Unilag Journal of Mathematics and Applications,
Volume 2, Issue 1 (2022), Pages 60--.73
ISSN: 2805 3966. URL: http://lagjma.edu.ng

COMPARATIVE STUDY OF BAYESIAN AND ORDINARY LEAST SQUARES APPROACHES

OGUNDEJI, ROTIMI*, ONYEKA-UBAKA, JOSEPHINE AND YINUSA, EMMANUEL

ABSTRACT. Frequentist (Classical) and Bayesian Statistics are two major approaches to data analysis in statistics; however, the difference is how both see a parameter. Frequentists see a parameter as constant value while the Bayesians see it as random variable. Resent research has witnessed increase in the application of Bayesian methods to statistical problems and in other fields. For linear regression modelling, frequentists use more often the Ordinary Least Squares (OLS) method despite violation of some assumptions. Bayesian approach can be used when assumptions in linear regression model using OLS are not met. The study objective is to ascertain if datasets provided meet the classical OLS assumptions or not, and demonstrate the prominence of Bayesian methods where OLS assumptions are violated. Two different data sets were adopted in this paper for a comparative study using both OLS and Bayesian approaches to linear regression modelling. The analysis showed that the resulting linear regression model using OLS does not meet all required assumptions for a good model. The Bayesian approach as an alternative to regression modelling was further established based on results of smaller standard errors. The results showed that linear regression modelling using Bayesian approach is better than Frequentist method using OLS regression modelling.

1. INTRODUCTION

The world of Statistics is divided into two schools of reasoning based on their respective paradigms or philosophies. This includes the Classical (often referred to as Frequentist) paradigm and Bayesian paradigm. The key difference between Bayesian statistical inference and frequentist concerns the nature of the unknown parameters. In the frequentist framework, a parameter of interest is assumed to be unknown, but fixed (e.g. regression coefficient). Hence, many analysts of social and behavioural sciences regression models favoured ordinary least squares regression modelling for its simplicity [1]. [2] investigated least squares method, non-

---

2010 *Mathematics Subject Classification*. Primary: 22E30. Secondary: 58J05.
Key words and phrases. Bayesian Methods, Frequentist, Regression Models, Assumptions Violation, Diagnostics Tools.

Submitted: October 26, 2021. Revised: June 6, 2022. Accepted: June 30, 2022.

*Correspondence

parametric method and robust regression methods to estimate the parameters of multiple regression models.

To evaluate these methods, measurements of body weight, total length and fork length of fishes collected from Serranus cabrilla were used. In these regression models, body weight was dependent variable whereas total length and fork length were independent variables. [3] investigated residual analyses as OLS regression without stretching assumptions.

In the Bayesian view of subjective probability, all unknown parameters are treated as uncertain and therefore should be described by a probability distribution. Bayesian statistics studies and applies probabilities to statistical problems, providing the tools to update beliefs about a statistic in the evidence of new data. Bayesian statistics is founded on Bayes' theorem which uses a method of revising probability estimates as new information becomes available. The probability before the new data becomes available is referred to as the prior probability, and the revised probability using the new data is referred to as the posterior probability. Whenever newer data becomes available, the current posterior probability becomes the new prior probability [4].

Today, Bayesian statistical softwares are widely used by researchers in diverse fields due to significant computational advancements including Markov Chain Monte Carlo (MCMC) [5], OpenBUGS [6], JAGS [7], [8], MLUS [9] and WinBUGS software [10]. Researchers in many fields have embraced the Bayesian approach due to its capacity to handle complexity in real world problems. [11] investigated the results of applying a Bayesian deconvolution method to several XRF spectra and compare them to conventional methods. The spectra are selected to illustrate both the advantages and disadvantages of this method. [12] compared the prediction accuracy of 92 infrared prediction equations obtained by different statistical approaches. The statistical methods used to develop the prediction equations were partial least squares regression (PLSR), Bayesian ridge regression, Bayes A, Bayes B, Bayes C, and Bayesian least absolute shrinkage and selection operator. In validation sets, Bayesian regression models performed significantly better than PLSR for the prediction of 33 out of 92 traits. [13] showed that a Bayesian model is preferable compared to the frequentist approach for multiple linear regression using the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Mean Absolute Deviance (MAD) as criterion for comparison. The Bayesian approach has many attractive features over frequentist statistics. In particular, missing data, outlying and latent variables often pose no difficulties in Bayesian analyses. This makes it a better option for parameter estimation [11]. Conceptually, Bayesian approach is a more straightforward method for making inferences than other methods. The Bayesian perspective offers a number of advantages over the conventional frequentist perspective. [14] showed the Bayesian modelling process from model development, through development of an MCMC algorithm to estimate its parameters, through model evaluation, and through summarization and inference. He also opined that although parameter estimates obtained via the Bayesian approach are often very consistent with those that could be obtained via a classical approach, there are many cases in which a Bayesian approach and a classical approach will not coincide.

