* **MONTE-CARLO STUDY OF SOME ROBUST ESTIMATORS**

**THE SIMPLE LINEAR REGRESSION CASE**

**ADEWOLE, AYOADE I.1, BODUNWA, OLUWATOYIN K.2 and SADIKU, OPEYEMI B.3**

*1Department of Mathematics, Tai Solarin University of Education Ijagun Ogun State Nigeria.*

*2Department of Statistics, Federal University of Technology Akure.Ondo State Nigeria.*

Corresponding Author Email Address:

hayorhade2005@gmail.com, P.M.B 2118, Ijebu-Ode, Ogun State, Nigeria, Tel: +2348055124368

Author Email Address:

[okbodunwa@futa.edu.ng](mailto:okbodunwa@futa.edu.ng) P.M.B 704, Akure, Ondo State, Nigeria, Tel: +2347031351004

Author Email Address:

[opesadiku@gmail.com](mailto:opesadiku@gmail.com) P.M.B 704, Akure, Ondo State, Nigeria, Tel: +2349026066696

**ABSTRACT**

*In this study, Least Trimmed Squares (LTS), Theil’s Pair-wise Median (Theil) and Bayesian estimation methods (BAYES) are compared relative to the OLSE via Monte-Carlo Simulation. Variance, Bias, Mean Square Error (MSE) and Relative Mean Square Error (RMSE) are computed and used to evaluate and rank the estimators’ performance. The Simple Linear Regression model is investigated for the conditions in which the error term is assumed to be drawn from three error distributions: unit normal, lognormal and Cauchy. Theil’s non-parametric estimation procedure was found to have the strongest and most reliable performance. The second-best results are obtained from LTS method. Though it was observed that the Bayesian estimators are affected by deviation of the dataset from normality, yet it is established from the results that the Bayesian estimators performed optimally more than all other competitors, even under non normal situations (especially under the standard lognormal distribution) in some cases, except whenever the error is drawn from a heavy tail distribution (Lognormal and Cauchy). OLSE is only reliable as long as the normality assumptions hold.*

Keywords: Robust estimation, Monte-Carlo, Lognormal and Cauchy

I**NTRODUCTION**

The frequentist perspective of linear regression is probably the most familiar. It gives a single estimate for the model parameters based only on the assumption that the model is completely informed (or influenced) by the data set alone. However, the classical (frequentist) estimations, especially the OLSE have been noticed to breakdown under small sample condition. It is well demonstrated that outliers, even just one, in the sample data heavily influence estimates using OLSE regression. Thus, alternative methods, which have considerable advantages over OLSE procedure and are practical to use, have been developed(Birkes and Dodge, 1993; Dietz, 1987; Iman and Conover, 1979; Bolstad, 2010).

The alternatives to OLSE method considered in this research paper include the Least Trimmed Squares (LTS), which was introduced by Rousseeuw in 1984, Theil’s pairwise median procedure introduced in 1950., which is expected to perform well without regards to the distribution of the error terms and the Bayesian inference method claimed to have been developed by Thomas Bayes in 1763, and further developed by Pierre-Simon Laplace (Meenai and Yasmeen, 2008).

The Bayesian perspective requires the formulation of linear regression using probability distributions rather than point estimates. It advocates the update of prior beliefs in the evidence of new data. The aim is to determine the posterior distribution, and not to find the single “best” value, for the model parameters. The concept of Bayesian regression is a simple application of Bayes Theorem, the fundamental bedrock of Bayesian Inference. Since the 1980s, the Bayesian methods have seen increasing use within statistics and have found application in many fields due to the discovery of Markov chain Monte Carlo methods (Bolstad, 2004).

Hussain and Sprent (1983) presented a simulation study in which they compared the OLSE regression against the Theil pairwise median and weighted Theil estimators in a study using one hundred (100) replications per condition. Agullo (2001) proposed two algorithms to compute the LTS estimators. Rousseeuw and Leroy (1987) introduced a Monte Carlo simulation study to compare regression estimators including OLSE, Least Trimmed Squares (LTS) and Theil, alongside other frequentist regression estimators, for the Simple Linear Regression model when the distribution of the error terms is Generalized Logistic. Meenai and Yasmeen (2004) applied non-parametric regression methods to some real and simulated data.

