BAYESIAN ESTIMATION OF OPTIMAL HYPERPARAMETERS OF LINEAR REGRESSION MODELS USING QUANTILE RANGES

AYOADE I. ADEWOLE[1] and OLUSOGA A. FASORANBAKU[2]

[1]Department of Mathematics,
Tai Solarin University of Education Ijagun Ogun State Nigeria.
[2]Department of Statistics, Federal University of Technology Akure

Author Email Address: oafasoranbaku@yahoo.com,
P.M.B 704, Akure, Ondo State, Nigeria,
Tel: +2348033932972
Corresponding Author Email Address: hayorhade2005@gmail.com,
P.M.B 2118, Ijebu-Ode
Ogun State, Tel: +2348055124368

ABSTRACT

Bayesian estimations have the advantages of taking into account the uncertainty of all parameter estimates which allows virtually the use of vague priors. Most importantly if the data are sparse, there is a need to specify prior distributions for all unknown parameters in analyzing the data from a Bayesian view. This study focused on determining the quantile range at which optimal hyperparameter of normally distributed data with vague information could be obtained in Bayesian estimation of linear regression models. Normally distributed data of 200 sample sizes were generated through Monte Carlo Simulation approach. Confidence Intervals for the regression parameters were obtained from the Ordinary Least Squares analysis. The variances were divided into 10 equal parts to obtain the hyperparameters of the prior distribution. Observation Precisions, Posterior Precisions were estimated from the regression output to determine the posterior means estimate for each model to derive the new dependent variables. Average Absolute Deviation was employed for model selection are used to validate the adequacy of each model. The process was repeated 10000 times to determine the quantile range for the optimal hyperparameters. The study established that the optimal hyperparameters are located within 5^{th} and 7^{th} deciles. The research simplified the process of selecting the hyperparameters of prior distribution from the data with vague information in empirical Bayesian inferences.

KEYWORD: Optimal Hyperparameters, Quantile Ranges, Bayesian Estimation and Vague Prior.

INTRODUCTION

Linear Regression represents the dependent variable as a linear function of one or more independent variables, subject to a random “disturbance ‘’ or error term. It estimates the mean value of the dependent variables for given level of independent variables, the relationship is modeled using linear predictor functions whose unknown model parameters are estimated from the data. (Hilary 1967, Schaalje and Rencher 2007). Linear regression focuses on the conditional probability distribution of Y given X, rather than on the joint probability distribution of Y and X, which is in the domain of multivariate analysis, (Yan 2009, Rencher and William 2012).

In statistics, Bayesian linear regression is an approach to linear regression in which the statistical analysis is undertaken within the context of Bayesian inference. The Bayesian approach provides a complete paradigm for both statistical inference and decision making under certainty. Bayesian methods make it possible to incorporate scientific hypothesis in the analysis (by means of prior distribution) and may be applied to problems whose structure is too complex for conventional methods to handle. While the objectivity of frequentist statistics has been obtained by disregarding any prior knowledge of the process being measured, the Bayesian approach allows direct probability statements about the parameters which are much more useful than the confidence statements. The essence of Bayesian approach is to provide a mathematical rule explaining how existing beliefs can be changed in the light of new evidence. In other words, it allows scientist to combine new data with their existing knowledge or prior. Researchers considered Bayesian approach to be superior to frequentist approach through the application of prior information. The argument is that the introduction of prior distributions violates the objective view point of convectional statistics, Lunn *et al.* (2013)..

This study is to investigate the claim of Atkinson *et al.* (1993) that prior distribution could be suggested by data to reduce the uncertainty around the determination of Bayesian prior and to use the prior to determine the superiority of Bayesian regression analysis over frequentist regression analysis. Raftery *et al.* (1997) considered the problems of accounting for model uncertainty in linear regression model. Conditioning on a single selected model ignores model uncertainty, and thus leads to underestimation of uncertainty when making inferences about quantities of interest. A Bayesian solution to this problem involves averaging over all possible models when making inferences about quantities of interest.