In this paper a gentle introduction to Bayesian analysis and ordinary least squares (OLS) regression analysis is provided. In Linear Regression modelling, the classical assumptions of the OLS sometimes do not hold. The study objective is to ascertain if datasets provided meet the classical OLS assumptions or not, and demonstrate the prominence of Bayesian methods where OLS assumptions are violated. In this empirical study, two datasets were used for demonstrations: performance indicators of Nigeria's GDP and factors that impact Wine Quality.

2. MATERIALS AND METHODS

2.1 Data

Two secondary datasets were used for the empirical study:

(i) Gross Domestic Product (GDP) data with indicators variables as independent variables. The GDP dataset was obtained from the National Bureau of Statistics (NBS) website.

(ii) Wine Quality dataset with factors impacting quality as independent variables was sourced from a research paper by [15].

2.2 Ordinary Least Squares Regression (OLS)

The OLS involves minimizing the error sum of squares with respect to the regression parameters, *β's.* The fitting of an OLS model, estimation of parameters and testing of hypothesis properties of the estimators are based on the following major assumptions: (i) the relationship between the study variable and explanatory variables is linear, at least approximately, (ii) the error term has zero mean, (iii) the error term has constant variance, (iv) the errors are uncorrelated, (v) the errors are normally distributed, [16].

The validity of these assumption is needed for the results to be meaningful. If these assumptions are violated, the result can be incorrect and may have serious consequences.

According to [1], a regression model is defined as:

$$\boldsymbol{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i} \quad (1)$$

Let each of the k-predictor variables $\boldsymbol{X_1, X_2, X_3, \cdots, X_k}$ have n-levels. The $\boldsymbol{X_{ij}}$ representing the $\boldsymbol{i}$th level of the $\boldsymbol{j}$th predictor variables. For instance, $\boldsymbol{X_{23}}$ represents the second level of the third predictor variable $\boldsymbol{X_3}$. Observations of the response variable for the n-levels include $\boldsymbol{y_1, y_2, y_3, \cdots, y_n}$. This gives us a new system of linear equations;

$$\boldsymbol{y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + \varepsilon_i}$$

$$\boldsymbol{y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + \varepsilon_i}$$

$$\boldsymbol{y_3 = \beta_0 + \beta_1 X_{31} + \beta_2 X_{32} + \cdots + \beta_k X_{3k} + \varepsilon_i}$$

$$\cdots$$

$$\boldsymbol{y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_k X_{nk} + \varepsilon_i} \quad (2)$$

The system of equations represented in matrix form is:

$$Y = X\beta + \varepsilon$$

Where $Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$ $\quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ 1 & x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nn} \end{bmatrix}$

$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_n \end{bmatrix}$ and $\varepsilon = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

The matrix X is known as the *design matrix*. It contains information about the levels of the predictor variables at which the observations are obtained. The vector $\boldsymbol{\beta}$ contains all the regression coefficients. To fit the regression model, $\boldsymbol{\beta}$ should be known. $\boldsymbol{\beta}$ is estimated using least square estimates. The following equation is used:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{3}$$

The estimated regression model (fitted model) is given as: $\hat{y} = X\hat{\beta}$.
The observations, $\boldsymbol{y_i}$, is different from the fitted values $\boldsymbol{\hat{y}_i}$ obtained from this model. The difference between these two values is the residual, $\boldsymbol{e_i}$. The vector of residuals is obtained as: $\boldsymbol{e = y - \hat{y}}$.
The fitted model is then written as;

$$\hat{\beta} = (X'X)^{-1}X'y \qquad \text{but} \qquad \hat{y} = X\hat{\beta}$$

$$\therefore \quad \hat{y} = X((X'X)^{-1}X'y) \tag{4}$$

$$\hat{y} = Hy \qquad \text{and} \qquad \text{where } H = X((X'X)^{-1}X')$$

H is called the HAT matrix, it transforms the vector of the observed response values, y_i , to the vector of fitted values, $\boldsymbol{\hat{y}_i}$.
The OLS estimation makes use of the normality distribution of error i.e. $\varepsilon \sim N(0, \sigma^2)$. Since the error is distributed normally, the variables $(Y|X, \beta, \sigma^2)$ are also distributed normally. The variables $(Y|X, \beta, \sigma^2) \sim \mathrm{N}(X\beta, \sigma^2)$ and probability density function of the variables are given below as;

$$p(Y|X, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2\sigma^2}(Y - X\beta)^T((Y - X\beta)\right] \tag{5}$$

Based on the probability density above, the likelihood function of the variables is defined as follows;

$$p(Y|X, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2\sigma^2}(Y - X\beta)^T((Y - X\beta)\right] \tag{6}$$

$$p(Y|X, \beta, \sigma^2) = (\sigma^2)^{-n/2} exp\left[-\frac{1}{2\sigma^2}(Y - X\beta)^T((Y - X\beta)\right] \tag{7}$$

$$p(Y|X,\beta,\sigma^2) \propto (\sigma^2)^{-\frac{v}{2}} exp\left[-\frac{vs^2}{2\sigma^2}\right] \times (\sigma^2)^{-n/2} exp\left[-\frac{1}{2\sigma^2}(Y-X\beta)^T((Y-X\beta)\right] \quad (8)$$

