$. The case that the data does not contain outliers value the error is generated according to the following equation $\varepsilon = \tau \sim \exp(0,3)$ and in the case of the presence of outliers value the error is generated according to the following equation

$$\varepsilon = \tau \sim \exp(0,3) + (\tau) \sim \exp(10) \tag{4.4}$$

where the rate of outlier values are ($\tau = 10\%, 20\%, 30\%, 40\%$)
Calculating the dependent variable (y), the following exponential model are applied

$$y_i = \beta_0 + \beta_1 e^{\beta_2 x_i} + \varepsilon$$

3. The estimation methods are compared on the statistical error scale of error squares (MSE).

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{4.5}$$

4.1 Analyzing the results

1. It was noted that the values of the mean squares of error were close for all the estimation methods and for all ratio of outliers. The methods can be considered equivalent, and all the methods are characterized by the accuracy of the estimation because they gave reliable results.
2. It is clear from Table 1 the results of the sample size (n = 25) and through the (MSE) scale of the Meyer7 model, it is clear that the forward search (FS) method is the best estimation method for all ratio outliers because it outperformed all methods and achieved the lowest (MSE), and the results indicate The simulation clearly indicates the robust of the forward search method in relation to other methods when the sample size is (n = 25) and the presence of influential outliers.
3. It is clear from Table 1 the results of the sample size (n = 50) and through the (MSE) scale of the Meyer7 model, it is clear that the forward search method (FS) is the best estimation method for all ratio outliers because it outperformed all methods and achieved the lowest (MSE), and the results indicate The simulation clearly indicates the robust of the forward search method compared to other methods when the sample size (n = 50) and the presence of influential outliers.
4. It is clear from Table 1 the results of the sample size (n = 100), and through the scale (MSE) of the Meyer7 model, it is clear that the forward search method (FS) is the best estimate method for ratio outliers (10%, 20%, 30%, 40%) because it It outperformed all methods and achieved the lowest (MSE), while the Least Absolute Quantile Estimates method was the best for the percentage of pollution (0%) because it outperformed all methods and achieved the lowest (MSE), and the simulation results clearly indicate the robust of the forward search method and the Least Absolute Quantile Estimates method For other methods, when the sample size is (n = 100) and the presence of influential outliers.
5. It is clear from Table 1 the results of the sample size (n = 150), and through the (MSE) scale of the Meyer7 model, it is clear that the forward search method (FS) is the best estimation method for all ratio outliers because it outperformed all methods and achieved the lowest (MSE), and the results indicate The simulation clearly indicates the robust of the forward search method in relation to other methods when the sample size (n = 150) and the presence of influential outliers.

5 The Real Data Analysis

In order to achieve the objective of the research and verify the performance of robust methods for treating outliers in real data, this part is devoted to the study of real data, especially for revenues (x) Public revenues are defined as all amounts collected by the government during a certain period of time, whether they are revenues from taxes, fees, grants or loans. and expenditures (y) Public expenditures are defined as all expenditures incurred by the government during a certain period of time, obtained from the Ministry of Finance and then apply the parameter estimation methods in the exponential regression model to the real data to reach the best method through the mean squares error (MSE).

A sample size (155 views) was selected from January 2009 to November 2021, where the variable y expenditures were tested by (easyfit) program, where the dependent variable is distributed exponentially with a parameter equal to $\lambda = 0.03$ as in Figure 1

By plotting the boxplot in Figure 2, we notice that there are outliers in the real data for variable y and variable x.

Some tests and measures have been conducted for the Meyer7 model to ensure the integrity of the data before estimating the parameters of the model using robust methods and to ascertain whether the data follow a normal distribution or not, and to detect the problem of heterogeneity of error variance and to detect the presence of abnormal values in the data. By the least squares method, the model was as follows:

$$y_i = 37.217 + 0.0022e^{2.181e-6*xi}, \quad i = 1, ..., 155$$

We note that an increase in revenues by one unit affects an increase in expenditures by $22 = 37.217 + 0.0022e^{2.181e-6}$