Aside from investigating Atkinson *et al.* (1993), this research work also aimed to determine the quantile range at which optimal hyperparameters could be obtained when Bayesian estimation is employed to solve regression analysis of normally distributed data with vague information. Agresti (2006) examined Bayesian Inference for categorical Data Analysis, with primary emphasis on contingency table analysis. Several application of Bayesian analyses have yielded evidence that some hyperparameters indeed are much important. Through minimizing an empirical error criterion, Adankon and Cheryl (2009) used a gradient descent method to automatically select hyperparameter values for the least squares support vector machine. (Bergstra and Bengio, 2012, Hutter *et al*. 2013) also worked on speeding up automatic selection of hyperparameter value for neural work. Liseo andMacaro (2013) considered the problem of deriving objective priors for the causal stationary autoregressive model of orderp.Guido *et al.* (2018) provided review of prior distributions for objective Bayesian analysis. However, not much work has been done on determining the range of obtaining the optimal hyperparameter in Bayesian estimation of linear regression models. With this gap in mind, this study opts to determine the optimal quantile range of the prior parameters and establish the prior parameters from ordinary least squares (OLS) model confidence intervals.

MATERIALS AND METHODS

The simple linear regression model is given by

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad (1)$$

for i= 1,2,…,n, y_i is the response variable, x_i is the explanatory variable, α *and* β are the unknown parameters that are to be estimated and ε_i is the residual term. The term ε_i is independent and identically distributed with a normal distribution with mean 0 and unknown, σ^2.

Baye's theorem is constantly summarized by

Posterior ∝ prior × likelihood

Hence, there is a need to determine the prior and the likelihood distribution for the model.

The joint likelihood is factored into a product of two individual likelihood of α *and* β. It is simplified as

$$Likelihood_{sample}(\alpha_x, \beta) \propto likelihood_{sample}(\alpha_x) \text{ X } likelihood_{sample}(\beta) \qquad (2)$$

where;

$$likelihood_{sample}(\beta) \propto e^{-} \frac{1}{2\sigma^2 / ss_x} (\beta - B)^2 \qquad (3)$$

and

$$likelihood_{sample}(\alpha_x) \propto e^{-} \frac{1}{2\sigma^2 / n} (\alpha_x - A_x)^2 \qquad (4)$$

The independent likelihood is independent, the likelihood of the slope β has a normal shape with mean B and the variance $\frac{\sigma^2}{ss_x}$. Similarly, the likelihood of α_x also has normal shape with mean A_x and variance$\frac{\sigma^2}{n}$.

If the joint likelihood is multiplied by joint prior, it is proportional to the joint posterior. Using independent priors for each parameter, the joint prior of the two parameters is the product of the two individual priors.

$$g(\alpha_x, \beta) = g(\alpha_x) \text{ X } g(\beta) \qquad (5)$$

The joint prior follows a normal distribution.

The joint posterior is proportional to the joint prior multiplied by the joint likelihood.

$$g(\alpha_x, \beta | data) \propto g(\alpha_x, \beta) \text{ X } likelihood_{sample}(\alpha_x, \beta) \qquad (6)$$

where the data is the set of ordered pair (x_i, y_i), …, (x_n, y_n)

Regression analysis was run for a data set, the standard errors obtained from the linear regression result for the output from the parameters $\beta_0, .., \beta_4$ were used to obtain the lower and upper limits for the prior variance required for thestudy analysis using chi- square with $(n-1)$ degree of freedom.

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2, n-1}} \quad \text{and} \quad \frac{(n-1)s^2}{\chi^2_{1-\alpha/2, n-1}} \quad \text{respectively} \qquad (7)$$

The confidence interval for each parameter is used to determine the prior mean range

$$\overline{\beta_s} \pm t_{\alpha/2,\, n-2} \text{ X } SE \qquad (8)$$

Difference between lower and upper intervals obtained above is divided into 9 grid points. Each pair of the points is the mean and variance for the prior distribution, these grid points are referred to as the hyperparameters of the prior distribution.

$$\beta \sim N(m_\beta, s^2_\beta) \qquad (9)$$

The likelihood function is of the form:

$$\overline{X}/\beta \sim N(\beta, \frac{\sigma^2}{n}) \qquad (10)$$

where $\bar{X}$ is the average of X_i ,…, X_n

The likelihood of the regression parameters $\beta_0, .., \beta_4$ were estimated, for the intercept (β_0), the mean square error from the regression analysis and the sample size n were used while the sum of squares of the independent variables $(X_s), s = 1, \dots, 4$ and the mean square errorobtainedwere used to determine the likelihood of β_1 ,β_2,β_3,and β_4.