2.3 Bayesian Linear Regression

The Bayesian approach takes cognizance of Prior, Likelihood and Posterior Distributions. The regression coefficients β(s) are assumed to be random with a specified prior distribution. Bayesian estimation does not give a point estimate rather it produces a posterior distribution. Parameter estimation using this approach is done by the product of the prior distribution and the likelihood [17]. There are several prior distributions that can be used, one of them include the distribution of the prior conjugate. This estimation of regression parameters is usually done by an iteration process on the marginal posterior [18], [19]:

$$Posterior \propto Likelihood \times Prior \quad (9)$$

$$p(\beta,\sigma^2|Y,X) \propto p(Y|X,\beta,\sigma^2)p(\sigma^2)p(\beta|\sigma^2) \quad (10)$$

$$p(\beta,\sigma^2|Y,X) \propto (\sigma^2)^{-n/2} exp\left[-\frac{1}{2\sigma^2}(Y-X\beta)^T((Y-X\beta)\right] \times (\sigma^2)^{-(\frac{v}{2}+1)} exp\left[-\frac{vs^2}{2\sigma^2}\right] \times (\sigma^2)^{-k/2} exp\left[-\frac{1}{2\sigma^2}(\beta-\mu)^T\Lambda(\beta-\mu)\right] \quad (11)$$

Markov-Chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution [20], [21]. Gibbs sampling is a special case of the Metropolis-Hastings method and one of the techniques in MCMC where all the samples of parameters are drawn from or generated from distributions with a 100% acceptance rate [22], [23]. Thus, Bayesian estimation of parameters follows from the model [14]:

$$(y|\beta,\sigma^2) \sim N_n(X\beta,\sigma^2 I_n). \quad (12)$$

A conjugate prior distribution, the conditional distribution of β is given by

$$(\beta|\sigma^2) \sim N_p(\tilde{\beta},\sigma^2 M^{-1}) \quad (13)$$

where M is a (p, p) positive definite symmetric matrix, and the marginal prior on σ^2 is an inverse Gamma distribution

$$\sigma^2 \sim IG(a,b) \qquad a,b > 0$$

where *IG(a ,b)* refers to Inverse Gamma Distribution with Parameters a *and* b.

Taking advantage of the matrix identities

$$(M + X^TX)^{-1} = M^{-1} - M^{-1}(M + (X^TX)^{-1})^{-1}M^{-1}$$ (14)

$$= (X^TX)^{-1} - (X^TX)^{-1}(M^{-1} + (M^{-1} + (X^TX)^{-1})^{-1}(X^TX)^{-1} \quad (15)$$

and $$X^TX(M + X^TX)^{-1}M = (M^{-1}(M + X^TX)(X^TX)^{-1})^{-1} \quad (16)$$

$$= (M^{-1} + (X^TX)^{-1})^{-1} \quad (17)$$

This establishes that:

$$\beta|y, \sigma^2 \sim N_p((M + X^TX)^{-1}\{(X^TX)\hat{\beta} + M\tilde{\beta}\}, \sigma^2(M + X^TX)^{-1}) \quad (18)$$

Where $\hat{\beta} = (X^TX)^{-1}X^Ty$ and

$$\sigma^2|y \sim IG(\frac{n}{2} + a, b + \frac{s^2}{2} + \frac{(\tilde{\beta} - \hat{\beta})^T(M^{-1} + (X^TX)^{-1})^{-1}(\tilde{\beta} - \hat{\beta})}{2}) \quad (19)$$

Where $s^2 = (y - \hat{\beta}X)^T(y - \hat{\beta}X)$ are the correct Posterior distributions.

To get the $(1- \propto)$ highest posterior density (HPD) region on β,

From the prior distribution (equation 13):

$$\beta\big|\sigma^2, X \sim N_{k+1}(\tilde{\beta}, \sigma^2M^{-1}) \;,\; \sigma^2\big|X \sim IG(a, b) \quad (20)$$

The posterior distribution is

$$\pi(\beta, \sigma^2 | \hat{\beta}, s^2, X) \propto \sigma^{-k-1-2a-2-n} \exp\frac{-1}{2\sigma^2}\left\{(\beta - \tilde{\beta})^T M(\beta - \tilde{\beta}) + (\beta - \tilde{\beta})^T(X^TX)(\beta - \tilde{\beta}) + S^2 + 2b\right\} \quad (21)$$

$$= \sigma^{-k-n-2a-3} \exp\frac{-1}{2\sigma^2}\{\beta^T(M + X^TX)\beta - 2\beta^T(M\tilde{\beta} + X^TX\hat{\beta}) + \tilde{\beta}^TM\hat{\beta} + \hat{\beta}^T(X^TX)\hat{\beta} + S^2 + 2b\}$$