Table 1: It shows the (MSE) values of the Meyer7 model

size	percentage outliers	MSE			
		ols	MOM	FS	LA
N=25	0%	4.427e-7	1.373e-7	1.383e-8	1.930e-5
	10%	1.273e-4	2.995e-7	2.885e-7	0.002659
	20%	0.003535	1.756e-6	6.301e-7	0.820640
	30%	5.811e-4	1.756e-6	6.301e-7	0.053860
	40%	1.507e-4	1.756e-6	6.302e-7	0.971253
N=50	0%	1.310e-7	6.342e-8	3.777-8	1.292e-6
	10%	0.009134	0.004328	0.004311	0.052011
	20%	0.007972	0.001158	0.001137	0.052046
	30%	0.005926	0.001158	0.001137	0.052081
	40%	0.003641	0.001158	0.001137	0.135734
N=100	0%	0.029548	9.107e-7	2.294-7	1.755e-8
	10%	3.551e-4	2.627e-5	2.588e-5	7.523e-4
	20%	0.007915	0.004973	0.004365	0.026331
	30%	0.007421	0.004789	0.004365	0.921585
	40%	0.006970	0.004528	0.004365	0.921620
N=150	0%	7.954e-9	1.227e-8	7.802e-9	2.562e-5
	10%	1.322e-4	9.826e-6	9.251e-6	0.076162
	20%	8.770e-5	6.156e-4	5.791e-5	8.184e-4
	30%	5.414e-5	7.035e-4	6.020e-5	0.028644
	40%	1.083e-4	0.002111	6.021e-5	0.229152

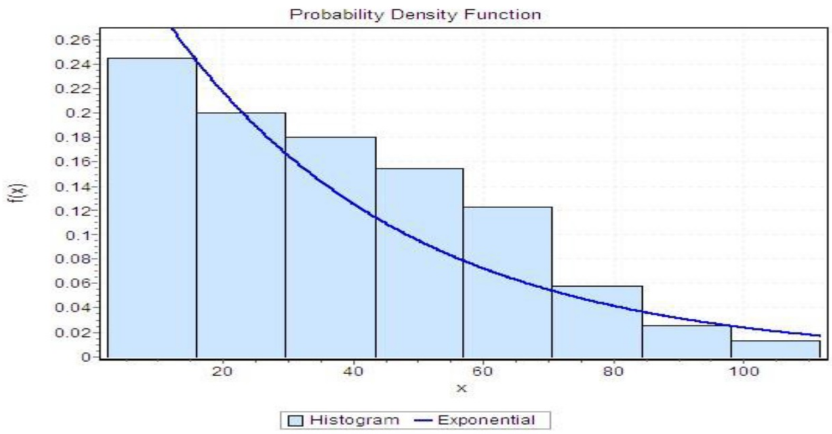

Figure 1: Shows the exponential distribution of the dependent variable for (n = 155)

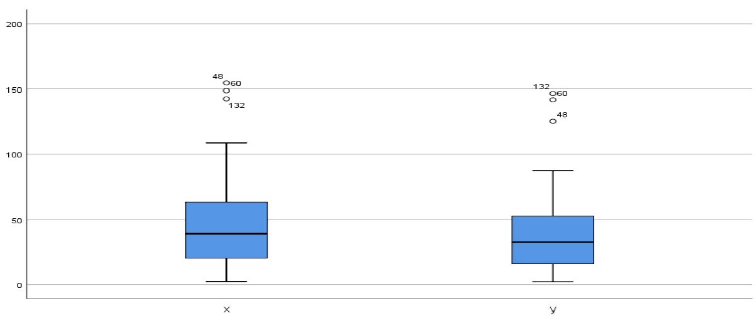

Figure 2: Shows the outliers of the dependent variable and the independent variable

The Jarque-Bera test was used to test the normal distribution of random error for the Meyer7 model and it was found that the random error does not follow the normal distribution, where the test value was (h =1), which means rejecting the null hypothesis with a degree of freedom (0.05), that is The data is incorrect, i.e. suffers from the presence of heterogeneity of variance or the presence of outliers values in the data that were detected by the boxplot, so we resort to the robust methods for estimating the parameters of Meyer7 model.