$$likelihood\ of\ \ \beta_0 = \frac{MSE}{n} \tag{11}$$

$$likelihood\ of\ \ \beta_1 = \frac{MSE}{SSx_1} \tag{12}$$

$$likelihood\ of\ \ \beta_2 = \frac{MSE}{SSx_2} \tag{13}$$

$$likelihood\ of\ \ \beta_3 = \frac{MSE}{SSx_3} \tag{14}$$

$$likelihood\ of\ \ \beta_4 = \frac{MSE}{SSx_4} \tag{15}$$

The posterior distribution:

$$\beta \big| X \sim N(m'_{\beta}, (S'_{\beta})^2) \tag{16}$$

The posterior precisions of the regression parameters $\beta_0, .., \beta_4$are the prior precisions plus the observation precision for the parameters.

$$\text{for } \beta_0, \frac{1}{(S'\beta_0)^2} \quad = \quad \frac{1}{S^2\beta_0} + \frac{n}{\sigma^2} \tag{17}$$

$$\text{for } \beta_1, \frac{1}{(S'\beta_1)^2} \quad = \quad \frac{1}{S^2\beta_1} + \frac{ssx_1}{\sigma^2} \tag{18}$$

$$\text{for } \beta_2, \frac{1}{(S'\beta_2)^2} = \frac{1}{S^2\beta_2} + \frac{ssx_2}{\sigma^2} \qquad (19)$$

$$\text{for } \beta_3, \frac{1}{(S'\beta_3)^2} = \frac{1}{S^2\beta_3} + \frac{ssx_3}{\sigma^2} \qquad (20)$$

$$\text{for } \beta_4, \frac{1}{(S'\beta_4)^2} = \frac{1}{S^2\beta_4} + \frac{ssx_4}{\sigma^2} \qquad (21)$$

The Bayes estimates of $\widehat{\beta_0}, \ldots, \widehat{\beta_4}$ were obtained using the equations below:

$$\frac{prior\ \ precision}{posterior\ \ precision} \text{ X } prior\, mean + \frac{observation\, precision}{posterior\, precision}\ OLS\ \ Estimate \qquad (22)$$

$$\text{for } \beta_0, \frac{\frac{1}{s^2\beta_0}}{\frac{1}{s^2\beta_0} + \frac{n}{\sigma^2}} \text{ X } PM\beta_0 + \frac{\frac{n}{\sigma^2}}{\frac{1}{s^2\beta_0} + \frac{n}{\sigma^2}} \text{ X } \beta_0\ OLS\ \ Estimate \qquad (23)$$

$$\text{for } \beta_1, \frac{\frac{1}{s^2\beta_1}}{\frac{1}{s^2\beta_1} + \frac{ssx_1}{\sigma^2}} \text{ X } PM\beta_1 + \frac{\frac{ssx_1}{\sigma^2}}{\frac{1}{s^2\beta_1} + \frac{ssx_1}{\sigma^2}} \text{ X } \beta_1\ OLS\ \ Estimate \qquad (24)$$

$$\text{for } \beta_2, \frac{\frac{1}{s^2\beta_2}}{\frac{1}{s^2\beta_2} + \frac{ssx_2}{\sigma^2}} \text{ X } PM\beta_2 + \frac{\frac{ssx_2}{\sigma^2}}{\frac{1}{s^2\beta_2} + \frac{ssx_2}{\sigma^2}} \text{ X } \beta_2\ OLS\ \ Estimate$$

(25)

$$\text{for } \beta_3, \frac{\frac{1}{s^2\beta_3}}{\frac{1}{s^2\beta_3}+\frac{ssx_3}{\sigma^2}} \text{ X } PM\beta_3 + \frac{\frac{ssx_3}{\sigma^2}}{\frac{1}{s^2\beta_3}+\frac{ssx_3}{\sigma^2}} \text{ X } \beta_3 \ OLS\ Estimate$$

(26)

$$\text{For } \beta_4, \frac{\frac{1}{s^2\beta_4}}{\frac{1}{s^2\beta_4}+\frac{ssx_4}{\sigma^2}} \text{ X } PM\beta_4 + \frac{\frac{ssx_4}{\sigma^2}}{\frac{1}{s^2\beta_4}+\frac{ssx_4}{\sigma^2}} \text{ X } \beta_4 \ OLS\ Estimate \quad (27)$$

SIMULATION STUDY

The simulation study applied a Monte Carlo simulation approach that generated the data set used for the extensive analysis to determine optimal hyperparameters and the optimal quantile range.