(22)

$$= \sigma^{-k-n-2a-3} \exp\frac{-1}{2\sigma^2}\{(\beta - \mathbb{E}[\beta|y, X])^T(M + X^TX)(\beta - \mathbb{E}[\beta|y, X])\} + \beta^TM\tilde{\beta} + \hat{\beta}^T(X^TX)\hat{\beta} - \mathbb{E}[\beta|y, X]^T(M + X^TX)\mathbb{E}[\beta|y, X] + S^2 + 2b \quad (23)$$

with $$\mathbb{E}[\beta|y, X] = (M + X^TX)^{-1}(M\tilde{\beta} + X^TX\hat{\beta}) \quad (24)$$

Therefore, integrating out β leads to.

$$\pi(\sigma^2|\hat{\beta}\,, S^2, X) \propto \sigma^{-n-2a-2} exp \frac{-1}{2\sigma^2}\left\{\beta^T M\tilde{\beta} + \hat{\beta}^T(X^TX)\hat{\beta} - \mathbb{E}[\beta|y,X]^T(M + X^TX)\mathbb{E}[\beta|y,X] + S^2 + 2b\right\}$$

$$= \sigma^{-n-2a-2} exp \frac{-1}{2\sigma^2}\left\{\beta^T M\hat{\beta} + \hat{\beta}^T(X^TX)\hat{\beta} + S^2 + 2b - \left(M\tilde{\beta} + X^TX\right)^T(M + X^TX)^{-1}(M\tilde{\beta} + X^TX)\right\} \quad (25)$$

using the matrix identity, gives:

$$\left(M\tilde{\beta} + X^TX\hat{\beta}\right)^T(M + X^TX)^{-1}(M\tilde{\beta} + X^TX\hat{\beta})$$

$$= \tilde{\beta}TM\tilde{\beta} - \tilde{\beta}^T(M^{-1} + (X^TX)^{-1})^{-1}\tilde{\beta} + \hat{\beta}^T(X^TX)\hat{\beta} - \hat{\beta}^T(M^{-1} + (X^TX)^{-1})\hat{\beta} + 2\hat{\beta}^T(X^TX)(M + X^TX)^{-1}M\tilde{\beta} \quad (26)$$

$$= \tilde{\beta}^T M\tilde{\beta} + \hat{\beta}^T(X^TX)\hat{\beta} - \left(\tilde{\beta} - \hat{\beta}\right)^T(M^{-1} + (X^TX)^{-1})^{-1}(\tilde{\beta} - \hat{\beta}) \quad (27)$$

by the virtue of the second identity. Therefore,

$$\pi(\sigma^2|\hat{\beta}\,, S^2, X) \propto \sigma^{-n-2a-2} exp \frac{-1}{2\sigma^2}\left\{\left(\tilde{\beta} - \hat{\beta}\right)^T(M^{-1} + (X^TX)^{-1})^{-1}(\tilde{\beta} - \hat{\beta}) + S^2 + 2b\right\}$$

$$\text{Thus} \quad (\beta|y, X) \sim \mathfrak{I}_{k+1}(n + 2a\,, \hat{\mu}, \hat{\textstyle\sum})$$

This means that

$$\pi(\beta|y, X) \propto \frac{1}{2}\left\{1 + \frac{(\beta - \hat{\mu})^T\hat{\sum}^{-1}(\beta - \hat{\mu})}{n+2a}\right\}^{n+2a+k+1} \quad (28)$$

And therefore, that an HPD region is of the form,

$$\mathfrak{H}_\propto = \left\{\beta, (\beta - \hat{\mu})^T\hat{\textstyle\sum}^{-1}(\beta - \hat{\mu}) \le K_\propto\right\} \quad (29)$$

Where $K_\propto$ is determined by the coverage probability α.

Now, $(\beta - \hat{\mu})^T\hat{\sum}^{-1}(\beta - \hat{\mu})$ has the same distribution as $||Z||^2$ when

$$Z \sim \mathcal{J}_{K+1}(n + 2a, 0, I_{k+1}) \quad (30)$$

This distribution is Fisher, $\digamma\ (k+1, n+2a)$ distribution which means that the bound $K_{\propto}$ is determined by the quantiles of this distribution. More details on Bayesian Linear Regression in [24].

3. RESULT

As stated in methodology section, two datasets were sourced for this empirical study, namely; Gross Domestic Product (GDP) dataset and Wine Quality dataset.

3.1 Results Using GDP Data

3.1.1 Summary Statistics of GDP Data

Table 1 shows the summary statistics of the Nigerian GDP data, taking the GDP (Y) as the response variable and Private Consumption (X_1), Gross Investment (X_2), Government Investment (X_3), Export (X_4), Import (X_5) respectively as independent variable.