Sunita (1999) explained the drawbacks of the multiple regression approach for software engineering data and discusses the Bayesian approach which alleviates a few of the problems faced by the multiple regression approach. A Variational Bayesian Gaussian Mixture Regression (VBGMR) method for soft sensing key quality-related variables in a non-Gaussian industrial process was developed by Jinlin et al. (2017).

The present study explores the behavior of robust regression and Bayesian inference approaches to Simple Linear Regression under three different situations with respect to contaminated data and non-normal error distributions. Very little research, if any, exists in which the Bayesian method to linear regression is directly compared to Robust-non-parametric alternatives.

**METHODOLOGY**

**Simulation Design**

The study design uses a Monte-Carlo simulation technique developed with VISUAL FORTRAN. The design variable is generated as a sequential model of the form while the response variable Y is generated as a linear model . 150,000 sets of random data of sample sizes were simulated, and were crossed with three types of error distributions (Normal, Lognormal and Cauchy) from which the random component εt was assumed to be drawn.

**Choice of Error Distributions**

The Lognormal and Cauchy distributions are heavy tailed distributions. However, the Cauchy distribution has much heavier tails than the Log-normal distribution. Hence, the two distributions are used to observe the performance of each estimator as the dataset deviates from normality. Sensitivity of each estimation method to outliers were put to test by constructing alternative forms of the error distributions (mixture, outlier and contamination).

Detailed information on this approach, and algorithm for drawing random deviates from each of the error distributions mentioned above, are given in Evans, Hastings, and Peacock (1993).

**Estimation Procedures**

For each simulated data set, the estimators of α and β are calculated using the four estimation techniques earlier described. The y-intercept α and slope β estimators are computed and for each estimator mean, variance, bias and mean square error (MSE) are calculated where . Details on the algorithms for calculating these estimators have been discussed in previous literatures (Evans, Hastings and Peacock, 1993; Hussain and Sprent, 1983). However, each procedure is briefly reviewed below:

**Ordinary Least Squares Method**

The Ordinary Least Squares Estimation (OLSE) method was introduced by Gauss in 1794. Ordinary least Squares (OLS) is a standard approach to specify a linear regression model and estimate its unknown parameters by minimizing the sum of squared errors. This leads to an approximation of the mean function of the conditional distribution of the dependent variables. OLS achieves the property of Best, Linear and Unbiased Estimator (BLUE), if

 1.4

Where;

Y is the response variable,

 is the product of the transpose of matrix and the matrix of slope parameter,

 is the error term.

The following assumptions hold:

i. The relationship between Y and X requires that the dependent variable (Y) is a linear combination of explanatory variable and error term

1. The explanatory variable is non Stochastic.
2. The expectation of the error term is zero i.e .
3. Homoscedacity – the variance of the error terms ei is constant i.e.
4. No autocorrelation i.e. ij.
5. Normality of errors i.e the error are normally distributed with mean 0 and variance .

However, frequently one or more of these assumptions are violated; resulting in that OLS is not any more the best linear unbiased estimator. However, these assumptions are stringent such that if any one of the assumptions is not met, OLSE procedure breaks down. The two, of the six assumptions which most often cause issues are the assumption of homoscedasticity of the residuals and normally distributed residuals. *When these two assumptions fail, OLSE estimates will still be unbiased and consistent; however, the estimates will be inefficient (OLSE will give incorrect estimates of the parameter standard errors)****.***

**The Least Trimmed Squares (LTS)**

The Least Trimmed Squares (LTS), which was introduced by Rousseeuw in 1984(Sunita Chulani, 1999), aims at minimizing by choosing a subsample of *h* observations, computing some α and β that minimize the sum of squared residuals for the selected subsample and then deleting data points corresponding to a specific percentage of the largest residuals under an initial OLSE to reduce their adverse effects on the estimators (Rousseeuw and Leroy, 1987). The estimate, out of estimates for both α and β, which makes the objective function smallest is the final estimate. Thus, *the only difference between OLSE and LTS estimation is that in LTS the largest squared residuals are not used* and therefore the fit is not so much affected by the outliers(Michael and Christopher, 1999; Rousseeuw and Leroy, 1987).