Let y_i and x_i denote the simulated data of the dependent and k explanatory variable, $x_{i1}, \dots, x_{ik}$,

for i = 1,…,n.

the linear regression model is given by:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \dots, + \beta_k x_{ik} + \varepsilon_t \quad (28)$$

where k = 4 and 10000 sets of random data of sample size 100 were simulated for the four independent variables x_1 to $x_4$with $y$generated from a linear regression model with normally distributed error term. The parameters used for the simulation were chosen arbitrarily, they were given below as

$$\beta_0 = 0.7, \beta_1 = 0.4, \beta_2 = 0.5, \beta_3 = 0.1, \beta_4 = 0.2$$

$x_1 \sim N(4, 0.3)$, $x_2 \sim N(2, 0.1)$, $x_3 \sim N(5, 3)$, and $x_4 \sim N(3, 1)$ and $\varepsilon \sim N(0, 1)$,

(29)

Regression analysis was run on each set of the simulated data, intervals for the prior mean and prior variance of the parameters determined were divided into 9 grid points, each pair of the grid points is therefore used to determine the posterior Bayesian estimates for the determinationof the optimal hyperparameter. The average absolute deviation for the posterior estimates for each of 9 grid points of the models was computed and the model with the least Average Absolute Deviations (AAD) ischosen as the best model with the corresponding hyperparameters as the optimal hyperparameters. The process was repeated for the remaining 9999 data set and the optimal hyperparameters for the 10000 data set were considered to determine the optimal quantile range. The Average Absolute Deviation was computed as

$$\frac{1}{N}\sum_{i,j,k=1}^{N,S,T}\left|Y_{i,j,k}-\hat{Y}_{i,j,k}\right| \tag{30}$$

where N is the sample size.

S is the number of grid points.

T is the number of simulations

Y_i is the simulated dependent variables.

$\hat{Y}_i$ is the estimated dependent variables derived from posterior estimates.

RESULTS AND DISCUSSION

LINEAR REGRESSION ANALYSIS

The ordinary least squares parameters' estimates obtained from the linear regression analysis of each set of simulated data was presented in Table 1 below

TABLE 1: REGRESSION STATISTICS

Multiple R			0.6523		
R- Square			0.8836		
Adjusted R Square			0.7106		
Standard Error			0.8529		
Observations			200		
	Df	SS	MS	F	Significance F
Regression	4	50.3721	38.3386	38.1944	3.3372E-11
Residual	195	93.3560	0.9032		
Total	199	144.2281			

	Coefficients	Standard Error	t statistics	P- value	Lower 95%	Upper 95%
Intercept	0.4814	0.03672	1.2638	0.1946	0.4089	0.5538
x_1	0.2409	0.2778	0.6582	0.3671	-0.3069	0.7887
x_2	0.5275	0.3415	3.0945	0.9923	-0.1459	1.20094
x_3	0.7423	0.1892	0.4789	0.2514	0.3692	1.1154
x_4	0.3364	0.4096	4.1139	0.8356	-0.4713	1.14413

The standard error of the regression parameters $\beta_0, .., \beta_4$ from the regression analysis results of the simulated data are 0.03672, 0.2778, 0.3415, 0.1892, and 0.4096 respectively. The chi square values for $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ were obtained from the statistical table as 161.826 and 239.960 respectively for sample size 200. These values were used to obtain the lower and upper limits of the prior variances for the regression parameters$\beta_0, .., \beta_4$.

Precision measures statistical variability, D_0 were obtained as the differences between the limits divided by 10 for the parameters $\beta_0, .., \beta_4$. The D_0 are the incremental values added to the lower limit of each of the prior variance to obtain D_i and where i = 1 to 9. The prior precision obtained is the reciprocal of the prior variance. Table 2 to 5 shows the derivation of priorvariances, prior precisions for the parameter estimates for the 9 grid points.