Table 1: Summary Statistics of GDP Dataset

Summary Statistics	GDP (Y)	Private Consumption (X_1)	Gross Investment (X_2)	Govt. Invest. (X_3)	Exports (X_4)	Imports (X_5)
Minimum	69147	11351	54.09	4606	4000	8000
1st Quartile	769367	101246	55.63	26944	21500	45500
Median	4582127	331057	56.46	193413	375000	567000
Mean	8815976	918275	57.25	227561	1394419	3451161
3rd Quartile	17684947	1924130	58.38	413456	1678000	4123500
Maximum	29205783	4007832	61.57	677957	9893000	22444000

3.1.2 OLS and Bayesian Regression Modelling Using GDP Data

The multiple linear regression model based on the GDP dataset was fitted using OLS parameter estimation method. The results of the estimation process for the fitted model is presented in Table 2. Similarly, results of Bayesian parameter estimation requires the use of Normal distribution as the prior distribution for the β parameter and the Gamma Inverse distribution for the σ^2 parameter. The Markov Chain Monte Carlo (MCMC) employed the Gibbs Sampler. Table 2 also gives the results of the Bayesian estimation of regression parameters. The parameter estimates for both Bayesian and OLS methods are approximately the same. To determine a better method, standard errors of the two methods were considered.

As indicated in Table 2; the smaller standard errors of estimated parameters of the Bayesian regression model show that Bayesian approach is a better modelling method.

Table 2: Parameter Estimates for OLS and Bayesian Regression Modelling

	OLS		Bayesian	
	Estimates	**Std. Error**	**Estimates**	**Std. Error**
Intercept	7.247×10^{6}	2.757×10^{1}	7.245×10^{6}	2.744×10^{1}
Private Consumption (X_1)	0.496×10^{1}	3.075×10^{-2}	0.496×10^{-1}	3.052×10^{-2}
Gross Investment (X_2)	-1.133×10^{5}	2.100×10^{-1}	-1.130×10^{5}	2.091×10^{-1}
Govt. Investment (X_3)	1.050×10^{1}	8.414×10^{-3}	1.043×10^{1}	8.358×10^{-3}
Exports (X_4)	4.443×10^{-1}	6.441×10^{-1}	4.433×10^{-1}	6.479×10^{-1}
Imports (X_5)	-2.415×10^{-1}	4.040×10^{-4}	-2.438×10^{-1}	4.058×10^{-4}

Multiple R-squared: 0.8143 **Adjusted R-squared: 0.7771**

3.1.3 Assumption Checks for the OLS

The violation or otherwise of the OLS assumptions was established using different diagnostics plots as shown in Figure 1. The residual-fitted values plots show that the residuals have linear patterns, that is, a linear relationship between the residuals and the fitted values. The near-equal spread of values around the line of best fit (broken line) indicates the fitted model is approximately adequate. The Q-Q plot tests normality of the residuals. It is observed that some of the residuals do not follow the straight line, thus the residuals of the fitted model are not normally distributed.

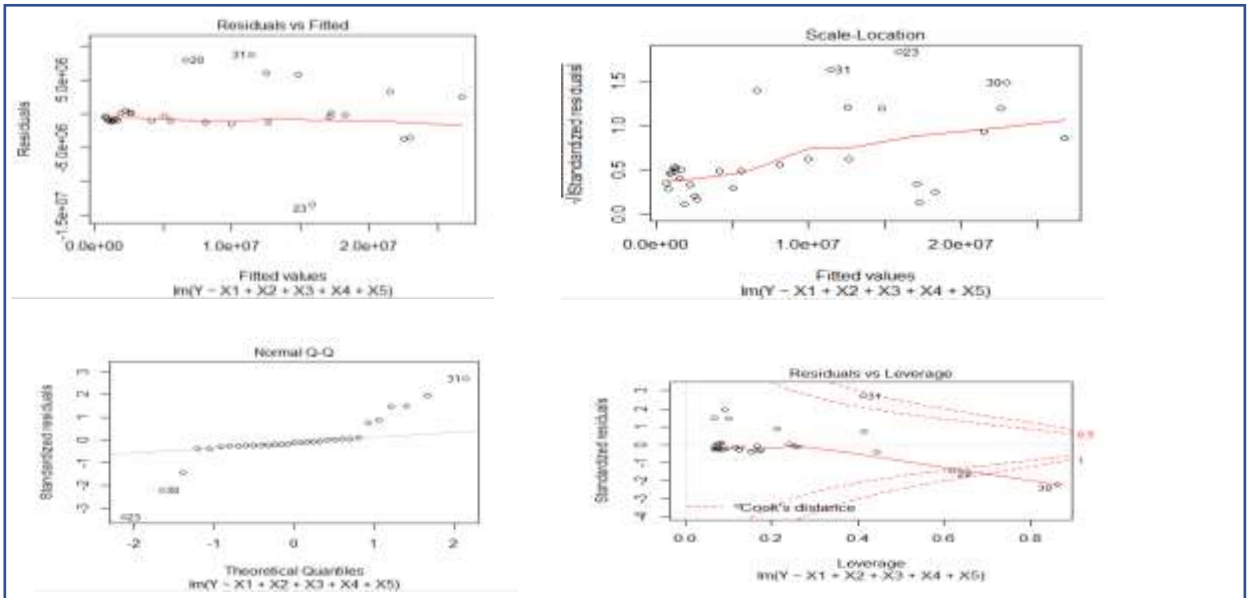

Figure 1: Diagnostic Checks for the Fitted Model based on GDP Data using Residual Plots

