USMANU DANFODIYO UNIVERSITY, SOKOYO
(POSTGRADUATE SCHOOL)

ASSESSMENT OF ORDINARY LEAST SQUARE, POLYNOMIAL REGRESSION, THEIL's REGRESSION AND KERNEL REGRESSION ON THEIR PREDICTIVE PERFORMANCE

A Dissertation

Submitted to the

Postgraduate School,

USMANU DANFODIYO UNIVERSITY, SOKOTO, NIGERIA

In Partial Fulfillment of the Requirements

For the Award of the Degree of

MASTER OF SCIENCE (STATISTICS)

By

GARBA, Nasiru

Adm. No16210311007

DEPARTMENT OF MATHEMATICS

February, 2020

CERTIFICATION

This dissertation by GARBA, Nasiru (16210311007) has met the requirements for award of the Degree of Master of Science (Statistics) of Usmanu Danfodiyo University, Sokoto, and is approved for its contribution to knowledge.

……………………….. ………………………

Prof Date

External Examiner

………………………… ……………………

Dr. U. Usman Date

Major Supervisor

………………………… ……………………

Dr. A. B. Zoramawa Date

Co-supervisor I

…………………………. ……………………...

Dr. (Mrs.) H. Usman Date

Co-supervisor II

………………………… ……………………...

Dr. A. I. Garba Date

Head of Department

DEDICATION

I dedicated this Dissertation to my entire family

ACKNOWLEDGMENTS

I would like to express my sincere thanks to the Almighty Allah who enabled the completion of this research.

I am grateful to my supervisory team lead by Dr. U. Usman for their useful ideas, valuable comments, suggestions, encouragement, guidance and patience throughout the period of this study. Without their support the study would have been more difficult journey. I would like to extend my appreciation to all lecturers in the department especially those in statistics unit who had also contributed greatly toward the successful completion of this study. I also appreciate the support and cooperation of my course mates, i also appreciate the effort and support of my entire friends especially Nura Sambo Danchadi and Shamsu Alhaji who one way or the other contributed towards the succesful completion of this study.

I would like to thanks to my entire family especially my mother, my wife and my brothers whose were always in touch praying, given courageous words during difficult and good times, not only for this study but for the whole my entire life.

Finally my special thanks go to Sokoto State University management and entire staff of the University especially those in Academic planning Unit.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

This study investigates the assessment of Ordinary Least Squares (OLS) regression, polynomial regression (PR), theil-sen regression (TLS) and kernel regression (KE) on their predictive performance was undertaken. The assessment was done by examining the sample performances of the estimators when the observation follow normal and log-normal distributions using simulation with 1000, 2000, 3000, 4000 and 5000 samples respectively the mean square error and root mean square error was applied to find out the most efficient among the estimated models. The result shows that when n = 1000 and n=2000 the OLS is efficient than the PR and TLS because it has the least estimates of $\hat{\beta}_0,\ \hat{\beta}_1,\ and\ (\hat{y})\ of\ MSE\ and\ RMSE$ *on both distributions, As the number of observations increases from 3000 to 5000 the(PR) performed better than the OLS and TLS regressions for MSE and RMSE of* $\hat{\beta}_{0,}\ \ and\ \hat{\beta}_1$ *on both distributions, the performance of the estimators MSE*$(\hat{y})$ *and RMSE* $(\hat{y})$ *increases with the increases in the number of observations, however, its shows that the kernel regression estimate performed better than the OLS, TLS and PR regression estimate, because it has the least estimates of MSE*$(\hat{y})$ *and RMSE* $(\hat{y})$ *on both distributions.*

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, and the other independent variables are held fixed. Regression analysis can be defined as the expression of a mean relationship between dependent and independent variables in the form of a mathematical function. This definition assumes a linear relationship between independent and dependent variables. A regression analysis in which only one independent variable is used, is called univariate regression analysis whereas, a regression analysis in which more than one independent variable is used, is called multivariate regression analysis (Erilli and Alakus, 2014).

Regression analysis depends on some assumptions. The most important of these assumptions is that the type of relationship between dependent and independent variables is known. Estimates which are made when there are no assumptions cannot be good estimates. Under such circumstances, in order to make better assumptions, regression methods which enable flexibility in the linearity assumption of the parametric regression are needed. Therefore, these methods are regression models known as nonparametric and semi-parametric regression methods (Erilli and Alakus, 2014).

In general, there will be visible differences between a parametric and a nonparametric regression estimate. It is therefore quite natural to compare the two regression models (Hardle and Mammen, 1993); (Akpos and Jude, 2018) compared parametric and nonparametric regression using normal and non-normal data set with difference sample size in their paper, "Comparison of Theil-sen and Simple regression on normal and non-normal data set with different sample sizes"; (Shah *et al.,* 2016) compared parametric and nonparametric regression using normal distributions data, in their paper, "Comparative study of ordinary least squares regression and Theil-sen regression through simulation in the presence of outliers"; (Jude *et al.,* 2016) Compared parametric regression and nonparametric regressions with normal distribution, in their paper, "Comparison of parametric (OLS) and (Theil's) linear regression" (Erilli and Alakus, 2014) Compared parametric and nonparametric using normal distribution data in their paper, "nonparametric regressions estimations of data with equal values".

1.2 Statement of the Problem

Several researches have been proposed for comparative study of parametric and nonparametric regression Such as: ((Al -Noor and Mohammad, 2013); (Shah *et al.*, 2016)), assessed the performance of classical parametric estimation OLS with the classical Nonparametric Theil's regression. They used (OLS) method while considering outliers data, which has strong influence on the (OLS) method, also the MSE of the (OLS) become very sensitive to these outliers (Gad and Qura, 2016). And (Jude *et al.*, 2016; Akpos and Jude 2018), studied parametric (OLS) and nonparametric Theil's regression using AIC and BIC for measuring the predictive performance, these two traditional model selection criterions favor only the (OLS) since the maximum likelihood cannot be applied to complete nonparametric estimation (Geman and Hwang, 1982) and (Martin and Hjort, 2017)

In this research work Ordinary Least Square (OLS), Polynomials Regression (PR), Kernel Regression (KE) and Theil-sen Regression (TLS) was assessed by applying MSE and RMSE to measure their efficiency using simulated data that follows normal and Log-normal distributions.