TABLE 2: THE PRIOR VARIANCE AND PRECISION FOR THE PARAMETER ESTIMATE $\boldsymbol{\beta_0}$

		$\frac{(n-1)s^2}{\chi^2_{\alpha/2,,n-1}}$	$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,,n-1}}$	D_0	D_i	Prior Variance of β_0	Prior Precision of β_0
N	200	0.00112	0.00166	0.000054	$D_1 = LL + D_0$	0.001174	851.7887
S	0.03672				$D_2 = D_1 + D_0$	0.001228	814.3322
$\chi^2{}_{\alpha/2}$	161.826				$D_3 = D_2 + D_0$	0.001282	780.0312
$\chi^2{}_{1-\alpha/2}$	239.960				$D_4 = D_3 + D_0$	0.001336	748.5029
					$D_5 = D_4 + D_0$	0.00139	719.4244
					$D_6 = D_5 + D_0$	0.001444	692.5207
					$D_7 = D_6 + D_0$	0.001498	667.5567
					$D_8 = D_7 + D_0$	0.001552	644.3298
					$D_9 = D_8 + D_0$	0.001606	622.6650

TABLE 3: THE PRIOR VARIANCE AND PRECISION FOR THE PARAMETER ESTIMATE $\boldsymbol{\beta_1}$

		$\frac{(n-1)s^2}{\chi^2_{\alpha/2,,n-1}}$	$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,,n-1}}$	D_0	D_i	Prior Variance of β_1	Prior Precision of β_1
N	200	0.06399	0.09490	0.003090	$D_1 = LL + D_0$	0.06708	14.9075
S	0.2778				$D_2 = D_1 + D_0$	0.07017	14.2511
${\chi^2}_{\alpha/2}$	161.826				$D_3 = D_2 + D_0$	0.07326	13.6500
${\chi^2}_{1-\alpha/2}$	239.960				$D_4 = D_3 + D_0$	0.07635	13.0975
					$D_5 = D_4 + D_0$	0.07944	12.5881
					$D_6 = D_5 + D_0$	0.08253	12.1168
					$D_7 = D_6 + D_0$	0.08562	11.6795
					$D_8 = D_7 + D_0$	0.08871	11.2726
					$D_9 = D_8 + D_0$	0.09180	10.8932

TABLE 4: THE PRIOR VARIANCE AND PRECISION FOR THE PARAMETER ESTIMATE $\boldsymbol{\beta_2}$

		$\frac{(n-1)s^2}{\chi^2_{\alpha/2,,n-1}}$	$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,,n-1}}$	D_0	D_i	Prior Variance of β_2	Prior Precision of β_2
N	200	0.0967	0.14341	0.00497	$D_1 = LL + D_0$	0.1016	9.8425
S	0.3415				$D_2 = D_1 + D_0$	0.1066	9.3809
$\chi^2{}_{\alpha/2}$	161.826				$D_3 = D_2 + D_0$	0.1116	8.9606
$\chi^2{}_{1-\alpha/2}$	239.960				$D_4 = D_3 + D_0$	0.1165	8.5837
					$D_5 = D_4 + D_0$	0.1215	8.2305
					$D_6 = D_5 + D_0$	0.1265	7.9051
					$D_7 = D_6 + D_0$	0.1315	7.6045
					$D_8 = D_7 + D_0$	0.1365	7.3260
					$D_9 = D_8 + D_0$	0.1414	7.0721

TABLE 5: THE PRIOR VARIANCE AND PRECISION FOR THE PARAMETER ESTIMATE $\boldsymbol{\beta_3}$

		$\frac{(n-1)s^2}{\chi^2_{\alpha/2,,n-1}}$	$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,,n-1}}$	D_0	D_i	Prior Variance of β_3	Prior Precision of β_3
N	200	0.02968	0.04401	0.001433	$D_1 = LL + D_0$	0.03111	32.1440
S	0.1892				$D_2 = D_1 + D_0$	0.03254	30.7314
$\chi^2{}_{\alpha/2}$	161.826				$D_3 = D_2 + D_0$	0.03397	29.4377
$\chi^2{}_{1-\alpha/2}$	239.960				$D_4 = D_3 + D_0$	0.03541	28.2406
					$D_5 = D_4 + D_0$	0.03685	27.1370
					$D_6 = D_5 + D_0$	0.03827	26.1301
					$D_7 = D_6 + D_0$	0.03971	25.1825
					$D_8 = D_7 + D_0$	0.04114	24.3072
					$D_9 = D_8 + D_0$	0.04257	23.9072

TABLE 6: THE PRIOR VARIANCE AND PRECISION FOR THE PARAMETER ESTIMATE $\boldsymbol{\beta_4}$

