INTRODUCTION

Stephen and Senthamari (2017) defined Regression analysis as a statistical technique for analyzing and modeling the relationship between dependent variable and one or more independent variables. This technique uses the mathematical equation to establish the relationship between variables. It is a predictive modeling technique used for forecasting and to find causal effect relationship between the variables. Outlier is a data point that differs significantly from other observations. An outlier may be due to variability in the measurement or it may indicate experimental error, the latter are sometimes excluded from the dataset. An outlier can cause serious problem in statistical analysis (Zimek and Filzsimner, 2018). Outlier detection has many applications, such as data cleaning, fraud detection and network intrusion. The existence of outliers can indicate individuals or groups that have behavior very different from most of the dataset. Frequently, outliers are removed to improve accuracy of the estimators. But sometimes the presence of an outlier has a certain meaning, which explanation can be lost if the outlier is deleted (Hawkins 1980).

The Classical Linear Regression Model

The classical linear regression model is a statistical model that describes a data generation process. The classical linear regression model of the form

\[ Y = X\beta + \varepsilon \]  

Where \( Y \) is an \( n \times 1 \) vector of observed response values, \( X \) is the \( n \times p \) matrix of the predictor variables, \( \beta \) is the \( p \times 1 \) vector which contains the unknown parameters and needs to be estimated, and \( \varepsilon \) is the \( n \times 1 \) vector of random error terms.

Assumptions of classical linear regression model

1. The dependent variable is linearly related to the coefficients of the model and the model is correctly specified. i.e. \( Y = X\beta + \varepsilon \)
2. The independent variable(s) is/are uncorrelated with the equation error term. i.e. \( \text{Cov}(X,\varepsilon) = 0 \)
3. The mean of the error term is zero. i.e. \( E(\varepsilon) = 0 \)
4. The error term has a constant variance (Homoscedastic error). i.e. \( \text{var} (\varepsilon) = \sigma^2 I \)
5. The error terms are uncorrelated with each other. No autocorrelation or serial correlation. i.e. \( \text{cov}(\varepsilon_i,\varepsilon_j) = 0 \) for \( i \neq j \)
6. There is no perfect linear relationship between the independent variables.
7. The error term is normally distributed.
8. There is absence of outlier in the dataset

Applying Ordinary Least Squares Estimators (OLSE) in simple or multiple linear regressions always calls for some assumptions: normality of the error terms; equal variance of the error terms, and absence of outliers, leverage points and Multicollinearity. According to Hampel (2001) and Huber (1972), normality of the error distributions finds its basis from the central limit theorem; which is a limit theorem based on approximations. Additionally, outliers in the dependent variable, lead to large residual values which further results in the failure of the normality assumption of the error terms. Therefore, in regression analysis, the ordinary Least Squares estimation is the best method if the assumptions are met. However, if these assumptions are not satisfied, the results can easily be affected (Alma, 2011). One of the first steps towards obtaining a coherent analysis is the detection of outlying observations. Although outliers are often considered as an error or noise, they may carry important information. Detected outliers are candidates for aberrant data that may otherwise adversely lead to model misspecification.

ABSTRACT

The paper examined the performance of five estimation methods using six different outlier percentages (0%, 5%, 10%, 20%, 30% and 40%) and five different sample sizes (20, 40, 60, 100 and 200) were used to investigate effect of sample size on the performance of each of the estimation methods. The study adopted absolute bias, variances, relative efficiency and root mean square errors as comparison criteria through Monte-Carlo experiment and real life data was used to validate the simulation results. The study found that, under 5%, 10%, 20% and 30% outlying condition Robust-MM is the most preferred estimator across all criteria and sample size except using relative efficiency and when the sample size is 20 and 200 under 5%, 20% and 30% outlying condition and using absolute bias criterion respectively while Robust-LS is the least preferred estimator except when the sample size is 40, 20 ; 40, 20 ; 20, 200 under 5%, 20% and 30% outliers and using absolute bias, variance and root mean square error respectively. Under 40% outlying condition Robust-MM is the most preferred estimator across all criteria and sample size except using relative efficiency and when the sample size is 20. Furthermore, Robust-MM is the most consistent estimator across the comparison criterion except when using relative efficiency and sample size has little or no effect on the performance of the estimators across all the different outlier levels. R Statistical package was used for the data analysis. This study therefore recommends the used of Robust-MM estimator.