3.2 Results Using Wine Quality Data

3.2.1 Summary Statistics Using Wine Quality Data

Table 3 shows the summary statistics of the Nigerian GDP data taking the Wine Quality (Y) as the response variable and Fixed Acidity, Volatile Acidity, Residual Sugar, Chlorides, Total Sulphur Dioxide, Density, pH, Sulphates, Alcohol variables respectively as independent variable.

Table 3: Summary Statistics using Wine Quality Data

	Fixed Acidity X_1	Volatile Acidity X_2	Residual Sugar X_3	Chlorides X_4	Total Sulfur Dioxide X_5	Density X_6	Ph X_7	Sulphates X_8	Alcohol X_9	Quality Y
Minimum	3.8	0.08	0.6	0.01	9	0.99	2.72	0.22	8	3
1st Quartile	6.3	0.21	1.7	0.04	108	0.99	3.09	0.41	9.5	5
Median	6.8	0.26	5.2	0.04	134	0.99	3.18	0.47	10.4	6
Mean	6.86	0.28	6.39	0.05	138.4	0.99	3.19	0.49	10.51	5.88
3rd Quartile	7.3	0.32	9.9	0.05	167	1	3.28	0.55	11.4	6
Maximum	14.2	1.1	65.8	0.35	440	1.04	3.82	1.08	14.2	9

3.2.2 OLS and Bayesian Regression Modelling Using Wine Quality Data

The multiple linear regression model based on the Wine Quality dataset was fitted using OLS parameter estimation method. The results of the estimation process for the fitted model is presented in Table 4. Similarly, results of Bayesian parameter estimation requires the use of Normal distribution as the prior distribution for the β parameter and the Gamma Inverse distribution for the σ^2 parameter. The Markov Chain Monte Carlo (MCMC) employed the Gibbs Sampler. Table 4 gives the results of the Bayesian estimation of regression parameters with standard errors.

Table 4: Parameter Estimates for OLS and Bayesian Regression Modelling

	OLS		Bayesian	
	Estimates	**Std. Error**	**Estimates**	**Std. Error**
(Intercept)	1.628×10^{2}	1.857×10^{1}	1.627×10^{2}	1.844×10^{1}
Fixed Acidity (X_1)	6.646×10^{-2}	2.075×10^{-2}	6.637×10^{-2}	2.052×10^{-2}
Volatile Acidity (X_2)	-1.966	1.100×10^{-1}	-1.967	1.091×10^{-1}
Residual Sugar (X_3)	8.731×10^{-2}	7.414×10^{-3}	8.731×10^{-2}	7.358×10^{-3}
Chlorides (X_4)	-1.533×10^{-1}	5.441×10^{-1}	-1.531×10^{-1}	5.479×10^{-1}
Total SulfurDioxide (X_5)	7.152×10^{-4}	3.040×10^{-4}	7.147×10^{-4}	3.058×10^{-4}
Density (X_6)	-1.629×10^{2}	1.884×10^{1}	-1.628×10^{2}	1.871×10^{1}

Ph(X_7)	7.073×10^{-1}	1.051×10^{-1}	7.067×10^{-1}	1.042×10^{-1}
Sulphates (X_8)	6.377×10^{-1}	1.005×10^{-1}	6.379×10^{-1}	1.008×10^{-1}
Alcohol (X_9)	1.837×10^{-1}	2.405×10^{-2}	1.839×10^{-1}	2.390×10^{-2}

Multiple R-squared: 0.279 **Adjusted R-squared: 0.2776**

As indicated in Table 4; the smaller standard errors of estimated parameters of the Bayesian regression model show that Bayesian approach is a better modelling method.

3.2.3 Assumption Checks for the OLS

The violation or otherwise of the OLS assumptions was established using different diagnostics plots as shown in Figure 2. Multiple R-squared and Adjusted R-squared values indicate there is no good linear relationship between the variables and the model is not well-fitted. A violation of linearity between the response variable and independent variables. The Residual-Fitted Values plot shows that residuals have linear patterns, that is a linear relationship between the residuals and the fitted values. The near-equal spread of values around the line of best fit (broken line) indicates a linear relationship. Based on this plot, the model is approximately adequate, hence the model is fairly good. The Q-Q plot tests normality of the residuals. It is observed that some of the residuals do not follow the straight line, thus the residuals of the fitted model are not normally distributed.