		$\frac{(n-1)s^2}{\chi^2_{\alpha/2,,n-1}}$	$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,,n-1}}$	D_0	D_i	Prior Variance of β_4	Prior Precision of β_4
N	200	0.13913	0.20631	0.00672	$D_1 = LL + D_0$	0.1458	6.8587
S	0.4096				$D_2 = D_1 + D_0$	0.1525	6.5573
$\chi^2{}_{\alpha/2}$	161.826				$D_3 = D_2 + D_0$	0.1593	6.2775
$\chi^2{}_{1-\alpha/2}$	239.960				$D_4 = D_3 + D_0$	0.1660	6.0240
					$D_5 = D_4 + D_0$	0.1727	5.7904
					$D_6 = D_5 + D_0$	0.1794	5.5741
					$D_7 = D_6 + D_0$	0.1862	5.3705
					$D_8 = D_7 + D_0$	0.1929	5.1840
					$D_9 = D_8 + D_0$	0.1996	5.0100

.

The sum of the squares of the deviation from the mean of the independent variables SSX_1, SSX_2, SSX_3 and SSX_4 for the simulated data were calculated as 32.19, 24.09, 414. 61 and 291.52 respectively. The observation precisions for $\beta_{0,}\beta_{1,}\beta_{2,}\beta_{3,}$ *and* β_4 were given by $\frac{n}{\sigma^2} = 221.4348$, $\frac{SSX_1}{\sigma^2} = \frac{32.19}{0.9032} = 35.6399$, $\frac{SSX_2}{\sigma^2} = \frac{24.09}{0.9032} = 26.6718$, $\frac{SSX_3}{\sigma^2} = \frac{414.61}{0.9032} = 459.0456$ and $\frac{SSX_4}{\sigma^2} = \frac{291.52}{0.9032} = 39.4595$ respectively.

The Posterior Precision for the Regression Parameters

The prior precision of the regression parameters, the mean square error, the sum of squares of the deviation were substituted in equations (17-21) to derive the posterior precision for the regression parameters. Table 7 below presents the posterior precisions of the 9 grid points.

TABLE 7: THE POSTERIOR PRECISION OF THE PARAMETER ESTIMATE

Posterior Precision of $\beta_{0,}$	Posterior Precision of $\beta_{1,}$	Posterior Precision of $\beta_{2,}$	Posterior Precision of $\beta_{3,}$	Posterior Precision of $\beta_{4,}$
1073.2235	50.5474	36.5143	491.1896	46.3182
1035.767	49.891	36.0527	489.7770	46.0168
1001.466	49.2899	35.6325	488.4833	45.7370
969.9377	48.7374	35.2555	487,2862	45.4835
940.8592	48.2280	34.9023	486.1826	45.2499
913.9555	47.7567	34.5769	485.1757	45.0336
888.9915	47.3194	34.2763	484.2281	44.8300
865.7646	46.9125	33.9978	483.3528	44.6435
844.0990	46.5331	33.7439	482.9528	44.4695

The unstandardized regression coefficients and the standard errors of the regression parameters $\beta_0,..,\beta_4$, were obtained from the linear regression analysis statistics obtained in Table 1 and the critical values of the parameters were also obtained from the student's t statistical table. These values were substituted into equation (8) to obtain the lower and upper limits for the prior means of the regression parameters.

The prior means for the 9 grids points of the regression parameters are presented below in Table 8

TABLE 8: THE PRIOR MEANS OF THE PARAMETER ESTIMATE $\boldsymbol{\beta_0}$,

LB	UB	$D_0 = \frac{UB - LB}{10}$	D_i	Prior mean of $\beta_{0,}$
0.4089	0.5538	0.01449	$D_1 = LB + D_0$	0.42339
			$D_2 = D_1 + D_0$	0.43788
			$D_3 = D_2 + D_0$	0.45237
			$D_4 = D_3 + D_0$	0.46686
			$D_5 = D_4 + D_0$	0.48135
			$D_6 = D_5 + D_0$	0.49584
			$D_7 = D_6 + D_0$	0.51033
			$D_8 = D_7 + D_0$	0.52482
			$D_9 = D_8 + D_0$	0.53931

TABLE 9: THE PRIOR MEANS OF THE PARAMETER ESTIMATE $\boldsymbol{\beta_1}$,

LB	UB	$D_0 = \frac{UB - LB}{10}$	D_i	Prior mean of $\beta_{1,}$
-0.3069	0.7887	0.10956	$D_1 = LB + D_0$	-0.19734
			$D_2 = D_1 + D_0$	-0.08778
			$D_3 = D_2 + D_0$	0.02178
			$D_4 = D_3 + D_0$	0.13134
			$D_5 = D_4 + D_0$	0.24090
			$D_6 = D_5 + D_0$	0.35046
			$D_7 = D_6 + D_0$	0.46002
			$D_8 = D_7 + D_0$	0.56958
			$D_9 = D_8 + D_0$	0.67914