Keywords: Estimation, Estimator, Outliers, Performance, Regression, Robust
biased parameter estimation and incorrect results. It is therefore important to identify them prior to modeling and analysis (Williams et al. 2002; Liu et al. 2004).

Among methods used in detecting the presence of outliers are graphical methods and scatter plot. In the situation where the assumptions of the linear regression are not met, robust regression estimator is an important estimation technique for analyzing data that are contaminated with outliers or data with non normal error term.

Many estimation methods such as Least Trimmed Squares Estimator (LTSE), M-Estimator (ME), S-Estimator (SE), Modified Maximum Likelihood Estimator (MMLE) have been proposed which are more efficient than the Ordinary Least Squares (OLS) when there is outlier.

In regression analysis, the application of ordinary least square method works well if the assumptions of the regression model, variables and the error terms are met. However, the presence of outliers or failure of the assumptions renders the ordinary least square method of estimation unreliable. This is because good leverage points, vertical outliers and good leverage points can influence the coefficients in the model, the residuals, as well as the standard errors of the model and the coefficients (David, 2014).

Several estimation procedures have been proposed in literature to handle the problem of outliers during parameter estimation. Therefore, the paper examined the performances of five robust estimators using different percentages of outlier conditions with varying sample sizes say; 20, 40, 60, 100 and 200 and identified the most preferred estimator based on each of the selected comparison criteria.

Least Absolute Deviation (LAD)
This estimator obtains a higher efficiency, instead of minimizing the sum of squared errors; it minimizes the sum of absolute values of errors. The LAD method is not sensitive to outliers and produces robust estimates. (DasGupta and Mishra, 2004).

\[
\min \sum_{i=1}^{n} |e_i| \tag{2}
\]

\[
\min \sum_{i=1}^{n} |y_i - x_i \beta| \tag{3}
\]

M-Estimator (ME)
One of the robust regression estimation methods is the M estimation. M estimation is an estimation of the maximum likelihood type. M-estimation is an extension of the maximum likelihood estimate method and a robust estimation.

\[
\hat{\beta}_1 = (X^T WX)^{-1} X^T W y \tag{4}
\]

Least Trimmed Squares Estimator (LTSE)
Rousseeuw (1984) developed the least trimmed squares estimator (LTSE) given by,

\[
\hat{\beta} = \min \sum_{i=1}^{n} (e_i^2) \tag{5}
\]

Where \( q = \lfloor n (1 - \alpha) + 1 \rfloor \) is the number of observations included in the calculation of the estimator, and \( \alpha \) is the proportion of trimming that is performed. Using \( q = \left( \frac{5}{7} \right) + 2 \) ensures that the estimator has a breakdown point of 50%.

S – Estimation (SE)
The S-estimation is a high breakdown method introduced by Rousseeuw and Yohai (1984) that minimizes the dispersion of the residuals. The S-estimator was introduced to take care of the low breakdown point of the M-estimators.

\[
\frac{1}{n} \sum_{i=1}^{n} \rho \left( \frac{e_i}{s} \right) = k \tag{6}
\]

MM- Estimation (MME)
MM estimation is a special type of M-estimation developed by Yohai (1987). MM-estimation is a combination of high breakdown value estimation and efficient estimation. MM estimator was the first estimate with a high breakdown point and high efficiency under normal error.

\[
\frac{1}{n-1} \sum_{i=1}^{n} \rho \left( \frac{y_i - X_i \beta}{s} \right) = 0.5 \tag{7}
\]

MATERIALS AND METHOD
Data Generation and Model Formulation Procedure
The dataset used for this study was simulated using Monte Carlo in the environment of R statistical package (www.cran.org).

Mechanism for Generating the Independent Variables
In this study, three regressors were simulated, where two of the regressors were simulated to be normally distributed with mean zero and variance 1 and the other one was simulated with different percentages of outlier. The procedure is:

\[ X_{t1} \sim N(0,1) \]

Where \( t = 1, 2, 3,..., n; i = 1, 2, 3 \).

\[ X_{t2} \sim (1 - n1\%) N(0,1) + n1\% N(0,500) \]

Where \( n1\% \in \{0\%, 5\%, 10\%, 20\%, 30\%, 40\%\} \) is the percentage of outliers to be injected in the third predictor variable.