**Theil’s Pair-Wise Median Methods**

The Theil’s method, as well as many other non-parametric procedures, is based on using the ranks of the observed data rather than the observed data themselves (Hussain and Sprent, 1983; Rousseeuw and Leroy, 1987; Bolstad, 2010). The complete Theil’s slope estimate is computed by comparing each data pair to all others in a pairwise fashion. A data set of pairs will result in pairwise comparisons. For each of these comparisons a slope is computed. The median of all possible pairwise slopes is taken as the non-parametric Theil’s slope estimate, , where; . The y-interceptis obtained by calculating; and taking the median of these values as the y-intercept. Conover suggested estimating by using the formula:, which is analogous to OLSE, where the fitted line always goes through .

**Bayesian Inference Method**

The Simple Linear Regression model has three fundamental assumptions (Bolstad, 2004; Wolpert et al., 2004). The application of Bayesian estimation requires a probabilistic reformulation of the Simple Linear Regression model based on these assumptions. To achieve this, the response variable Y, and the model parameters α, β and ε are assumed to come from a predetermined (prior) distribution. Additionally, an appropriate likelihood function is specified. The likelihood component is the part that incorporates the data. The Bayes' rule is the working instrument with which the update of the prior distribution to the posterior distribution is achieved in the light of new data. It is the tool that gets us from the probability of the data given the model to the likelihood of the model given the data (Laplace, 1800).

The posterior probability of the model parameters is conditional upon the training inputs and outputs;

. (1)

Here; is the posterior probability distribution of the model parameters given the inputs and outputs. This is equal to the likelihood of the data, , multiplied by the prior probability of the parameters and divided by a normalization constant.

The joint likelihood of the , denoted , observation is its probability density function as a function of the two parameters , where are fixed at the observed values; it gives relative weights to all possible values of both parameters from the observations:

(2)

The joint likelihood of the whole sample of all observations is the product of the independent likelihoods:

. (3)

This simplifies to:

, (4)

Where; (5)

; (6) (7)

Thus the joint likelihood of sample is;

α (8)

Factorizing and completing the squares;

(9)

where; (the least squares slope); and (the least squares intercept)

The joint prior of and is the product of the individual prior;

. (10)

By Bayes’ rule, the joint likelihood multiplied by the joint prior is proportional to the joint posterior:

(11)

Where the data is the set of ordered pair . The joint posterior factors to

(12)

The marginal posteriors are independent and can be found by the simple updating rules for normal distributions:

Given prior for β, then the posterior is; ; where

(13)

And

(14)

Also, Given prior for , then the posterior is; ; where

(15)

And

(16)

For the specific cases of this research, the probabilistic reformulation of the regression model parameters assumes a conjugate Normal prior distribution with mean 0on both the regression slope parameter βand the regression intercept parameter α. The prior standard deviation was computed as where σ is the OLSE sample estimate, while adopting the Normal probability distribution density) as the likelihood of the data across all the cells of the simulation.

**SIMULATION RESULTS AND DISCUSSIONS**

Tables 1, 2, 3 and 4 reveals the results across sample sizes, estimator variances for both the intercept and the slope parameters decreased with increasing sample size. This pattern of decreasing variance and bias holds for all estimators under all error distributions. The pattern seen in the variances are also exhibited in the estimators MSE values. Because the results for the n=30 sample size are intermediate to those of n=10 and n=50 sample sizes, they are not reported here.

A close look at that all the selected slope estimators are approximately unbiased i.e. they have negligible bias. This pattern of performance is observed to improve consistently as sample size increases and across all error distributions and respective alternative models. Results of the work reveals that OLSE and Bayesian slope estimators are always biased whenever the error term is Cauchy distributed regardless of sample size (though the performance improve with increasing sample size).