TABLE 10: THE PRIOR MEANS OF THE PARAMETER ESTIMATE $\boldsymbol{\beta_2}$

LB	UB	$D_0 = \frac{UB - LB}{10}$	D_i	Prior mean of $\beta_{2,}$
-0.1459	1.20094	0.13468	$D_1 = LB + D_0$	-0.01122
			$D_2 = D_1 + D_0$	0.12346
			$D_3 = D_2 + D_0$	0.25814
			$D_4 = D_3 + D_0$	0.39282
			$D_5 = D_4 + D_0$	0.52750
			$D_6 = D_5 + D_0$	0.66218
			$D_7 = D_6 + D_0$	0.79686
			$D_8 = D_7 + D_0$	0.93154
			$D_9 = D_8 + D_0$	1.06622

TABLE 11: THE PRIOR MEANS OF THE PARAMETER ESTIMATE $\boldsymbol{\beta_3}$,

LB	UB	$D_0 = \frac{UB - LB}{10}$	D_i	Prior mean of β_3
0.3692	1.1154	0.07462	$D_1 = LB + D_0$	0.44382
			$D_2 = D_1 + D_0$	0.51844
			$D_3 = D_2 + D_0$	0.59306
			$D_4 = D_3 + D_0$	0.66768
			$D_5 = D_4 + D_0$	0.74230
			$D_6 = D_5 + D_0$	0.81692
			$D_7 = D_6 + D_0$	0.89154
			$D_8 = D_7 + D_0$	0.96616
			$D_9 = D_8 + D_0$	1.04078

TABLE 12: THE PRIOR MEANS OF THE PARAMETER ESTIMATE $\boldsymbol{\beta_4}$

LB	UB	$D_0 = \frac{UB - LB}{10}$	D_i	Prior mean of $\beta_{4,}$
-0.4713	1.14413	0.16154	$D_1 = LB + D_0$	-0.30975
			$D_2 = D_1 + D_0$	-0.14182
			$D_3 = D_2 + D_0$	0.01332
			$D_4 = D_3 + D_0$	0.17486
			$D_5 = D_4 + D_0$	0.33640
			$D_6 = D_5 + D_0$	0.49794
			$D_7 = D_6 + D_0$	0.65948
			$D_8 = D_7 + D_0$	0.82102
			$D_9 = D_8 + D_0$	0.98256

The prior precision, posterior precision, observation precision, prior mean and ordinary least squares estimates of the regression parameters $\beta_0,..,\beta_4$, obtained were substituted into equations (23-27) to derive the 9 Bayes Estimates.

The posterior mean (Bayes Estimates) and OLS estimates of the regression parameters for the 9 grids points were presented below in Table 13

TABLE 13: THE BAYES ESTIMATES OF THE PARAMETER ESTIMATES

BAYES ESTIMATES OF $\beta_{0,}$	BAYES ESTIMATES OF β_1	BAYES ESTIMATES OF β_2	BAYES ESTIMATES OF β_3	BAYES ESTIMATES OF β_4
0.4353	0.1117	0.3821	0.72270	0.2434
0.4471	0.1470	0.4224	0.72826	0.2682
0.4587	0.1802	0.4598	0.73330	0.2858
0.4770	0.2115	0.4947	0.73797	0.2942
0.4814	**0.2409**	**0.5275**	**0.74232**	**0.3364**
0.4923	0.2686	0.5583	0.74632	0.3564
0.5031	0.2949	0.5873	0.75090	0.3711
0.5137	0.3198	0.5991	0.75417	0.3974
0.5240	0.3435	0.6407	0.76052	0.4111

Table 14: Ordinary Least Squares Estimates of the Parameters.

Regression Parameters	$\beta_{0,}$	β_1	β_2	β_3	β_4
OLS Estimates	0.4814	0.2409	0.5275	0.7423	0.3364

From the Table 13 and 14 above, the posterior means produced at the average quantile level were the same as the ordinary least squares estimates of the parameters.