Mechanism for simulating Model with Outlier(s) in the Error Term.
The error term \( e_t \) was simulated to be distributed according to a Gaussian mixture, i.e \( e_t \sim (1 - n1\%) N(0,1) + n1\% N(0,500) \).

Where \( n1\% \in \{0\%, 5\%, 10\%, 20\%, 30\%, 40\%\} \) is the percentage of outliers injected in the error term.

Mechanism for Generating the Dependent Variable
The dependent variable \( Y_t \) was simulated to be distributed according to a Gaussian mixture, i.e \( Y_t \sim (1 - n1\%) N(0,1) + n1\% N(0,500) \).

The response variable is obtained from the relation given by:

\[ Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + ... + \beta_p X_{tp} + e_t \tag{8} \]

\( t = 1,..., n \) and \( i = 1,2,3 \).

Data Simulation
A Monte-Carlo experiment of 1000 trials was carried out for five sample sizes (20, 40, 60, 100 and 200) each with different percentage of outliers (0%, 5%, 10%, 20%, 30%, 40%). Five Robust estimators were used to estimate parameters that were fitted to this simulated data. Real life data was used to validate our findings from the simulation study.
Criteria for Evaluating the Estimators
The assessments of the estimators considered in this work was based on the following criteria.

Root Mean Square Error (RMSE)
Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). The formula is given by:

\[ \text{RMSE} = \sqrt{E(\beta - \hat{\beta})^2} = \sqrt{MSE(9)} \]

Bias
The bias is measured by

\[ \text{Bias} = \hat{\beta} - \beta \]

The smaller the bias of an estimator the better.

Variance
The variance of an estimator \( \beta \) for \( \beta \) is defined as

\[ \text{Var}(\^\beta) = E(\hat{\beta} - E(\beta))^2 \]

(10)

Relative efficiency of an estimator
The efficiency of an estimator is its minimum possible variance to its actual variance, and this estimator is efficient when the ratio gives one 1.

\[ \text{Efficiency}(T, T_n) = \frac{E(T_n - \beta)^2}{E(T - \beta)^2} \]

(11)

RESULTS AND DISCUSSION

Table 1: RMSE of Coefficient estimates from the various estimators under different outlying condition

| Outlier percentage | Sample Size | Estimator          | Robust -M | Robust MM | Robust-S | Robust-LTS | Robust-LAD |
|--------------------|-------------|--------------------|-----------|-----------|-----------|------------|------------|
|                    |             |                    |           |           |           |            |            |
| 0                  | 20          | 0.07064922         | 0.07270832| 0.2603092| 0.28988819| 0.09925639|
|                    | 40          | 0.03585571         | 0.03113030| 0.09679149| 0.14748303| 0.04427831|
|                    | 60          | 0.02084533         | 0.02109373| 0.07648414| 0.11972958| 0.03112109|
|                    | 100         | 0.01157941         | 0.01166778| 0.04406762| 0.07257024| 0.01790462|
|                    | 200         | 0.00561002         | 0.00563080| 0.02609702| 0.04567185| 0.008525684|

Table 2: Bias of Coefficient estimates from the various estimators under different outlying condition

| Outlier percentage | Sample Size | Estimator          | Robust -M | Robust MM | Robust-S | Robust-LTS | Robust-LAD |
|--------------------|-------------|--------------------|-----------|-----------|-----------|------------|------------|
|                    |             |                    |           |           |           |            |            |
| 0                  | 20          | 0.2093655          | 0.2126916| 0.3539176| 0.4124872| 0.2499738|
|                    | 40          | 0.1395238          | 0.1403076| 0.245689| 0.3006825| 0.1674017|
|                    | 60          | 0.116049|         | 0.1167313| 0.2193111| 0.2757462| 0.1413239|
|                    | 100         | 0.08580933         | 0.08605832| 0.16643276| 0.21315400| 0.10401642|
|                    | 200         | 0.05969014         | 0.05977360| 0.12845390| 0.17035070| 0.07242573|

|                    | 20          | 0.2092223          | 0.1808144| 0.2851223| 0.3687487| 0.2375808|
|                    | 40          | 0.1351381          | 0.1154182| 0.2021793| 0.2687385| 0.1519090|