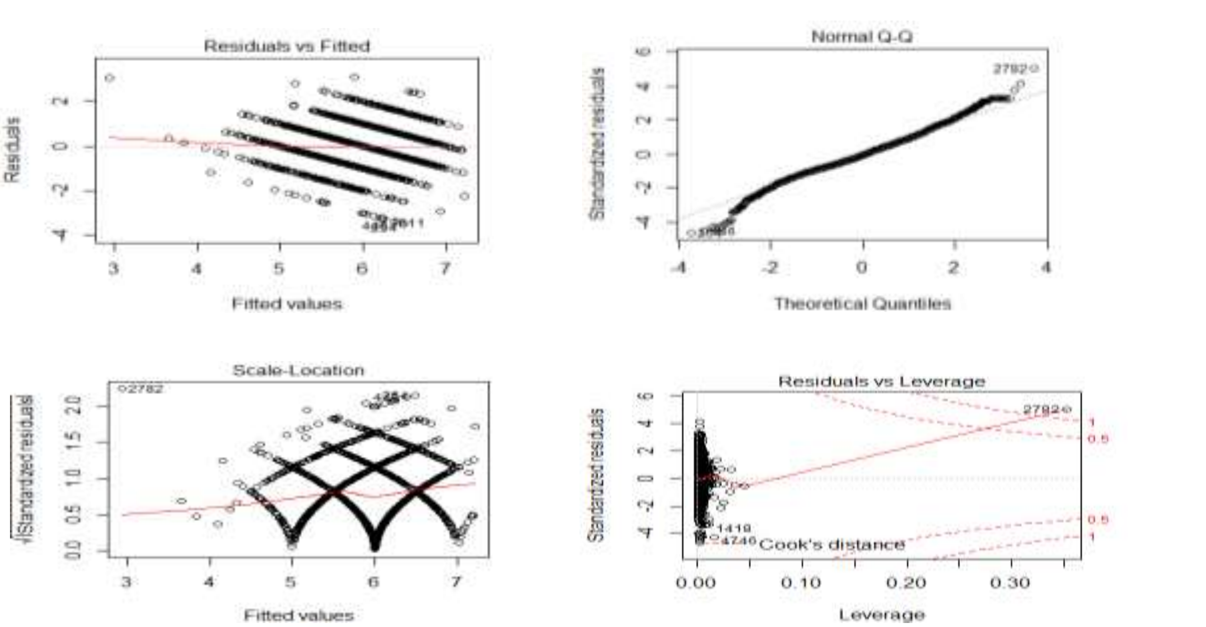

Figure 2: Diagnostic Checks of Fitted Model based on Wine Quality Data using Residual Plots

4. DISCUSSION

Based on the results of analysis of the two datasets namely; Gross Domestic Product (GDP) dataset and Wine Quality dataset, discussions are respectively presented below:

4.1 Gross Domestic Product (GDP)

Thus, the assumption of ordinary least square is violated because not all the residuals are lined on the straight dashed line of the Q-Q plot. The Spread-Location plot checks the assumption of equal variance (homoscedasticity). In Figure 1, it can be seen that the spread of residual points is not equal above and beneath the line, the data is not homoscedastic. This is another assumption violation of OLS regression modelling. In addition to the smaller standard errors of estimated parameters, since the assumptions of the linear regression using the OLS were not met, the conclusion is that the Bayesian method is better than the OLS method for this dataset.

4.2 Wine Quality

Similarly, the assumption of OLS is violated because not all the residuals are lined on the straight dashed line of the Q-Q plot. The Spread-Location plot checks the assumption of equal variance (homoscedasticity). In Figure 2, it can be seen that the spread of residual points is not equal above and beneath the line, the data is not homoscedastic. This is another assumption violation of OLS regression modelling. On the Wine Quality dataset, violation of assumptions of linear regression analysis using the OLS cannot be overlooked. It is noted that standard error of estimated parameters for the Bayesian fitted model are smaller than the standard error of estimated parameters for the OLS fitted model. Therefore, Bayesian linear regression analysis should be preferred.

5. CONCLUSION

Beyond accuracy in forecasting and prediction from regression models, underlying assumptions for analysis must be valid and not violated. In the fulfilment of these assertions, consideration of Bayesian approach becomes imperative as a robust regression modelling method when compared to classical OLS method. Thus, this work successfully showed that the Bayesian approach gives a more robust, accurate and reliable results in the areas of regression analysis than the frequentist approach. Hence, this is an improvement on the work of [25] which stated that the results of empirical Bayesian method were largely consistent with OLS results. Furthermore, Bayesian approach thrives on iterative steps involving prior distribution to give better regression model analysis.

REFERENCES

[1] J. Cohen, P. Cohen, S. G. West, & L. S.Aiken (2013). *Applied multiple regression/correlation analysis for the behavioral sciences.* Routeledge.

[2] C. G. Soner, T. K. Levent, T. Yalcin, & A. Mustapha (2006). A Comparative Study of Estimation Methods for Parameters in Multiple Linear regression Model. *Journal of Applied Animal Research,* **29**(1), pp. 43-47.