**TABLE 1: GENERAL ESTIMATORS’ PERFORMANCE RESULTS CHART FOR SOME REGRESSION ESTIMATORS (SAMPLE SIZE n=10)**



**TABLE 2: GENERAL ESTIMATORS’ PERFORMANCE RESULTS CHART FOR SOME REGRESSION ESTIMATORS (SAMPLE SIZE n=50)**



**TABLE 3: GENERAL ESTIMATORS’ PERFORMANCE RESULTS CHART FOR SOME REGRESSION ESTIMATORS (SAMPLE SIZE n=100)**



**TABLE 4: GENERAL ESTIMATORS’ PERFORMANCE RESULTS CHART**

**FOR SOME REGRESSION ESTIMATORS (SAMPLE SIZE n=300)**



Also, it is noticed that the Theil’s estimator gave the most satisfactory performance across all the cells of the simulation followed closely by LTS estimator (except whenever the sample size is very small). The Bayesian estimator maintained consistent relative unbiased-ness only under the standard and mixture normal error model. This confirms the elusiveness of the OLSE when the data set is heteroscedastic.

With the introduction of contaminations into the dataset, it is observed that all estimators remained relatively unbiased. However, OLSE and Bayesian estimators were noticed to maintain consistent unbiased-ness only when the sample size is either very small or very large (for outliers’ model) and consistent biased-ness (for contamination model) sample size notwithstanding. Under this condition, LTS estimator stayed on top most of the time (most especially whenever the error term is Cauchy distributed) with the least bias value, followed closely by Theil’s estimator and then the Bayesian point estimator.

**Based on variance and RMSE criteria**

The Bayesian point estimator of the slope parameter β, followed by the OLSE estimator, had the least variance and hence the least MSE under the normal distribution across all cells of the simulation and all sample sizes, with the exception of under mixture and contamination error model for large sample size (n=300) where Theil’s and LTS estimators led the way. As sample size increases, Bayesian and OLSE gained precision i.e. the variance and the MSE values of both estimators decrease with increasing sample size. Ultimately the Bayesian point estimator converges to the OLSE as the sample size tends to infinity. However, the Bayesian point estimator stays consistently more efficient than the OLSE sample size notwithstanding.

As the dataset deviates from normality, however, Bayesian estimator and OLSE begin to loose precision to LTS and Theil’s estimators with almost 400% increase in MSE across all sample sizes under the lognormal distribution and for Cauchy distribution, OLSE and Bayesian point estimator gave outrageously large value for MSE compare to other estimators. This pattern confirms that deviations from normality cause the OLSE to be poor estimator and thus its inappropriateness under non-normal conditions.

As long as the error distribution is normal, Theil’s and LTS estimators gave negative RMSE values without regards for the sample size except under outlier error model. But as the sample size increases and the error term deviates further from normality, Theil’s gains precision and its MSE values were approximately equal with that of OLSE when the sample size is very large (n≥50). The Least Trimmed Squares estimation method came next in view to Theil’s under the normal distribution for sample size n=10 and as has been usually observed, it experiences a consistent decrease in its MSE as sample size gets larger. For the non-normal distribution case, LTS compete rigorously with Theil’s. It was observed that the father away the error distribution deviated from normality, the better LTS becomes as it gains precision (55.35%, n=10; 42.97%, n=100 under lognormal distribution and 100% when n=10, 99.96% when n=100 under Cauchy distribution) following closely after Theil’s while displacing Bayesian estimator and OLSE out rightly.

**Effect of contamination**

For the first case of contamination considered in this paper (outliers case), the Theil’s slope estimator outperformed every other estimator regardless of sample size or distribution. OLSE slope estimator was better than LTS when sample size is small (n<50) and the error distribution is normal. Bayesian point estimator stayed afloat under the normal distribution but gave way whenever the error term is lognormal or Cauchy distributed. Unlike the OLSE and Bayesian estimators, Theil’s and LTS estimators generally improved in their precision compared to the standard model case.

It is clear from results that as long as the error distribution is normal, Bayesian and OLSE estimator win the game. But as the sample size gets larger (n≥100), Theils estimator defeats Bayesian estimator, and hence OLSE, by a slim chance of 4.37% reduction in its MSE relative to OLSE. However, when the error term is non-normal, Theils always take the crown followed closely by LTS.