FUDMA Journal of Sciences (FJS) Vol. 6 No. 1, March, 2022, pp 368 - 376
### Table 3: Variance of Coefficient estimates from the various estimators under different outlying condition

| Outlier percentage | Sample Size | Robust -M | Robust MM | Robust-S | Robust-LTS | Robust-LAD |
|--------------------|-------------|-----------|-----------|----------|------------|------------|
| 0                  | 20          | 0.02578261| 0.02704752| 0.07948785| 0.11809840 | 0.03613602 |
|                    | 40          | 0.01106449| 0.01136920| 0.02637458| 0.05629089 | 0.01619366 |
|                    | 60          | 0.00734661| 0.00734261| 0.002827225| 0.043510126| 0.011103152|
|                    | 100         | 0.00421405| 0.00259672| 0.016327368| 0.027113597| 0.006286764|
|                    | 200         | 0.02003417| 0.00205388| 0.008539546| 0.016641168| 0.003007219|
| 5                  | 20          | 0.22092921| 0.02658536| 0.06723951| 0.10814025 | 0.25342874 |
|                    | 40          | 0.05255633| 0.01095720| 0.03374572| 0.05720576 | 0.05165415 |
|                    | 60          | 0.009607724| 0.007524900| 0.00421786| 0.04308893 | 0.012414408|
|                    | 100         | 0.00469025| 0.003698088| 0.014454423| 0.026015922| 0.005904862|
|                    | 200         | 0.001819106| 0.001684848| 0.008252968| 0.039808275| 0.002509478|
| 10                 | 20          | 0.08416814| 0.02069145| 0.05571222| 0.09385618 | 0.08946993 |
|                    | 40          | 0.02868267| 0.01040385| 0.027887874| 0.05017417 | 0.03269002 |
|                    | 60          | 0.032749e-03| 0.06705552| 0.0062520-03| 0.3610506-02| 1.04516482-02|
|                    | 100         | 0.00526441| 0.003694931| 0.012827771| 0.022814382| 0.006405154|
|                    | 200         | 0.00170412| 0.001667176| 0.007762443| 0.011431618| 0.002766718|
| 30                 | 20          | 0.0511762e+01| 2.716386e-02| 4.484080e-02| 7.316037e-02| 9.310993e-01|
|                    | 40          | 0.03918098| 0.01014939| 0.02340542| 0.04092913 | 0.02287583 |
|                    | 60          | 1.536502e-02| 6.948244e-03| 1.760075e-02| 2.9228502e-02| 1.391265e-02|
|                    | 100         | 8.269337e-03| 4.058920e-02| 1.245086e-02| 1.976530e-02| 8.120719e-03|
|                    | 200         | 0.003718949| 0.001890064| 0.007604865| 0.010627911| 0.003633944|
|                    | 20          | 3.439819e+02| 3.238792e-02| 4.235182e-02| 6.808967e-02| 3.443174e+01|
|                    | 40          | 12.85508379| 0.01244392| 0.02352072| 0.03278415 | 1.21380707 |
|                    | 60          | 3.275245e-01| 8.177677e-03| 1.633525e-02| 2.618616e-02| 2.014467e-02|
|                    | 100         | 2.65809e-02| 4.633651e-03| 1.284705e-02| 1.67255e-02 | 1.072177e-02|
|                    | 200         | 7.411217e-03| 2.122629e-03| 8.598949e-03| 9.203125e-03| 4.917510e-03|
|                    | 20          | 1.199181e+03| 3.346673e-02| 1.853301e+01| 4.092227e+02| 3.961937e+02|
|                    | 40          | 193.17325883| 0.01890083| 0.02642862| 0.02562976 | 2.81680165 |
|                    | 60          | 60.034814515| 0.01179076| 0.02092649| 0.01969546 | 0.04738662 |
|                    | 100         | 5.88690e+02| 5.893713e-03| 1.506998e-02| 1.303252e-02| 1.656233e-02|
|                    | 200         | 2.509650e-02| 2.662694e-03| 1.033502e-02| 7.874868e-03| 6.773354e-03|
Table 4: Relative efficiency of Coefficient estimates from the various estimators under different outlying condition