Gold field Quandt test was used to detect the problem of heterogeneity of error variance of the Meyer7 model. It turns out that the error suffers from the problem of heterogeneity of variance, as in Table 2:

Table 2: shows the heterogeneity test for Meyer7 model

test	**F**	**F(78,78,0.05)**
Gold field-Quant	4.0162	1.47

Arithmetic methods were used to detect and check the data the presence of outliers in the Meyer7 model, where Cook's method was used, as shown in Table 3.

Table 3: shows the outliers and influential observations in the Meyer7 model

data	**1**	**4**	**25**	**35**	**36**	**43**	**46**	**47**	**48**	**60**	**70**	**72**	**74**	**97**	**109**	**120**	**132**	**145**	**155**
Meyer7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

It is clear from Table 3 that there are nineteen values affecting the parameters of the model the Meyer7 according to the Cook scale. Which affects the estimation of parameters of the Meyer7 model, so we use the robust methods.

The parameters of the Meyer7 model are estimated using the best methods that were produced by the experimental side of the research through the mean of the error boxes of the model, which was the method of forward search with different the ratio of outliers, because it gave less (MSE) for the model than other methods when (n=155). As shown in Table 4

Table 4: shows the value of the parameters of the model the Meyer7 model of the data, expenditures and revenues

method	**B0**	**B1**	**B2**	**MSE**	**Comparison**
OLS	37.217	0.0022	2.181e-6	15.936	
MOM	46.43	0.5427	0.0025	10.218	
FS	46.431	1.0583	0.0151	9.171	best estimator
LA	35.952	1.141e-61	-7.39e-64	16.759	

That the robust forward research method is characterized by its ability of strength and robustness and achieved the highest resistance to outliers than the rest of the methods and is characterized by the accuracy of the estimate because it gave reliable results. We note that an increase in revenues by one unit affects an increase in expenditures by $46.431 + 1.0583e^{0.0151} = 47.505$ as shown in the equation below:

$$y_i = 46.431 + 1.0583e^{0.0151x_i}, i = 1, ..., 155$$

6 Conclusions

1. The simulation results showed, in general, that the forward search method proved its efficiency in estimating exponential regression model in case of data outliers, as it achieved the lowest mean square error criterion compared to the rest of the robust estimation methods with an increase in the ratio of outliers.
2. It appeared to us through the analysis of expenditure and revenue data and using the exponential regression model, we note the fit of the data to the model
3. The mean squared error decreases as the sample size increases.

References

[1] J.A. Achcar and S.R.C. Lopes, *Linear and non-linear regression models assuming a stable distribution*, Colomb. J. Statist. **39** (2016), no. 1, 109–128.

[2] A.C. Atkinson, A. Corbellini and M. Riani, *Robust Bayesian regression with the forward search: theory and data analysis*, Test **26** (2017), no. 4, 869–886.

[3] A.C. Atkinson, M. Riani and A. Cerioli, *The forward search: Theory and data analysis*, J. Korean Statist. Soc. **39** (2010), 117–134.

[4] R. Koenker and B.J. Park, *An interior point algorithm for nonlinear quantile regression*, J. Econometr. **71** (1996), 265–283.

[5] M. Lerasle, *Selected topics on robust statistical learning theory lecture notes*, Paris-Saclay University, 2019.

[6] P. Liu, M. Zhang, R. Zhang and Q. Zhou, *Robust estimation and tests for parameters of some nonlinear regression models*, J. Math. **9** (2021).

[7] R. Meyer and P.M. Roth, *Modified damped least squares: an algorithm for non-linear estimation*, J. Inst. Maths. Appl. **9** (1972), 218–233.

[8] M. Rian and A.C. Atkinso, *Robust diagnostic data analysis: transformations in regression' technometrics*, Amer. Statist. Assoc. Amer. Soc. Qual. **42** (2000), no. 4, 384–394.

[9] G.A.F. Seber and C.J. Wild, *Nonlinear regression*, John Wiley & Sons Inc., 2003.

[10] Y. Zhang and P. Liu, *Median-of-means approach for repeated measures data*, Commun. Statistics-Theory Meth. **50** (2021), no. 17, 3903–3912.