The general pattern of estimators’ behavior, for the contamination case is similar to that of the outliers’ error model case, only that for small sample size (n≤30) and whenever the error term is normally distributed, Bayesian and OLSE estimators are the most efficient of all. But whenever n≥50, Theils estimator takes the stage irrespective of the distribution type and under non-normal error distribution situation, Theils and LTS kept their status-quo.

**Y-intercept estimators’ performance**

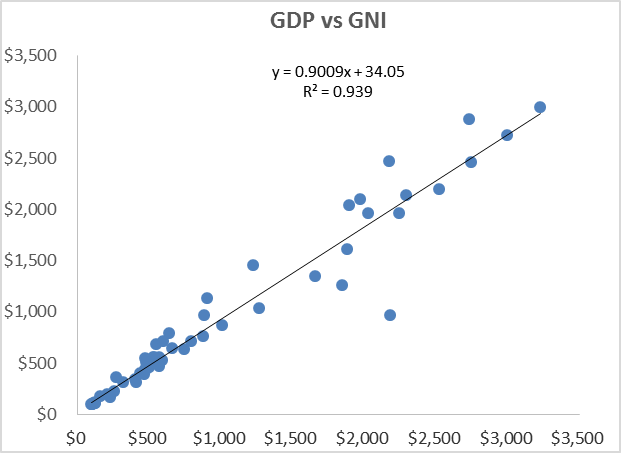
The performance of the regression y-intercept estimators for each of the methods was found to follow the same pattern as those of the slope estimators, but for a few significant variations. The variance and MSE for all the estimators are significantly larger than those of their respective slope estimators, especially as sample size increases.

It is worthy of note, also, that the Bayesian and OLSE intercept estimators, as long as the normality assumptions hold, are by far much more efficient than LTS and Theils intercept estimators regardless of sample size. But, under non-normal situations, and especially under the Cauchy error distribution, Theils and LTS take the stage as usual. Nevertheless, estimates of variance and MSE are larger and more consistent than the outliers’ case for all estimators across all cells of the simulation. The Bayesian estimators consistently maintains a higher precision about the true value of the regression parameters compared to all other estimators under consideration

For the contamination model, the patterns of estimators’ behavior for the intercept estimators are similar to the outliers’ case. However, estimates of variance and MSE are as usual larger than those of the respective slope estimators.

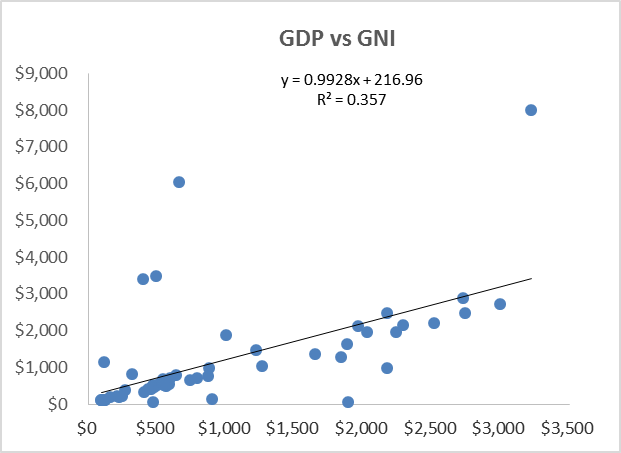
**TABLE 5: NIGERIA GROSS DOMESTIC PRODUCT AND GROSS NATIONAL INCOME (1969-2018)**

**EMPIRICAL DATA ANALYSIS**

 Table 5 presents a data set on Gross Domestic Products and Gross National Income of Nigeria for the year 1969-2018. Figure 1A shows the scatter plot of the dataset, the coefficient of determination value suggests a strong positive linear relationship between GDP and GNI. Figure 1B shows the scatter plot of the same dataset with 20% of the dataset contaminated, the drastic reduction in the coefficient of determination value suggests a sharp drop in the predictive power of the regression model due to contaminations in the dataset. Figure 2A and figure 2B show clearly that the required OLSE assumption of normally distributed residuals and homogeneous variances (homoscedasticity) are violated.