| Outlier percentage | Sample Size | Estimator | Robust -M | Robust MM | Robust-S | Robust-LTS | Robust-LAD |
|-------------------|-------------|-----------|-----------|-----------|----------|------------|------------|
| 0                 | 20          | 0.9502803 | 0.9142981 | 0.3234201 | 0.2313309 | 0.669599   |
|                   | 40          | 0.9499412 | 0.9340235 | 0.3005521 | 0.1976194 | 0.656619   |
|                   | 60          | 0.9310824 | 0.9202790 | 0.2543356 | 0.1629528 | 0.623610   |
|                   | 100         | 0.9468383 | 0.9369493 | 0.2495488 | 0.1512752 | 0.642881   |
|                   | 200         | 0.9467297 | 0.9432245 | 0.2035174 | 0.1162836 | 0.642603   |
| 5                 | 20          | 4446.582  | 8857.413  | 6591.848  | 4437.430   | 8969.458   |
|                   | 40          | 7359.158  | 9558.256  | 5926.395  | 3781.082   | 9119.084   |
|                   | 60          | 7048.575  | 8156.033  | 5337.219  | 3265.661   | 9986.442   |
|                   | 100         | 7025.968  | 8146.707  | 2470.017  | 1315.383   | 5182.173   |
|                   | 200         | 8910.908  | 8543.236  | 1652.990  | 1133.492   | 6819.636   |
| 10                | 20          | 8532.333  | 14553.032 | 5502.181  | 2629.700   | 8969.458   |
|                   | 40          | 11051.584 | 16458.735 | 5926.395  | 3781.082   | 9119.084   |
|                   | 60          | 11941.264 | 16964.892 | 5337.219  | 3265.661   | 9986.442   |
|                   | 100         | 10638.954 | 19357.338 | 2470.017  | 1315.383   | 5182.173   |
|                   | 200         | 14610.677 | 20372.211 | 1652.990  | 1133.492   | 6819.636   |

Table 5: Rank of Performances of the estimators using RMSE criterion

| Sample size | Estimator | 0% | 5% | 10% | 20% | 30% | 40% |
|-------------|-----------|----|----|-----|-----|-----|-----|
| n=20        | Robust-M  | 1  | 4  | 3   | 5   | 5   | 5   |
|             | Robust-MM | 2  | 1  | 1   | 1   | 1   | 4   |
|             | Robust-S  | 4  | 2  | 2   | 2   | 2   | 2   |
|             | Robust-LTS| 5  | 3  | 5   | 3   | 3   | 1   |
|             | Robust-LAD| 3  | 5  | 4   | 4   | 4   | 3   |

| n=40        | Robust-M  | 1  | 2  | 2   | 4   | 5   | 5   |
|             | Robust-MM | 2  | 1  | 1   | 1   | 1   | 1   |
|             | Robust-S  | 4  | 4  | 4   | 3   | 2   | 3   |
|             | Robust-LTS| 5  | 5  | 5   | 5   | 3   | 2   |
|             | Robust-LAD| 3  | 3  | 3   | 2   | 4   | 4   |

| n=60        | Robust-M  | 1  | 2  | 2   | 3   | 5   | 5   |
|             | Robust-MM | 2  | 1  | 1   | 1   | 1   | 1   |
|             | Robust-S  | 4  | 4  | 4   | 4   | 2   | 3   |
|             | Robust-LTS| 5  | 5  | 5   | 5   | 4   | 2   |
|             | Robust-LAD| 3  | 3  | 3   | 2   | 3   | 4   |
Table 6: Rank of performances of the estimators using absolute bias criterion