1.3 Aim and Objectives

The aim of this research is to assess the performances of OLS, PR, KE and TLS regression

The above aim will be achieved through the following objectives:

i. To simulate data that follows normal and log-normal distributions

ii. To estimate the $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{y}$ of the OLS, PR,TLS and KE

iii. To apply MSE and RMSE for measuring the predictive performance of OLS, PR,TLS, and KE

iv. To assess the predictive performance of the (OLS, PR, KE and TLS)

1.4 Scope and Limitation

This research work focuses only on the assessment of OLS, PR, KE and TLS using normal and log-normal distributions

1.5 Parametric Regressions

Parametric regression is the expression of dependent and independent variables and the average relationship between them variables is expressed through a mathematical function and the clear representation of parameter vectors in this function. For a successful application of parametric regression analysis, assumptions such as normal distribution, homoscedasticity and autocorrelation should be provided. Thus, they are the most powerful regression methods in the event of assumptions' becoming a reality (Erilli and Alakus, 2014). Regression analysis is a tool for exploring the relationship between a response variable y and one or more explanatory variables$x_{1,}\ x_{2,} \cdots,\ x_p$. If p >1 it is called Multiple Regression, but when p =1 it is called bivariate regression or simple regression. If there is a linear relationship between y and p explanatory variables, where β_i 0, 1, 2, ..., p are the parameters of the model, also they are called regression coefficients, and ε_i represents the random error between the observed values of the response variable y and the Functions $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots . + \beta_p x_p$. Suppose in a single explanatory variable, then the simple linear regression model is: y = $\beta_0 + \beta_1 x_1 + \varepsilon_i$ where β_0 is the intercept point of the straight line and $\beta_i$is the slope of the line, this model makes some assumptions about the data.

1.6 Ordinary Least Squares Method

One of the most popular methods to model the function relationship between variables is the Ordinary Least Squares (OLS) estimation procedure, which is very simple and straightforward to apply. However, for OLS estimators to be ideal, some conditions are needed. One of these conditions is that the error terms ε_i are assumed to be independently, identically distributed (iid) random variables with mean zero and a constant variance δ^2. The simple linear regression model is the traditional equation representing the relationship between two variables; the dependent and the independent variables. What is of interest is the estimation of the parameters of the model? The most popular method is the Least Square Method (LSM). But when the data fail to fulfill the assumptions such as the normality, then this parametric method of estimation of the parameters fails to give a valid estimate. In the alternative, a nonparametric method becomes very effective. Non-parametric (or distribution-free) statistical methods are those, which make no assumptions about the population distribution from which the data are taken (Dan and Jude, 2014).

Suppose that the distribution of the errors is not normal. If the errors are coming from a population that has a mean of zero, then the OLS estimates may not be optimal, but they at least have the property of being unbiased. If we further assume that the variance of the error population is finite, then the OLS estimates have the property of being consistent and asymptotically normal. However, under these conditions, the OLS estimates and tests may lose much of their efficiency and they can result in poor performance (Mutan, 2004). To deal with these situations, two approaches can be applied. One is to try to correct non-normality, if non-normality is determined and the other is to use alternative regression methods, which do not depend on the assumption of the normality (Birkes and Dodge, 1993). It may be that a regression function estimate is consistent for a certain class of distributions of (*X*,*Y*), but not consistent for others. It is clearly desirable to have estimates that are consistent for a large class of distributions. In this situation we are interested in properties of m_n that are valid for all distributions of (*X*,*Y*), that is, in distribution free or universal properties. The concept of universal consistency is important in nonparametric regression because the mere use of a nonparametric estimate is normally a consequence of the partial or total lack of information about the distribution of (*X*, *Y*), since in many situations we do not have any prior information about the distribution, it is essential to have estimates that perform well for *all* distributions. Gyorfi *et al.*,(2002).

1.6.1 Non-Parametric Regression

The most important difference between parametric and nonparametric regression methods depends on the trust in the information taken from the researcher and the data about the regression function. In parametric regression, the researcher chooses a possible family of curves from all the curves and needs very special quantitative data about the form of regression function. Nonparametric regression techniques depend on data more than parametric regression techniques in order to get information about the regression function. Thus, they are suitable for inference problems. Also, it is more suitable to use nonparametric estimators when there is no parametric form for the regression function, because when the parametric model is valid, nonparametric models will be less efficient. In addition, nonparametric models can be used to test the validity of parametric models. These are methods used when some assumptions valid for parametric regression methods are not provided. They are effective methods for data which have contradictory sample. In statistical studies, there are robust parametric methods which can address the effects of outliers differently. However, since parameters are spoiled because of outliers, even these robust methods may not generate suitable solutions and the real form of the data may not be reflected in the model. Thus, nonparametric regression provides preliminary information (Härdle, 1994). Although, nonparametric regression does not have restrictive assumptions while making estimations, it has some disadvantages. When there are too many independent variables, it is difficult to make estimations and the graphics may become complicated. In addition, with nonparametric method, it is difficult to take discrete independent variables into consideration and to comment on the individual effects of dependent variables because of the increase in independent variables.

The traditional nonlinear regression model fits the model contains m, $\theta,$ x, ε and y, where θ is a vector of parameters to be estimated, and x is a vector of predictors; the errors ε are assumed to be normally and independently distributed with mean 0 and constant varianceσ^2. The function $m(x,\theta)$, relating the average value of the response y to the predictors, is specified in advance, as it is in a linear regression model. The general nonparametric regression model is written in a similar manner, but the function m is left unspecified: For the p predictors x = $(x_1, x_2 \cdots x_p)$. Moreover, the object of nonparametric regression is to estimate the regression function m(x) directly, rather than to estimate parameters. Most methods of nonparametric regression implicitly assume that m is a smooth, continuous function. As in nonlinear regression, it is standard to assume that$\varepsilon_i \sim NID(0,\sigma^2)$. An important special case of the general model is nonparametric simple regression, where there is only one predictor: Nonparametric simple regression is often called “scatterplot smoothing” because an important application is to tracing a smooth curve through a scatter plot of y against x. We frequently use (John and Weisberg, 2010).

All of these models extend straightforwardly to generalize nonparametric regression, much as linear models extend to generalize linear models. The random and link components are as in generalized linear models, but the linear predictor of the GLM

Advantages of Nonparametric

1. Nonparametric are more robust than parametric test they are valid in a broader range of situation (fewer conditions of validity).
2. Nonparametric techniques does not rely on assumptions (Györfi *et al.,* 2002)
3. Nonparametric is a distribution-free or universal properties (Györfi *et al.,* 2002).

4 With some of nonparametric method the problem of outlier’s data is solve (Sen, 1968).

5 Polynomial Regression

“The goal of polynomial regression is to model a non-linear relationship between the independent and dependent variables (technically between the independent variable and the conditional mean of the dependent variable).” In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as an *n*th degree polynomial in *x*. polynomial regression fits a nonlinear relationship between the value of *x* and the corresponding conditional mean of *y*, denoted E(*y* |x). Sometimes, a plot of the residuals versus a predictor may suggest there is a nonlinear relationship. On way to try to account for such relationship is through a polynomial regression model. Such a model for a single predictor, *X* is:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots \beta_h X^h + \in,$$

Where h is called the degree of the Polynomial, for lower degrees, the relationship has a specific name (i.e., h = 2 is called quadratic, h = 3 is called cubic, h = 4 is called quartic, and so on). Although this model allows for a nonlinear relationship between *Y* and *X*, polynomial regression is still considered linear regression since it is linear in the regression coefficients $\beta_1, \beta_2 \ldots \beta_h$ (Barbeau, 2003).