[3] V. E. Johnson, & H. E. Albert (1999). *Ordinal Data Modeling.* New York: Springer-Verlag.

[4] R. K, Ogundeji & J. N. Onyeka-Ubaka (2019). A Bayesian Approach for Validation of Election Results. *Nigerian Journal of Basic and Applied Science,* 27(2): 22-31.

[5] A. D. Martin & K. M. Quinn (2002). Dynamic Ideal Point Estimation via Markov chain Monte Carlo for the U.S. Supreme Court. *Political Analysis, 1953-1999,* pp. 134–153.

[6] A., O' Thomas, B. Hara, U. Ligges & S. Sturtz (2006). Making BUGS Open. *R News*, 12-17. Retrieved from http://CRAN.R-project.org/doc/Rnews/

[7] M. Plummer (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using. In K. Hornik, F. Leisch, & A. Zeileis (Ed.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Austria.* Vienna. Retrieved from http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/

[8] M. Plummer (2005). JAGS: Just Another Gibbs Sampler, Version 0.8. Retrieved from http://www-fis.iarc.fr/~martyn/software/jags/

[9] L. K. Muthen & B. O. Muthen (2008). *Mplus (version 5.1).* Los Angeles, CA: Muthen and Muthen.

[10] D. J. Spiegelhalter, A. Thomas, N. G. Best & D. Lunn (2003). WinBUGS Version 1.4 User Manual. Cambridge: MRC Biostatistics Unit. Retrieved from http://www.mrc-bsu.cam.ac.uk/bugs/

[11] T. W. Elam, B. Scruggs, F. Eggert & J. A. Nicolosi (2011). Advantages and Disadvantages of Bayesian Methods for Obtaining Xrf Net Intensities. *International Centre for Diffraction*, **25**(2), pp. 1097-2002.

[12] V. Bonfatti, F. Tiezzi, F. Miglior, & P. Carnier (2017). Comparison of Bayesian Regression Models and Partial Least Squares Regression for the Development of Infrared Prediction Equations. *American Dairy Science Association,* **100**(9), pp. 7306-7319.

[13] D. P. Syarifah & T. Heruna (2018). Linear Regression Model Using Bayesian Approach for Energy Performance of Residential Building. *Procedia Computer Science,* **135,** pp. 671-677.

[14] S. M. Lynch (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* Springer.

[15] P. Cortez, A. Cerdeira, F. Almedia, T. Matos & J. Reis (2009). Modeling Wine Preferences By Data Mining from Physicochemical Properties. *In Decision Support Systems,* **47**(4), pp. 547-553.

[16] E. A. Pena & E. H. Slate (2006). Global Validation of Linear Model Assumptions. *Jounal of America Statistics Association,* **101**(473), pp. 341-354.

[17] R. McElreath (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* Boca Raton FL: Chapman and Hall/CRC Press.

[18] R. K. Ogundeji, I. A. Adeleke & R. O. Okafor (2018) An Assessment of Predictive Performance of Zellner's g-Priors in Bayesian Model Averaging. *Model Assisted Statistics and Applications (IOS Press)*, 13(2018), 63–71.

[19] M. Chelsea, O. Zita & G. Jonah (2018). User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *The Quantitative Methods for Psychology*, **14**(2), pp. 99 - 119.

[20] W. R. Gilks, S. Richardson & D. J. Spiegelhalter (1996). *Markov Chain Monte Carlo In Practice.* Boca Raton: Chapman & Hall/CRC.

[21] W. K. Ching & K. N. Michael (2006). *Markov Chains: Models, Algorithms and Applications.* Springer.

[22] G. Casella & E. George (1992). Explaining the Gibbs sampler. *American Statistician*, **46**(3), pp. 167-174.

[23] A. Gelman, J. Carlin, H. Stern & D. Rubin (2003). *Bayesian Data Analysis.* Boca Raton: Chapman and Hall/CRC.

[24] J. Marin and C. P. Robert (2014). Time Series Analysis with Applications in R. Springer Texts in Statistics, Springer Science Business Media, LLC.

[25] R. Abdul & N. Nazia (2017). Effects of Mergers on Corporate Performance: An Empirical Evaluation Using OLS and the Empirical Bayesian Methods. *Borsa _Istanbul Review*, **17**(1), pp. 10-24.

OGUNDEJI, ROTIMI[1*]

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, UNIVERSITY OF LAGOS, AKOKA, NIGERIA.

Email: rogundeji@unilag.edu.ng;

ONYEKA-UBAKA, JOSEPHINE

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, UNIVERSITY OF LAGOS, AKOKA, NIGERIA.

Email: jonyeka-ubaka@unilag.edu.ng;

YINUSA, EMMANUEL

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE, UNIVERSITY OF LAGOS, AKOKA, NIGERIA.

Email: yinusaemmanuel@gmail.com