| Sample size | Estimators | Outlier 0% | Outlier 5% | Outlier 10% | Outlier 20% | Outlier 30% | Outlier 40% |
|-------------|------------|------------|------------|------------|------------|------------|------------|
| n=20        | Robust-M   | 1          | 2          | 2          | 5          | 5          | 5          |
|             | Robust-MM  | 2          | 1          | 1          | 1          | 1          | 1          |
|             | Robust-S   | 4          | 4          | 4          | 2          | 2          | 2          |
|             | Robust-LTS| 5          | 5          | 5          | 3          | 3          | 1          |
|             | Robust-LAD | 3          | 3          | 3          | 4          | 4          | 3          |
| n=40        | Robust-M   | 1          | 2          | 2          | 3          | 5          | 5          |
|             | Robust-MM  | 2          | 4          | 1          | 1          | 1          | 1          |
|             | Robust-S   | 4          | 5          | 4          | 3          | 2          | 3          |
|             | Robust-LTS| 5          | 2          | 5          | 5          | 3          | 2          |
|             | Robust-LAD | 3          | 3          | 3          | 2          | 4          | 4          |
| n=60        | Robust-M   | 1          | 2          | 2          | 3          | 5          | 5          |
|             | Robust-MM  | 2          | 1          | 1          | 1          | 1          | 1          |
|             | Robust-S   | 4          | 4          | 4          | 4          | 2          | 3          |
|             | Robust-LTS| 5          | 5          | 5          | 5          | 4          | 2          |
|             | Robust-LAD | 3          | 3          | 3          | 2          | 2          | 4          |
| n=100       | Robust-M   | 1          | 2          | 2          | 3          | 5          | 5          |
|             | Robust-MM  | 2          | 1          | 1          | 1          | 1          | 1          |
|             | Robust-S   | 4          | 4          | 4          | 4          | 3          | 4          |
|             | Robust-LTS| 5          | 5          | 5          | 5          | 4          | 2          |
|             | Robust-LAD | 3          | 3          | 3          | 2          | 2          | 4          |

Table 7: Rank of performance of the estimators using Variance criterion

| Sample size | Estimator | Outlier 0% | Outlier 5% | Outlier 10% | Outlier 20% | Outlier 30% | Outlier 40% |
|-------------|-----------|------------|------------|------------|------------|------------|------------|
| n=20        | Robust-M  | 1          | 4          | 3          | 5          | 5          | 5          |
|             | Robust-MM | 2          | 1          | 1          | 1          | 1          | 3          |
|             | Robust-S  | 4          | 2          | 2          | 2          | 2          | 2          |
|             | Robust-LTS| 5          | 3          | 5          | 3          | 3          | 1          |
|             | Robust-LAD| 3          | 5          | 4          | 4          | 4          | 4          |
| n=40        | Robust-M  | 1          | 4          | 2          | 4          | 5          | 5          |
|             | Robust-MM | 2          | 1          | 1          | 1          | 1          | 1          |
|             | Robust-S  | 4          | 2          | 4          | 3          | 2          | 3          |
|             | Robust-LTS| 5          | 5          | 5          | 5          | 3          | 2          |
|             | Robust-LAD| 3          | 3          | 3          | 2          | 4          | 4          |
| n=60        | Robust-M  | 1          | 2          | 2          | 3          | 5          | 5          |
|             | Robust-MM | 2          | 1          | 1          | 1          | 1          | 1          |
### Table 8: Rank of performances of the estimators using Relative Efficiency criterion

| Sample size | Estimator | Outlier 0% | Outlier 5% | Outlier 10% | Outlier 20% | Outlier 30% | Outlier 40% |
|-------------|-----------|------------|------------|------------|------------|------------|------------|
| n=20        | Robust-M  | 5          | 4          | 4          | 1          | 1          | 1          |
|             | Robust-MM | 4          | 5          | 5          | 5          | 5          | 5          |
|             | Robust-S  | 2          | 3          | 2          | 4          | 4          | 2          |
|             | Robust-LTS| 1          | 1          | 1          | 3          | 3          | 4          |
|             | Robust-LAD| 3          | 2          | 3          | 2          | 2          | 3          |
| n=40        | Robust-M  | 5          | 4          | 4          | 2          | 1          | 1          |
|             | Robust-MM | 4          | 5          | 5          | 5          | 5          | 5          |
|             | Robust-S  | 2          | 2          | 2          | 4          | 4          | 3          |
|             | Robust-LTS| 1          | 1          | 1          | 3          | 3          | 4          |
|             | Robust-LAD| 3          | 3          | 3          | 3          | 2          | 2          |
| n=60        | Robust-M  | 5          | 4          | 4          | 2          | 1          | 1          |
|             | Robust-MM | 4          | 5          | 5          | 5          | 5          | 5          |
|             | Robust-S  | 2          | 2          | 2          | 3          | 4          | 2          |
|             | Robust-LTS| 1          | 1          | 1          | 3          | 3          | 4          |
|             | Robust-LAD| 3          | 3          | 3          | 3          | 2          | 2          |
| n=100       | Robust-M  | 5          | 4          | 4          | 3          | 3          | 5          |
|             | Robust-MM | 4          | 5          | 5          | 5          | 5          | 4          |
|             | Robust-S  | 2          | 2          | 2          | 2          | 2          | 2          |
|             | Robust-LTS| 1          | 1          | 1          | 1          | 1          | 1          |
|             | Robust-LAD| 3          | 3          | 3          | 4          | 3          | 1          |
| n=200       | Robust-M  | 5          | 5          | 4          | 3          | 3          | 1          |
|             | Robust-MM | 4          | 4          | 5          | 5          | 5          | 5          |
|             | Robust-S  | 2          | 2          | 2          | 3          | 2          | 2          |
|             | Robust-LTS| 1          | 1          | 1          | 1          | 1          | 1          |
|             | Robust-LAD| 3          | 3          | 3          | 3          | 2          | 2          |