Even though, a polynomial of degree p is a nonlinear function of the variable x when $p > 1$, it can be regarded as a linear function (affine to be specific) of the variables x, x^2, . . ., and x^p. So, we could fit a linear model to the data as long as the observations $x_{1,} x_2$, . . .,x_n on the single univariate explanatory variable x are replaced by n observations of x and its powers which can be collectd together in a design matrix used for polynomial regression is lm(response poly(regressor, degree)) to which we may want to add the data frame by setting the parameter data (Manuel *et al.*, 2006).

Kernel Regression

The main philosophy of nonparametric regression is to estimate the regression function f using a weighted average of the raw data, where the weights are function of distance in the x -space. In particular, the weights are decreasing function of distance. A weighting scheme of this type is proposed by (Nadaraya –Watson, 1964), in which the weight associated with observations y_i for prediction at x_i is given where $K(u)$ is a decreasing function of u , and $h > 0$ is called the bandwidth or smoothing parameter. $K(u)$, the kernel function, may be taken to be a probability density function such as a Gaussian. The kernel function should be symmetric. The predictions of the kernel regression comes from the fact that the estimated regression function at x_i is obtained by taking a weighted average of the y_j values where the weights w_{ij} are produced by the kernel function, $K(u)$. It is concluded that the selection of smoothing parameter (bandwidth) is much more important than the selection of kernel function for the performance of the kernel regression estimator The kernel function, $K(u)$ is typically chosen to be nonnegative, symmetric about zero, continuous and twice differentiable, some alternative popular (Aydin, 2007).

Theil-sen regression estimation

In nonparametric statistics the theil-sen estimator is a method for robustly fitting a line sample points in the plane (simple linear regression) by choosing the median of the slopes of all lines through pairs of points, it has also been called Sen's slope. It named after Theil Henri and Sen, Parank K who published papers on this method in the year 1950 and 1986 respectively; this estimator can be computed efficiently and is insensitive outliers. It can be significantly more accurate than the non-robust simple linear regression method can compute well against least square even for normally distributed data. It has been called the most popular nonparametric technique for linear trend, the theil-sen estimator, Fundamental of modern statistical methods substantially improving power and accuracy (Hirsch *et al* 1982).

As defined by (Theil, 1950) the theil-sen estimator of a set of two dimensional points (x_i, y_i) is a median M of the slopes $(y_j - \mathrm{y_i})$ $(x_j - x_i)$ determined by all pairs of sample points (Sen, 1968), extended this definition to handle the case in which two data points have the same x-co ordinates. In Sen's definition, one takes the median of the slope defined only from pairs of points have distinct x co-ordinates once the slope M has been determined, one may determined a line from the sample points by setting the y-intercept b, to be median of the values $y_i - mx_i$. The fit line is then the line y = mx+b with coefficients in slope –intercept m and b equivalently $\beta_0 and\ \beta_i$ respectively. The theil-sen estimator is an unbiased estimator of the true slope in simple linear regression. For many distributions of the response error, this estimator has high asymptotic efficient relatively to least squares estimation. Estimator with low efficiency required more independent observations to attain the same sample variance of efficient unbiased estimators.

CHAPTER TWO

LITERATURE REVIEW

This chapter deals with review of some literatures. A comparison of parametric and nonparametric regression is presented in terms of practical problems and measurement on theoretical considerations (Anderson, 1961). Below are some of the researches published, comparing parametric and nonparametric regressions using several areas.

Akpos and Jude (2018) compared parametric and nonparametric regressions on normal and non-normal data set with different sample sizes. For the researchers to know the efficiency of one method over the other, they used Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and applied Mean Square Error (MSE). From the analysis, the result revealed that there is a significant relationship between dependent and independent variables for both the parametric OLS regression and non-parametric Theil's regression with and without residual normality validity. Hence, the parametric OLS regression performs better than its non-parametric Theil's regression since their Residual standard error, AIC and BIC values are all smaller for both the normal and non-normal real data, even though, both models are good in this study, but the OLS is more efficient. AIC and BIC were considered, however the information criterions favor only the (OLS) since the nonparametric has likelihood.

Shah *et al* (2016) compared ordinary least squares regression and Theil-Sen Regression through simulation in the presence of outliers, where they used Ordinary Least Square (OLS) parametric estimation method, simple linear regression is frequently used in agricultural sciences involving multi-location trails. In practice experimental data, obtained from multi-location trials contains outliers. In the conclusion, the OLS estimates are highly affected by the presence of outliers and become less efficient in the presence of outlier as a result lead to wrong conclusions. Theil-sen simple regression is a nonparametric estimation method which is robust to outliers present in the data. Theil-sen regression not only shows consistent performance in the presence of outliers but also a competitor of OLS, considering Ordinary Least Square (OLS) from parametric while using data contain outlier, which has strong influence on OLS method.

Jude *et al* (2016) worked on the comparison of parametric and non-parametric linear regression. First, the set of data was subjected to normality test and it was concluded that all errors in the y-direction are normally distributed (i.e. they follow a Gaussian distribution) concluded that the parametric OLS is better than its non-parametric Theil's regression, since their AIC and BIC are both lower than that of Theil's regression. AIC and BIC were considered in the research that favored only the (OLS)

Ohlson and Kim (2013) carried out a research on Linear Valuation, without OLS: The Theil-Sen Estimation Approach. According to them, OLS confronts two well-known problems in many archival accounting research settings. First, the presence of outliers tends to influence estimates excessively. Second, in the cross-sections, models often build in heteroscedasticity which suggests the need for scaling of all variables. Their study compared the relative efficiency of (Theil, 1950) and (Sen, 1968) (TS) estimation approach vs. OLS estimation in cross-sectional valuation settings. On both criteria, results showed that TS performed much better than OLS. The dominance was most apparent when OLS estimates have the "wrong" sign. TS estimations, by contrast, never lead to such outcomes. Conclusions remained intact even when variables have been scaled for size. Outliers' data was considered which affected the (OLS) method.

Erilli and Alakus (2014), worked on non-parametric regression estimation for data with equal values. The study proposed a new method for the estimation of nonparametric regression parameters with sample data. The method proposed and other nonparametric methods such as Theil, Mood-Brown, Hodges-Lehmann methods and OLS method were compared with the sample data. In the data set which the independent variable had outliers, the OLS estimators gave incorrect values as expected. The proposed method produced more successful results like other nonparametric regression methods. In addition, the proposed methods' results were close to OLS results in the data set which were close to normal distribution and in the data set which the dependent variable had outliers. It showed that the proposed method can be among the alternative nonparametric regression family. They researchers concluded that since the analysis were made without searching if the data had the linear regression assumptions for the OLS method or not, the analysed results were in favor of OLS. The normality assumption regarding error term was not considered as a result of that the (OLS) performed inefficiently.