**FIGURE 1A: SCATTER PLOT OF GROSS DOMESTIC PRODUCT ON GROSS NATIONAL INCOME**

**FIGURE 1B: SCATTER PLOT OF GROSS DOMESTIC PRODUCT ON GROSS NATIONAL INCOME WITH 20% CONTAMINATION**



Regression analysis was carried out on the dataset for years (2009-2018), years (1989-2018) and years (1969-2018) respectively.

Table 6 shows the regression parameter estimates for the dataset using the previously discussed methods. The prior parameter for the intercept and slope parameters are obtained as the OLSE estimates for the intercept and slope parameter of the data ten years before the years in view. For the year 2009-2018, for example, the OLSE estimates for the year 1999-2008 were used as the prior means for the intercept and slope parameters. The variances for both parameters were calculated as: where σ is the OLSE sample estimate.

**TABLE 6: EMPIRICAL ANALYSIS OF GROSS DOMESTIC PRODUCT AND GROSS NATIONAL INCOME USING SELECTED REGRESSION TECHNIQUES**



**Intercept Parameter Estimators’ Performance**

For a very small sample size, n=10, the Bayesian estimates of the regression intercept parameter α has the least standard deviation (60.115). The LTS estimator put up a better performance with standard deviation 318.5068 than the OLSE estimator with standard deviation 345.0132. The standard deviation (1030.231) of the Theil’s estimator is particularly very large compared to OLSE and LTS.

However, as the sample size increases, OLSE and LTS improved significantly with almost 85.6% and 83.5% reduction in standard deviation respectively while Theil’s estimator improved with 53.9% reduction in standard deviation (for n=30). When the sample size increased to 50, Theil’s estimator further improved with 71.2% reduction in standard deviation, OLSE and LTS had no significant improvement, while the Bayesian estimator remained consistently more efficient than all others regardless of sample size.

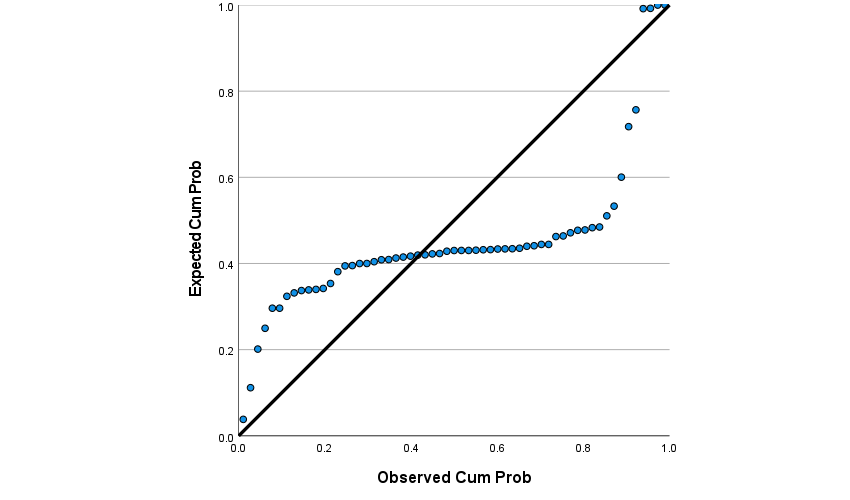
With 20% contamination in the data set, however, OLSE and LTS estimator lost significant precision with 468.0% and 526.8% increase in standard deviation respectively, while Theil’s estimators put up a better performance with 38.7% and the Bayesian estimator seems not to be significantly affected with just about 7.7% increase in the standard deviation. This affirms the robustness characteristic of the Bayesian estimator to contamination in the dataset.