### Table 9: Parameter Estimates of various Estimators under outlying condition using real life data

| Estimator | Robust-M | Robust-MM | Robust-S | Robust-LTS | Robust-LAD |
|-----------|----------|-----------|----------|------------|------------|
| $\beta_0$ | 173.3444913 | 133.7990838 | 63.3640509 | 78.3057784 | 141.1575281 |
| $\beta_1$ | 0.9975838 | 1.3573958 | 1.7498547 | 1.5652388 | 1.2717276 |
| $\beta_2$ | 1.1152763 | 0.9886164 | 0.8531579 | 0.8983310 | 1.0061706 |
| $\beta_3$ | -1.1158903 | -1.0523911 | -1.1621459 | -0.9232399 | -0.9359867 |
Comparative Study of Some Estimators

Table 10: Variance and Mean Square Error of Prediction (MSEP) of Coefficient estimates from the various estimators under outlying condition using real life data

| Estimator | Robust –M | Robust –MM | Robust –S | Robust –LTS | Robust –LAD |
|-----------|-----------|------------|-----------|-------------|-------------|
| Variance  | 7484.3535 | 4447.8708  | 990.0747  | 987.0263    | 4950.8103   |
| MSEP      | 21740.22  | 30862.24   | 35717.46  | 35415.3     | 29889.85    |

Assessing the Performance of Estimators under Outlying Condition Using Real Life Data

Table 11: Performance of estimators of real life data using variance and MSEP criterion

| Criteria | Robust-M | Robust-MM | Robust-S | Robust-LTS | Robust-LAD |
|----------|----------|-----------|----------|------------|------------|
| Variance | 5        | 3         | 2        | 1          | 4          |
| MSEP     | 1        | 3         | 5        | 4          | 2          |

Assessing the Performance of Estimators under Outlying Condition Using Real Life Data

Going by variance criterion, the estimator that best fit the dataset as evident from Table 11 is Robust-LTS estimator. Similarly, from Table 11 above it can be observed that Robust-M estimator is the best estimator in terms of predictions.

Based on the results about the estimators presented in Table 1 through Table 8, the paper found that under 0% outlying condition and using absolute bias, variance and root mean square error criterions, Robust-M is the most preferred estimator while Robust-LTS is the least Preferred at all sample sizes (20, 40, 60, 100, and 200). However, under 0% outlying condition and using Relative Efficiency criterion, Robust-LTS is the most preferred Estimator while Robust-M is the least preferred estimator at all sample sizes (20, 40, 60, 100, and 200).

In addition, under 5% outlying condition and using absolute bias, variance and root mean square error criterion, Robust – MM is the most preferred estimator except when the sample size is 40 using absolute bias criteria while Robust-LTS is the least preferred Estimator except when the sample size is 40 using absolute bias, 20 and 40 using variance and 20 using root mean square error criterion respectively.

However, under 5% outlying condition and using relative efficiency, Robust-LTS is the most preferred estimator across all sample size while Robust-MM is the least preferred estimator across all sample size except when the sample size is 200.

Under 10% outlying condition and using absolute bias, variance and root mean square error criteria, Robust-MM is the most preferred estimator while Robust-LTS is the least preferred Estimator across all sample size. However, under 10% outlying condition and using relative efficiency, Robust-LTS is the most preferred estimator while Robust-MM is the least preferred estimator across all sample size except when the sample size is 200.

Under 20% outlying condition and using absolute bias, variance and root mean square error criteria, Robust –MM is the most preferred estimator across all sample size except when the sample size is 200 using absolute bias criterion while Robust-LTS is the least preferred Estimator across all sample size except when the sample size is 20.