Dan and Jude (2014), worked on Estimation of Bivariate Regression Data via Theil's algorithm. The method was adopted since all errors in the y-direction are not normally distributed (i.e. do not follow a Gaussian distribution. From the analysis, the result revealed that there exist a significant relationship between weights and shoulder heights of primary school pupils, and the estimated fitted Theil's and it was observed that both the intercept and slope were significant, considering only the nonparametric Theil's regression.

Tim and Sergio (1999), compared parametric with the nonparametric model in their paper where they used multi-output distance functions to investigate technical inefficiency in European railways. Finally, the paper concludes with the suggestion that a parametric distance function can be used as an alternative method. Where they considered linear programming, data envelopment analysis (DEA), corrected least squares (COLS), input-oriented, output-oriented and constant returns to scale (CRS).

Park and Park (2003), compared the parametric and nonparametric models where the logit model is used. However, the exact form of the response curve is usually unknown and even very complicated, so it is likely that the true model does not follow the logit model. Therefore concluded that nonparametric model should be used rather than logit model of parametric. They considered locally weighted quasi-likelihood estimator which is very competitive in the small sample situation.

Jou *et al* (2009), compared parametric density functions with nonparametric Fourier series and estimated the annual precipitation for five old rain gauge stations (Bushehr, Isfahan, Meshed, Tehran and Jask) in Iran. The nonparametric approach is Fourier series method. Thus, the Fourier series of nonparametric can be used as a better alternative approach for precipitation frequency analysis. Where They considered the following density functions for parametric; normal, two and three parameters log-normal, two parameters gamma, Pearson and log-Pearson type 3 and Fourier series for nonparametric. However these methods have been successfully applied in many cases, but have some problem because of not fitting to the observed data very well or diverting from the extreme tails. Some other problems involved in selection of these methods are difficulties to choose the best distribution function and estimation of their parameters particularly for skewed data. And also considered data contains outlier

Fort *et al* (1996), worked on parametric and nonparametric estimation of speech formants; Application to infant cry. The performance comparison between some classical parametric and non-parametric estimation techniques, Concluded that parametric method to be more reliable and robust against noise. They considered robustness which capitalized on data with outliers.

Sharma *et al* (1999), analyzed Technical allocative and economic efficiencies in swine production in Hawaii: A comparison of parametric and nonparametric approaches in technical, allocative and economic efficiency measures are derived for a sample of swine producers in Hawaii using the parametric stochastic efficiency decomposition technique and nonparametric data envelopment analysis (DEA). Efficiency measures obtained from the two frontier approaches are compared. Firm-specific factors affecting productive efficiencies are also analyzed, where they find out that the Data Envelopment Analysis (DEA) of nonparametric is more robust than those from parametric. Considering data contains outliers which has strong influence on the (OLS) method.

Adamowski (1989), studied Monte Calor Comparison of parametric method and nonparametric estimation of flood frequencies by using the relationship of flood magnitude to frequency of occurrence which can be estimated from observed annual flood data by the parametric method of fitting any of various theoretical distributions to the data, or by the nonparametric method with variable kernel, the parametric and nonparametric methods were investigated in his paper and were compared. In his Conclussion that nonparametric methods are accurate, uniform, and particularly suitable for multimodal data. In addition, the variable kernel method provides more accurate estimates of the tail of a distribution, considering the log-Pearson type 3 distributions and data contain outliers

Jones *et al* (2012), compared the use of parametric and nonparametric approaches to adjust for heterogeneity in self-reported data. Where they concluded that in general Parametric and Nonparametric estimation produced similar results in term of first order stochastic dominance for domains of both mobility memory, neither method consistently explained discrepancies across countries between self reported data and objectives measures of health mobility and memory. However this research fail to find a consistent pattern in the result to suggest that either approach satisfactorily addresses the issue of differential reporting.

Diggle *et al* (2000), Compared nonparametric (design-based) and parametric (model-based) approaches to the analysis of data in the form of replicated spatial point patterns in two or more experimental groups. Basic questions for data of this kind concern estimating the properties of the underlying spatial point process within each experimental group, and comparing the properties between groups. Concluded that the parametric approach can be more efficient when the underlying model assumptions hold, but potentially misleading otherwise, considering K function using bootstrap testing procedure.

Al-Noor and Mohammad (2013), Model of robust regression with parametric and nonparametric method. They evaluated the performance of the classical parametric estimation method ordinary least squares with classical nonparametric estimation Theil's regression method, some robust estimation methods and two suggested methods for conditions in which varying degrees and direction of outliers are presented in the observed data. They concluded that nonparametric introduce better performance in the presence of the outliers x-direction and xy-direction compared to Ordinary least square (OLS), LAD and M-estimators Considering (OLS) of parametric while using data contains outliers which has strong influence on the method of OLS and may finally lead to wrong conclusions.

Mays *et al* (2001), worked on the model robust combining parametric and nonparametric regression procedures can improve upon the shortcomings of each when used individually. The results establish the accuracy of the theoretical formulas and illustrate the potential benefits of the model-robust procedures. They considered data contains outliers

Smith *et al* (2002), worked on comparison parametric and nonparametric model for traffic flow forecasting in Single point short-term traffic flow forecasting will play a key role in supporting demand forecasts needed by operational network models. Seasonal autoregressive integrated moving average (ARIMA), a classic parametric modelling approach to time series, and nonparametric regression models have been proposed as well suited for application to single point short-term traffic flow forecasting. They effort is to examine the theoretical foundation of nonparametric regression, the nonparametric regression base on heuristically improved forecast generation method approach of the single interval traffic flow prediction performance of seasonal ARIMA models. They considered ARIMA models

Wang and Small (2012), Compared of parametric and Non-parametric estimates of Attributable fraction for a semi-continuous exposure, the attributable fraction of a disease due to an exposure is the fraction of disease cases in a population that can be attributed to that exposure. They consider the attributable fraction for a semi-continuous exposure that is an exposure for which a clump of people have zero exposure and the rest of the people have a continuously distributed positive exposure. They concluded that the nonparametric regression estimator work well and improved considerably on the classical estimator. Considering classical method on sample averages parametric regression method such as logistic regression model and power model, for nonparametric regression local linear smoothing and isotonic regression were used.

Zamorano and Cervera (2001), compared Parametric frontier models and non-parametric methods have mono-polised the recent literature on productive efficiency measurement. The result show that parametric model are extremely consistent in ranking their fair wise correlation coefficients are not less than 99% the correlation is also high between parametric technique and DEAC. Considering data contains the outliers.