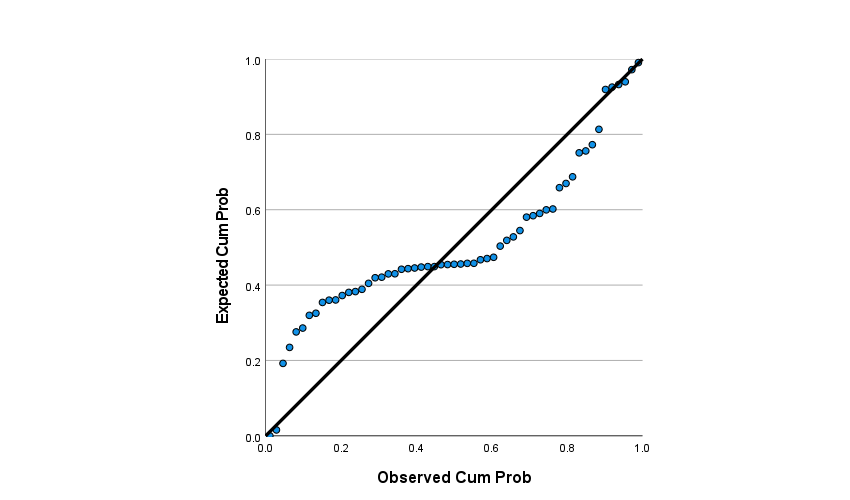
**Slope Parameter Estimators’ Performance**

For a very small sample size, n=10, it is noticed from Table 6 that the Bayesian estimate for the regression slope parameter β has the least standard deviation (0.13825) outperforming the LTS and Theil’s estimators by 12.4% and 206.1% respectively. The OLSE estimator for slope parameter seems to agree with the Bayesian estimator for all sample sizes. With significant increase in sample size (n=30), OLSE, LTS and Bayesian estimates improved significantly with 78.0%, 73.8% and 76.2% decrease in standard deviation respectively while the Theil’s estimator slightly improved with just about 29.7% reduction in standard deviation.

With further increase in sample size (n=50), the Theil’s estimator gained precision with a further 18.3% reduction in its standard deviation while the OLSE, LTS and Theil’s seem relatively consistent in their precision. With 20% contaminations in the dataset, however, the Theil’s estimator proved to be more robust than all others with just 37.6% reduction in its precision. This confirms the fact that OLSE estimator breaks down in the presence of outliers in the data set. Thus, the Theil’s slope estimator maintained an overall and consistent robustness over all other slope estimators across board.

**FIGURE 2A: NORMAL P-P PLOT OF STANDARDIZED RESIDUALS**





**STANDARDIZED RESIDUALS**

**FIGURE 2B: NORMAL P-P PLOT OF STANDARDIZED RESIDUALS FOR 20% CONTAMINATED DATA**

**Test of Hypothesis**

A two-sided hypothesis is set up for the regression slope parameter β as follows:

(17)

Versus

(18)

(Where ξ=0.8458, 0.2818, 0.4544 for n=10, 30, 50 respectively)

Looking at the range of the confidence interval, in Table 2 above, constructed for all estimators under consideration at α=0.05, the confidence interval for OLSE, LTS and Theil’s estimator spans the corresponding value of ξ for each sample size. Furthermore, the intervals showed that the value of β could possibly be zero and as well could be negative. Thus, the test is not significant at α=0.05 and we therefore cannot reject and conclude that there exists a significant predictive relationship between GDP and GNI at α=0.05 level of significance though the relationship is weakly quantified by the estimated values of the slope parameter respectively.

However, the Bayesian credible interval gave strong evidence, in the light of both the data and the prior knowledge that the values ξ=0.2818 and 0.4544 are not tenable values of , since the test is significant for both ξ=0.2818 and 0.4544, but ξ=0.8458 is a possible value of β since the test is not significant for ξ=0.8458 when the sample size is very small and very large. Also, the credible interval gave a very strong opinion that β can neither possibly be zero nor negative in value. Thus, is a credible range of β at and we could therefore conclude that there exists a significant predictive relationship between GDP and GNI at level of significance and the relationship is strongly quantified by the estimated values of the slope parameter respectively.