Consequently, under 20% outlying condition and using relative efficiency, Robust-LTS is the most preferred estimator except when the sample size is 20 while Robust-MM is the least preferred estimator across all sample size. Using 30% outlying condition and using absolute bias, variance and root mean square error criteria, Robust –MM is the most preferred estimator across all sample size except when the sample size is 200 using absolute bias criterion while Robust-LTS is the least preferred Estimator across all sample size except when the sample size is 200 using root mean square error.

However, under 30% outlying condition and using relative efficiency, Robust-M and Robust-LTS are the most preferred estimators when the sample sizes are 20, 40 and 60 and 100 and 200 respectively while Robust-MM is the least preferred estimator across all sample size. Going 40% outlying condition and using absolute bias, variance and Root Mean Square Error criteria, Robust –MM is the most preferred estimator across all sample size except when the sample size is 20 using Root Mean Square Error criterion while Robust-M is the least preferred Estimator across all sample size.

Also, under 40% outlying condition and using relative efficiency, Robust-M is the most preferred estimators except when the sample size is 100 while Robust-LTS and Robust-MM is the least preferred estimator at sample sizes 20 and 40, 60 and 200 respectively.

That sample size has little or no effect on the performance of the estimators across all the different outlier levels.

Conclusion

In conclusion, the study concludes that Robust-M is the most efficient estimator across all the comparison criteria in the absence of outlier except when using relative efficiency and

FUDMA Journal of Sciences (FJS) Vol. 6 No. 1, March, 2022, pp 368 - 376
that Robust-MM is the most consistent estimator across the comparison criteria except when using relative efficiency. Also, sample size has little or no effect on the performance of the estimators across all the different outlier levels.

REFERENCES
Alma, O. G. (2011). Comparison of robust regression methods in linear regression. *International Journal for Contemp Maths and Science*, 6: 409-421.

DasGupta, M. & Mishra, S.K. (2004). *Least absolute deviation estimation of linear econometric models: A literature review*. Available:http://mpra.ub.uni-muenchen.de/781.

David D. (2014). *Comparison of Robust Regression Estimators*. M.phil thesis, Kwame Nkrumah University of Science and technology, Ghana, Ghana.

Hampel, F. (2001). Robust statistics: A brief introduction and overview. Pages 1-5. In David D. (2014). Comparison of Robust Regression Estimators

Huber, P. J. (1972). The 1972 wald lecture: Robust statistics. *The Annals of Mathematical Statistics*, 43: 1041-1067. In David D. (2014). *Comparison of Robust Regression Estimator*.

Hawkins, D. (1980). Identification of Outliers. Chapman and Hall. London. In Edgar, A. & Caroline, R. (2004). On detection of outliers and their effect in supervised classification. *Conference paper*.

Liu H., Shah S. and Jiang W. (2004). On-line outlier detection and data cleaning. Computers and Chemical Engineering. In Maimon O. and Rockach L. (Eds.) (2005). *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*.

Rousseeuw, P. J. (1984). *Robust Regression and Outlier Detection*. JOHN WILEY and SONS.

Rousseeuw, P. J. & Yohai, V. J. (1984). Robust Regression by Mean of S Estimators. *Robust and Nonlinear Time Series Analysis*, 256-274, doi: 10.1007/978-1-4615-7821-5-15.

Stephen, R. & Senthamarai, K. K. (2017). Detection of Outliers in Regression Model for Medical Data. *International Journal of Medical Research &Health Sciences*, 6(7), 50-56.

Williams, G. J., Baxter, R. A., He H. X. & Hawkins S., Gu L. (2002). A Comparative Study of RNN for Outlier Detection in Data Mining. *IEEE International Conference on Data-mining*. In: Maimon O. and Rockach L. (Eds.) (2005). *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*.

Yohai, V. J. (1987). High Breakdown Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, 629-649, doi:10.1214/aos/1176350366.

Zimek, A. & Filzsmorer, P. (2018). “There and Back again: Outlier detection between statistical reasoning and data mining algorithm”. *Wiley Interdisciplinary Reviews: Data mining and Knowledge Discovery*, 8(6). Doi:10.1002/widm.1280

©2022 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via https://creativecommons.org/licenses/by/4.0/ which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.