Jonathan and Tam (1998), compared parametric regression (ordinary least squares) and nonparametric (Theil's) regression where the result shows that, the Theil's method of nonparametric are very strong alternative ordinary least squares of parametric in the presence of the outliers. Considering data contains outliers and condition in which the normality assumption regarding error distribution is violated.

Al-Homsi (2017), Compared three Non-parametric regression method Kernel Estimators, Theil Estimators and Siegel Estimators. Through simulation study, it is found that nonparametric regression methods are not sensitive to the presence of outliers, and performed well in many different contamination models, with advantage to Thiel-sen estimators, according to the quality of estimated parameters. In addition, it turns out that Kernel estimators using normal kernel function are the best estimators according to standard error of regression. Only nonparametric methods were considered and outliers' data was used.

Having reviewed the above literature, we are able to discover that different performance measures have been used to fish out the best-performed techniques, and many recommendations have been made where necessary.

.

CHAPTER THREE

3.1 MATERIALS AND METHOD

In this research work, Monte Carlo simulations with R Programming software using Normal N($\boldsymbol{\mu}, \boldsymbol{\sigma}$) and log-normal distributions for the dependent and independent variables was used to assesses the predictive performance of the Ordinary Least Squares (OLS), Polynomial Regression (PR), Kernel Regression (KE) and Theil's Regression (TLS)

3.2 Ordinary Least SQUARE Regression

Consider the model function

$$y = \beta_0 + \beta_1 x + \varepsilon_i \tag{3.1}$$

This relationship between the true (but unobserved) underlying parameters β_0 and $\beta_1$and the data points is called standard linear regression model:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{3.2}$$

Or equivalently:

$$\mathrm{Y} = \mathrm{X}\beta + \varepsilon, \qquad (3,3)$$

where *Y* is the vector of the n observed values of the response variable, *X* is a *n* x 2 matrix, β is a vector of the parameters, and ε is the vector of the errors. The estimates for β_0 and β_1 are $\hat{\beta}_0$ and $\hat{\beta}_1$ which can be find out by solving the two equations simultaneously

$$\textstyle\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0 \tag{3.4}$$

$$\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i = -\hat{\beta}_1 \sum_{i=1}^{n} {x_i}^2 = 0 \tag{3.5}$$

Finally result to

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sum_{i=1}^{n} {x_i}^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} \quad (3.6)$$

$$\beta_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3.7)$$

The design matrix is

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (3.8)$$

The difference between the observed value y and the corresponding fitted value y_i is the value $\hat{y}_i$ of the i^{th} residual, and it is denoted by e_i the sum of the residuals in linear regression model equals to zero $(\sum_{i=1}^{n} e_i) = 0$ (Montgomery *et al.*, 2012).

3.3 The polynomial regression

$$yi = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_m x_i^m + \in_i \ (i = 1, 2, \ldots . n) \quad (3.9)$$

Can be expressed in matrix form in terms of a design matrix X, a response vector $\vec{y}$, a parameter vector $\vec{\beta}$, and a vector $\vec{\in}$, of random errors. The i-th row of X and $\vec{y}$ will contain the *x* and *y* value for the i-th data sample.

Then the model can be written as a system of linear equations:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^m \\ 1 & x_2 & x_2^2 & \ldots & x_2^m \\ 1 & x_3 & x_3^2 & \ldots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \in_1 \\ \in_2 \\ \in_3 \\ \vdots \\ \in_n \end{bmatrix} \quad (3.10)$$

Which when using pure matrix notation is written as

$$\vec{Y} = X\vec{\beta} + \vec{\in} \quad (3.11).$$

The vector of estimated polynomial regression coefficients (using ordinary least squares estimation) is

$$\hat{\vec{\beta}} = (X^{T}X)^{-1} X^{T}\vec{y}, \qquad (3.12)$$

Assuming $m < n$ which is required for the matrix to be invertible; then since X is a Vandermonde matrix, the invertibility condition is guaranteed to hold if all the x_i values are distinct. This is the unique least squares solution.

3.4 Kernel Regression estimate

$$y = m(x,\theta) + \varepsilon \qquad (3.13)$$

Where θ is a vector of parameters to be estimated,Y and x is a vector of predictors; the errors ε are assumed to be normally and in dependently distributed with mean 0 and constant variance σ^2.

The function $m(x,\theta)$, relating the average value of the response y to the predictors, is specified in advance, as it is in a linear regression model.

The general nonparametric regression model is written in a similar manner, but the function m is left unspecified:

$$y = m(x) + \varepsilon \qquad (3.14)$$

The Kernel estimate of f is defined as

$$f_h = \frac{1}{nh}\sum_{j=1}^{n} k\left(\frac{x - x_j}{h}\right) \qquad (3.15)$$

kernel weights which is given by

$$W_i = \frac{\sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^{n} k\left(\frac{x-x_j}{h}\right)} \qquad (3.16)$$

$$\hat{f}_h(x_i) = w_i y_i \qquad (3.17)$$

$$\hat{f}_h(x_i) = \frac{\sum_{j=1}^{n} k\left(\frac{x-x_i}{h}\right) y_i}{\sum_{j=1}^{n} k\left(\frac{x-xj}{h}\right)}, \; i = 1, 2, \cdots, n. \qquad (3.18)$$

Where $K(.)$: is a Kernel function or shape function and it is non-negative function.

h: is called bandwidth, window width, or smoothing parameter, yi is a random variable where i = 1, 2, - - -, n., xi is a random variable where i = 1, 2 , - - -, n and xj is a random variable at J^{th} observation where j = 1, 2, - - -, k

In fact, if the space of X is very unevenly, then the Kernel estimator will give poor

Result. This problem was solved by (Nadaraya, 1964) and (Watson, 1964), they proposed

And the formula of kernel estimator by using this weight is defined as;

$$\hat{f}_h(x) = \frac{1}{n} \sum_{j=1}^{n} W_j \, y_j \, , \qquad (3.19)$$

The above estimator is called Nadaraya-Watson estimator, the shape of Kernel weights is Determined by the Kernel function K , and the size of weights is parameterized by Bandwidth h (Härdle, 1994).

Theil-sen regression estimate

$$y = mx + b; \qquad (3.20)$$

Where m is the slope y-intercept b to be the median of values: y_i – mx_i. The fit line is then line y = mx + b with coefficients m and b in slope and intercept form, ie. β_0 *and* β Respectively.

Mean Square Error

the mean squared error (MSE) or mean squared deviation (MSD) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors or deviations that is, the difference between the estimator and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate. The MSE is a measure of the quality of an estimator, it is always non-negative, and values closer to zero are better.

if $\hat{Y}$is a vector of n predictions, and Y is the vector of observed values corresponding to the inputs to the function which generated the predictions, then the MSE of the predictor can be estimated by

$$\text{MSE} \;=\; \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (3.21)$$

I.e., the MSE is the mean of the square of the errors. This is an easily computable quantity for a particular sample (and hence is sample-dependent).

Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) (also called the root mean square deviation, RMSD) is a frequently used measure of the difference between values predicted by a model and the values actually observed from the environment that is being modeled. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power.

The RMSE of a model prediction with respect to the estimated $\hat{y}_i$ variable is defined as the square root of the mean squared error:

$$\text{RMSE} \;=\; \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (3.22)$$

Where y_i is the observed value and $\hat{y}_i$ is the estimated value

Simulation Settings

In this research work, Monte Carlo simulations with R was used *to* evaluate the predictive ability of the parametric and nonparametric regression we considered random simulation, the Normal N($\boldsymbol{\mu}, \boldsymbol{\sigma}$) and log-normal distributions for the dependent and independent variables was also used.

CHAPTER FOUR

4.0 RESULT AND DISCUSSION

Table (4.1): The Result of MSE and RMSE of Point Estimates by Normal Distribution at n = 1000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	0.00152	0.032590	0.180527	0.08076	0.03216	0.17933	0.6310	0.7943
PR	0.00927	0.03989	0.199730	0.80613	0.03217	0.17936	1.0110	1.0054
TLS	0.00659	0.474650	0.688940	0.06678	0.43115	0.65662	0.9692	0.9844
KE	------	------	------	------	------	------	0.9923	0.9956

The result of the table (4.1) shows that the (OLS) method is efficient than the (PR) and (TLS), because it has the least estimate of MSE $(\hat{\beta}_0)$; MSE$(\hat{\beta}_1)$ and MSE$(\hat{y})$ with their root mean squares (RMSE) on normal distribution.

Table (4.2): The Result of MSE and RMSE of Point Estimates by Normal Distribution at n = 2000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	0.00522	0.15050	0.38794	0.030301	0.022621	0.15040	0.8956	0.9463
PR	0.002913	0.15097	0.38855	0.049610	0.07090	0.26627	1.0000	1.0000
TLS	0.021155	0.67590	0.82213	0.059926	0.38329	0.61910	0.9923	1.9961
KE	-----	-----	------	------	------	------	0.9639	0.9817

The result of the table (4.2) shows that the (OLS) method is efficient than the (PR) and (TLS), because it has the least estimate of MSE $(\hat{\beta}_0)$; MSE$(\hat{\beta}_1)$ and MSE$(\hat{y})$ with their root mean squares (RMSE) on normal distribution.

Table (4.3): The Result of MSE and RMSE of Point Estimates by Normal Distribution at n = 3000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE $(\hat{y})$
OLS	0.018924	0.018264	0.13514	0.007965	0.01322	0.13535	1.2898	1.1357
PR	0.014668	0.01420	0.11916	0.007415	0.01049	0.10242	0.9930	0.9965
TLS	0.016310	0.488428	0.69590	0.051930	0.38216	0.61819	0.9960	0.9979
KE	-----	----	-----	------	------	------	0.9910	0.9954

The result of the table (4.3) reveals that the (PR) is efficient than the (OLS) and (TLS) because it has the least estimates of MSE$(\hat{\beta}_0)$ and MSE$(\hat{\beta}_1)$ with their root mean squares errors (RMSE) on normal distribution. However the Kernel regression estimate (KE) is efficient than the (OLS) and (TLS) because it has the least estimate of MSE$(\hat{y})$ and RMSE$(\hat{y})$

Table (4.4): The Result of MSE and RMSE of Point Estimates by Normal Distribution at n = 4000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	0.008856	0.015740	0.12599	0.002058	0.00158	0.03974	1.0040	1.0019
PR	0.003880	0.013130	0.11458	0.00140	0.00095	0.03082	0.9922	0.9961
TLS	0.023900	0.453790	0.45379	0.013320	0.389890	0.62441	0.9963	0.9981
KE	-----	-----	------	------	------	------	0.9668	0.9832

The result of the table (4.4) shows that the (PR) is efficient than the (OLS) and (TLS) because it has the least estimates of MSE$(\hat{\beta}_0)$ and MSE$(\hat{\beta}_1)$ with their root mean squares errors (RMSE) on normal distribution. However the Kernel regression estimate (KE) is efficient than the (OLS) and (TLS) because it has the least estimate of MSE$(\hat{y})$ and RMSE$(\hat{y})$

Table (4.5): The Result of MSE and RMSE of Point Estimates by Normal Distribution at n = 5000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	0.001704	0.015589	0.12485	0.004652	0.01526	0.12353	1.9857	1.4091
PR	0.001520	0.00145	0.03820	0.001781	0.00243	0.04929	0.9883	0.9941
TE	0.008362	0.447953	0.66928	0.002720	0.41939	0.64760	1.0050	1.0024
KE	-----	------	-----	------	------	------	0.2044	0.1429

The result of the table (4.5) indicates that the (PR) is efficient than the (OLS) and (TLS) because it has the least estimates of MSE$(\hat{\beta}_0)$ and MSE$(\hat{\beta}_1)$ with their root mean squares errors (RMSE) on normal distribution. However the Kernel regression estimate (KE) is efficient than the (OLS) and (TLS) because it has the least estimate of MSE$(\hat{y})$ and RMSE$(\hat{y})$

Table (4.6): The Result of MSE and RMSE of Point Estimates by Log- Normal Distribution at n = 1000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	1.65783	0.08565	0.29266	0.02879	0.00064	0.02529	2.1100	1.4525
PR	1.85403	0.09899	0.31463	0.13952	0.00519	0.22789	2.1950	1.4815
TLS	1.03819	0.55204	0.74299	0.03684	0.26926	0.51890	2.3340	1.5277
KE	------	------	------	------	------	------	2.2089	1.4862

The result on the table (4.6) has shown that the (OLS) method is efficient than the (PR); (TLS) and (KE) because it has the least estimate of MSE$(\hat{\beta}_0)$; MSE $(\hat{\beta}_1)$ and MSE$(\hat{y})$ and their RMSE on log-normal distribution.

Table: (4.7): The Result of MSE and RMSE of Point Estimates by Log- Normal Distribution at n = 2000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	1.64300	0.00634	0.07962	0.00211	0.00213	0.04615	1.8963	1.3770
PR	0.64211	0.00831	0.09115	0.00103	0.00253	0.05029	2.0150	1.4195
TLS	0.95382	0.50103	0.70783	0.00383	0.29939	0.54716	2.2090	1.4862
KE	------	-----	------	-------	-------	-------	1.9094	1.3818

The result on the table (4.7) points out that the (OLS) method is efficient than the (PR); (TLS) and (KE) because it has the least estimate of MSE$(\hat{\beta}_0)$; MSE $(\hat{\beta}_1)$ and MSE$(\hat{y})$ and their RMSE on log-normal distribution.