Hence, the fitted Simple Linear Regression model for each of the estimation procedure is as follows:

Ordinary Least Square Estimation:

Least Trimmed Square Estimation:

Theil’s Estimation:

Bayesian Estimation:

**CONCLUSION**

The usual assumption of linear regression is that error terms have a normal distribution, which leads OLSE procedure to give good results. However, in real life it is nearly impossible to find a data set that satisfies the normality assumption. Since under these situations, OLSE results in loss of efficiency, alternative regression techniques are needed. In this paper, some robust procedures for a simple linear regression model under normal and non-normal error situations were studied. Results from tables for slope estimators show that for all cases, the Theils method demonstrates the strongest performance gains as compared to OLSE. That is to say, they have negligible bias and the smallest MSE. The second-best results are obtained from the Least Trimmed Squares method. This study, therefore, shows that for simple linear regression model, Bayesian linear regression is a more optimal and safer alternative for the OLSE and provides inferences that are conditional on the data and are exact without reliance on asymptotic approximation. Theil estimator has high small sample efficiency across board, but especially when the variance of the error terms is not constant. More so, Theils estimation has the strongest performance and most reliable results and can be used in varying circumstances. Also, the empirical results affirm the robustness characteristic of the Bayesian estimator to contamination in the dataset. Least Trimmed Squares is most especially applicable when the error term is confirmed to come from a heavy tailed distribution and the sample size is large. OLSE is only reliable as long as the normality assumption holds.

The current study has also confirmed the applicability of the Theils method to varying circumstances and its robustness over many other robust methods, especially the LTS and Bayesian methods. However, the Bayesian model used in this study assumes a normal likelihood and a conjugate normal prior, thus restricting the Bayesian model to reflect normal data distribution situation only. Also, the Bayesian model should be constructed and studied for non-normal data distribution cases, and compared for robustness against other robust methods, especially the Theils method.

**REFERENCES**

1. Agullo, J. (2001). A new algorithm for computing the Least Trimmed Squared regressionestimator. *Computational Statistics and Data Analysis*
2. Bayes, Thomas (1763). An essay towards solving a problem in the doctrine of chances.*In Philosophical Transactions.* London.
3. Birkes, D., and Dodge, Y. (1993). Alternative Methods of Regression*.* New York, NY: Wiley.
4. Dietz, E.J. (1987). A comparison of robust estimators in simple linear regression.*Communication in Statistics-Simulation*.
5. Evans, M., Hastings, N., & Peacock, B. (1993). Statistical Distributions(2nd edition).New York, NY: Wiley.
6. Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin (1995, 2003, 2013). Bayesian Data Analysis. Teaching Statistics.
7. Hussain, S.S., and Sprent, P. (1983). Nonparametric Regression. *Journal of the Royal Statistical Society. Series A.*
8. Iman, R.L., and Conover, W.J. (1979). The use of rank transformation in regression. *Technometrics*
9. Jinlin, Z., Zhiqiang, G., and Zhihuan, S. (2017). Variational Bayesian Gaussian Mixture Regression for Soft Sensing Key Variables in Non-Gaussian Industrial Processes*. IEEE Trans. Contr. Sys. Techn*.
10. Laplace, P.S. (1800). Seances Des Ecoles Normales.
11. Meenai, Y.A., and Yasmeen, F. (2008). Nonparametric regression analysis. *Proceedings of the 8thIslamic Coumtries* *Conference of Statiscal Sciences,* Lahore - Pakistan
12. Michael, E.T., and Christopher, M. B. (1999). Probabilistic Principal Component Analysis. Journal of the *Royal Statistical Society, Series B.*
13. Muntan O.C. (2004). Comparison of regression techniques via Monte Carlo Simulation. A Thesis (Master of Science). Graduate School of Natural and Applied Sciences, Middle East Technical University.
14. Rousseeuw, P.J., and Leroy, A.M. (1987). Robust regression and outliers detection. New York, NY: Wiley.
15. Sunita Chulani (1999). Bayesian analysis of empirical software engineering cost models. *IEEE transactions on Software Engineering*.
16. Tam, H.P. (1996, April). A review of nonparametric regression techniques. A paper presented at the annual meetings of the *American Educational Research Association,* New York .
17. Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae*
18. William, M. Bolstad. (2010). Understanding computational Bayesian Statistics. *Wiley series in computational Statistics*. John Wiley & Sons, Inc., Hoboken, New Jersey
19. William, M. Bolstad. (2004). Introduction to Bayesian Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey.

20.0 Wolpert, Robert L., and Mengersen, Kerrie L. (2004). Adjusted Likelihoods for Synthesizing EmpiricalEvidence from Studies that Differ in Quality and Design:Effects of Environmental TobaccoSmoke. *Statist. Sci.*