Table (4.8): The Result of MSE and RMSE of Point Estimates by Log- Normal Distribution at n = 3000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	1.62749	0.04719	0.21724	0.004069	0.01736	0.13174	2.1050	1.4663
PR	1.62299	0.03595	0.18960	0.000341	0.00227	0.04762	2.0450	1.4300
TLS	0.95425	0.52319	0.72330	0.006974	0.27659	0.52590	2.1550	1.4679
KE	------	------	------	------	------	------	2.0401	1.4283

The result on the table (4.8) indicates that the (PR) method is efficient than the (OLS) and (TLS) because it has the least estimate of MSE$\hat{\beta}_0$; MSE$(\hat{\beta}_1)$ and their RMSE on log-normal distribution, however the kernel regression estimate is efficient than the (OLS); (PR) and (TLS) because it has the least estimate of MSE$(\hat{y})$ and RMSE$(\hat{y})$ on log-normal distribution.

Table (4.9): The Result of MSE and RMSE of Point Estimates by Log- Normal Distribution at n = 4000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	1.59066	0.03857	0.19639	0.03268	0.04410	0.21000	2.8929	1.7008
PR	1.00302	0.01608	0.12680	0.01196	0.02763	0.01662	1.9214	1.3861
TLS	1.00539	0.52650	0.72560	0.01663	0.29663	0.54443	2.0431	1.4293
KE	------	------	-------	------	-----	------	1.9091	1.3817

The result on the table (4.9) shows that the (PR) method is efficient than the (OLS) and (TLS) because it has the least estimate of MSE$\hat{\beta}_0$; MSE $(\hat{\beta}_1)$ and their RMSE on log-normal distribution, however the kernel regression estimate is efficient than the (OLS); (PR) and (TLS) because it has the least estimate of MSE$(\hat{y})$ and RMSE$(\hat{y})$ on log-normal distribution.

Table (4.10): The Result of MSE and RMSE of Point Estimates by Log- Normal Distribution at n= 5000

Method	$\hat{\beta}_0$	MSE $(\hat{\beta}_0)$	RMSE $(\hat{\beta}_0)$	$\hat{\beta}_1$	MSE $(\hat{\beta}_1)$	RMSE $(\hat{\beta}_1)$	MSE $(\hat{y})$	RMSE$(\hat{y})$
OLS	1.61091	0.03402	0.18444	0.01509	0.01191	0.10913	2.3291	1.5261
PR	1.60666	0.03143	0.17728	0.01479	0.00077	0.02778	1.9631	1.4011
TLS	0.94621	0.52890	0.72725	0.02431	0.26685	0.51652	2.0640	1.4367
KE	------	------	------	------	------	------	1.6910	1.3004

The result on the table (4.8) shows that the (PR) method is efficient than the (OLS) and (TLS) because it has the least estimate of MSE$\hat{\beta}_0$; MSE $(\hat{\beta}_1)$ and their RMSE on log-normal distribution, however the kernel regression estimate is efficient than the (OLS); (PR) and (TLS) because it has the least estimate of MSE$(\hat{y})$ and RMSE$(\hat{y})$ on log-normal distribution.

4.11 Discussion of Result

The tables (4.1), (4.2), (4.6), and (4.7) shows that the estimate of (OLS) is efficient than the (PR); (KE) and (TLS) because they have the least estimates of MSE$(\hat{\beta}_0)$, MSE$(\hat{\beta}_1)$, MSE$(\hat{Y})$ and their mean square errors (RMSE), While for tables (4.3), (4.4), (4.5), (4.8), (4.9) and (4.10). The kernel estimators with normal kernel function have the least estimates of MSE$(\hat{y})$ and RMSE$(\hat{y})$ on both distribution as number of observation (n) increases from n=3000 to n=5000. However, PR has the least estimate of MSE$(\hat{\beta}_0)$; MSE$(\hat{\beta}_1)$ and their RMSE (Root Mean Square Errors) in both distributions

CHAPTER FIVE

5.0 SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

We have seen the performance of all the methods, in which the PR has good estimated parameters MSE and RMSE values for $\hat{\beta}_0 and\ \hat{\beta}_1$ as the sample size increases from 3000 to 5000 in all the distributions. Generally, in most cases the kernel regression estimates have the least estimated $MSE\ (\hat{y})$ and $RMSE\ (\hat{y})$ in both distributions

5.2 Conclusions

The result shows that when n = 1000 and n = 2000 the OLS is efficient than the (PR), (TLS) and (KE), simply because, it has the least estimates of $\hat{\beta}_0,\ \hat{\beta}_1 and\ (\hat{y})\ for\ MSE\ and\ RMSE$ on both distributions. As the (number of observations) sample size, n increases from 3000 to 5000, the (PR) performed better than the (OLS) and (TLS) regressions for MSE and RMSE of $\hat{\beta}_0\ and\ \hat{\beta}_1$ on both distributions. The performance of the estimators MSE$(\hat{y})$ and RMSE $(\hat{y})$ increases with the increase of the number of observations, however the kernel regression estimate performed better than the (OLS), (TLS) and (PR) regression estimate in both distributions.

5.3 Recommendations

Finally, it is recommended that nonparametric regression should be used when we have very large number of observation and also when the nature of the distribution is not known.

5.4 Contribution to knowledge

The study has contributed to knowledge in the following directions. The study identified the weakness in the efficiency of (Jonathan and Tam, 1998); (Shah *et al.*, 2016); (Al-Noor and

Muhammad, 2013); (Erilli and Alakus, 2014) in terms of considering data contains outliers while using ordinary least squares estimator (OLS).

It has also identified the weakness in the efficiency of Jude *et al.*, (2016); Akpos and Jude, (2018). On comparing parametric and Non-parametric regression considering AIC and BIC

5.5 Suggestion of Further studies

I therefore, suggested that future research should look at a similar work by considering three nonparametric regression methods, such as kernel estimate, using Normal and log-normal distributions, K-nearest neighbor estimates using Normal, log-normal distributions, smoothing Splines using Normal, log-normal distributions and one parametric other than ordinary least squares like least squares, absolute deviation or Generalize least squares, measure their performance by MSE, RMSE and compare the performances for both distribution using simulation.

References

Adamowski, K.(1989). A Monte Carlo comparison of parametric estimation of flood frequencies. *Journal of Hydrology, 108*(1), 295-308.

Akpos, E. P., & Jude, O. (2018). Comparison of Theil's and simple regression on normal and non-normal data set with different sample size. *International Journal of Managment and applied science, 4(1), 70-74*

Al-Homsi, Z. A. (2017). Comparison among some Nonparametric Methods in estimating Regression Models. (Master's Degree of Science in Mathematics), Zarqa University, for the Master's Degree of Science in Mathematics.

Al-Noor. N., & Muhammad, A.(2013). Model of Robust Regression with parametric and Nonparametric Methods. *Mathematical Theory and Modeling, 3*(1), 27-39.

Anderson, H. N. (1961). Scale and statistics: Parametric and Nonparametric *Psychological Bullettin, 58*(4), 305-316.

Aydin, D. (2007), Comparison of the Nonparametric Regression Models Using Smoothing Spline and Kernel regression, *International Journal of Mathematical and Computational Sciences*, *1*(12), 23-35.

Barbeau, E. J. (2003). Polynomial (Problem books in Mathematics) *Polynomial* (Second edition ed., Vol. 13). New York: Springer.

Birkes, D., & Dodge, Y.(1993). Alternative Method of Regression, in S. Weisberg (Ed.), *Probability and Statistics*. UK: Wiley Series.

Dan, E. D., & Jude, O.(2014). Estimation of Bivariate Regression data via Theil's Algorithm. *Journal of Emerging Trends In Engineering and Applied Sciences*, *5*(8), 29-34.

Diggle, P. J., Mateu, J., & Clough, H. T.(2000). Comparison between parametric and Nonparametric Approaches to the analysis of replicated spatial point patterns. *Advance In Applied Probability*, *32*(2), 331-343.

Erilli, A. N., & Alakus, K. A.(2014). Nonparametric Regression Estimation for data with equal values. *European Scientific Journal*, *10*(4), 70-80.

Fort, I. A., Manfred C., & B. P.(1996). Parametric and Nonparametric estimation of Speech formant application of infant cry. *Anal. of Statistics, 18(8), 677-691.*

Gad, A. M., & Qura, M. E. (2016). Regression estimation in the presence of outliers: A Comparative study, *International Journal of probability and Statistics, 9*(3), 56-72.

Geman, S., & Hwang, c.-R.(1982). nonparametric Maximum Likelihood estimation by the method of sieves. *Anal. of Statistics, 10*(2), 401-414.

Gyorfi, L., Zyzak, A. K., Kohler, M., & Walk, H,(2002). Distribution -free Theory of Nonparametric Regression. In P. Bickel (Ed.), *Springer Series in Statistics* (pp.662) New York: Springer-Verlag New York Berlin Heidelberg.

Hardle, W. (1994) applied nonparametric method. *Economic Research Cambridge Univerty press,* 426.

Hardle, W., & Mammen, E. (1993). Comparing parametric versus nonparametric regression fits. *Annals.of Statistics, 21*(4), 1926-1947.

Hirsch, R. M., Slack, J. R., & Smith, R. A. (1982). teaching of trend Analysis monthly Water quality data. *Hydrology Research, 18*(1), 32-35.

John, F., & Weisberg, S (2010). Nonparametric Regression in R In V. knight (Ed.), *An R Companion To Applied Regression* (2nd ed, Vol. 2). SAGE Publications, Ltd 1 Oliver's yard 55 City Road London Oliver

Jonathan, N., & Tam, H. P. (1998). Comparison of Rebust and Nonparametric Estimators Under Simple Linear Regression Model. *Multiple Linear Regression viewpoint, 25*(1), 28-33.

Jones, A. M., Rice, N.,& Robone, S. (2012). A Comparison of parametric and nonparametric adjustment using Vignettes for self reported data. *Health Econometrics and data group, 3*(5), 27-35.

Jou, P. H., Akhoondali, A. M., Behnia, A., & Chinipardaz, R. (2009). Comparison of parametric and Nonparametric Density function for estimating Annual precipitation in iran. *research journal of Environmental sciences, 3*(1), 62-70.

Jude, O., Iheagwara, A. I., & Idoch, O. (2016). Comparison of parametric (OLS) and Nonparametric (THEIL'S) linear regression *Advance Research Journal of multi-Disciplinary Discoveries, 2*(1), 24-29.

Manuel, B., Alicia, D., Emiris, Z. I. (2006). Solving polynomial Equations foundations *Algorithms & Applications, 4*(6), 34-43.

Martin, J. & Nils L. H.(2017). Parametric or Nonparametric the FIC Approach *Anal. of Statistics 951-981.*

Mays, J. E., Jeffrey B. Birch, B., & Alden, S. (2001). model robust regression combining parametric and non-parametric and semi-parametric methods. *Journal of Non parametric Statistics, 13(2), 245-277.*

Montgomery, D. C., & Peck, G., E. A. (2012). Introduction to Linear Regression Analysis. (pp 672), 5^{th} ed.

Mutan, O. C. (2004). A *Comparison of Regression Techniques via Monte Calro Simulation* (Master of Science in Statistics) A thesis Submitted to the School of Natural & Applied Sciences of Middle East Technical University.

Nadaraya, E.A. (1964), On estimating regression, *Theory of Probability and its Applications, 9(1), 141-142.*

Ohlson, James A. & Seil K. (2013). Linear valuation without OLS: Theil-sen Estimation Approach, *Electronic Journal. 20(1), 216-221*

Park, D., & Park, S.(2003). parametric and Nonparametric estimators ED100. *Journal of Statistical Computation Simulation 76*(8), 661-672.

Sen k. parank & Henri Theil 1968, Economic information Theory 19(2), 185-188.

Shah, S. H., Rashid, A. T., Kanim, J., & Shah, S. M. (2016). Comparative study of ordinary least square regression and Theil's regression in the presence of outliers. *Science and Technology, 5(11), 137-142.*

Sharma, K. R., Leung, P., & Zaleski, H. M.(1999). Technical, allocative and economic efficiencies in swine production in Hawaii: a comparison of parametric and nonparametric approaches. *Agricultural Economics, 20(1), 23-35.*

Smith, B. L., Williams, B. M., & Oswald, R. K. (2002). Comparison of parametric and Nonparametric Models for traffic flow forecasting. *Trasport Research part C emerging technologies, 10*(1), 303-321.

Theil, Henr (1950). A rank-invariant method of linear and polynomial regression I, II, III, *Nedel Akad, Wetensch, proc., 53, 386-392*

Tim, C., & Sergio, P. (1999). A. Comparison of parametric and nonparametric distance function. *Journal of Operational Research, 117*(2), 326-339.

Wang, W., & small, D. (2012). comparative study of parametric and nonparametric estimates of the attributable fraction for semi - continuous exposure. *International Journal of Biostatistics, 14*(2), 243-260.

Watson, G.S., (1964). Smoothing Regression Analysis Sankhya *The Indian Journal of Statistics Series A., 26*(4), 359-372.

Zamorano Luis R. Murillo, & Juan A. Vega – Cervera. (2001). the use of parametric and non-parametric frontier method to measure the productive efficiency in the industrial sector: A comparative study. *International Journal of Production Economics*. 69(3), 265-275