Average Absolute Deviation for the Grids Points

The Bayes estimates for each grid point derived were used to estimate new dependent variables$(\hat{Y})$. Average absolute deviation is determined between the simulated dependent variable (Y) and the Bayes estimated dependent$(\hat{Y})$. For each data set 9 Average absolute deviations were derived from which least Average absolute deviation is chosen. Table 15 presented the 9 Average absolute deviation values derived respectively for the data set.

TABLE 15: Average Absolute Deviation (AAD) of related quantiles

Quantile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Average Absolute Deviation	0.8825	0.8159	0.7996	0.7947	**0.7036**	0.7292	0.7145	0.8039	0.8193

From Table 15 above, the least Average Absolute Deviation was located at 5th quartiles of the simulated dependent variables.

Table 16: Frequency of the Quantiles with the Least Average Absolute Deviation.

Quantile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Average Absolute Deviation	0	0	35	268	4059	2193	3384	61	0

Table 16 above revealed the frequency of the quantiles with the Least Average Absolute Deviation for all the 10000 simulations. From Table 16, the optimal quantile range was located between quantiles 5 and 7, this implies that the optimal hyperparameters of the study lied mostly within 5th and 7thquantiles.

CONCLUSION.

This study worked on determining the quantile range at which optimal hyperparameters could be obtained when Bayesian estimation is employed to solve regression analysis of normally distributed data with vague information. The prior parameters were determined from ordinary least squares confidence intervals and the optimal quantiles were determined using the prior parameters. The least Average Absolute Deviations of the studyrevealed that the best model from 10000 exhaustive trials were within 5^{th} and 7^{th} grid points when the confidence intervals were divided into 10 quantiles.

The study, however revealed the following: the possibility of obtaining the prior distribution from data and distribution parameters from quantiles of the confidence intervals of OLS estimates; the optimal quantile range, where prior hyperparameter that produce the best model in regression analysis could be found. The research work minimizes the difficulties involved in identifying prior distribution when the true information of the data is vague. This is a valuable focus for research, given the increasing availability of alternative estimation methods within software packages.

REFERENCES

Acharya, M., Armaan, A, and Antony, A. (2019). A comparison of Regression Models for Prediction of Graduate Admissions. IEEE International Conference on Computational Intelligence in Data Science 2019.

Adankon Marthias and Cheietl Mohammed (2009). Least Squares Support Vector Machines- Worldscientific. Proceedings of 2009 International Joint Conference on Neural Networks. IJCNN'09, 1054-1061

Agresti. A. (2006). Bayesian inference for categorical data analysis. Department of Statistics University of Florida, Gainesville, Florida, USA. 3 2611-8545.

Atkinson, A. C., Chaloner, K., Juritz, J. and Herzberg, A.M. (1993). Optimum experimental designs forproperties of a computational model. *Biometrics. 49, 325-337.*

Bergstra, J. and Bengio Y. (2012). Random search for hyperparameter optimization. *Journal of MachineLearning Research, 13, 281-305.*

Brook S.P. MCMC methods and its application. *The Statistician 1998; 47, 69-100.*Casella, George (1985). An introduction to Empirical Bayes Data Analysis. American Statistician. *AmericaStatistical Association 39 (2): 83-87. Doi10.2307/2682801. JSTOR 2682801.*

Christle Biode Pierre (2012). Approximation of Improper Prior by Vague Priors. *Hal Publishers. 00880194*

Guido Consonni, Dimitris Fouskakis, Brunero Liseo and Ioannis Ntzoufras (2018). Prior Distributions for Objective Bayesian Analysis. *Bayesian Analysis 13,2,627-679.*

Hilary L. Seal (1967). 'The Historical development of Gauss Linear model'. Biometrica. 54 (1/2): 1-24.

Liseo, B. and Macaro, C. (2013). "Objective priors for causal AR(P) with partial autocorrelations." *Journalof statistical computation and Simulation 83:9, 1613-1628.*

Rencher, Alvin C.; Christensen, William F. (2012). Methods of Multivariate Analysis, Willey Series in Probability and Statistics, 709,(3rd ed), John Wiley & Sons, ISBN9781118391679.

Schaalje Bruce and Alvin C. Rencher. (2007). Linear models in Statistics. Wiley – Inter science, John Wiley & Sons, Inc., Hoboken, New Jersey.

Thornton C. , Hutter, H. Hoos H. and K. Leyton- Brown (2013) . Auto WEKA: Combined Selection andhyperparameter optimization of classification algorithms. In Proc. Of KDD'13

Yan, Xin (2009). Linear Regression Analysis: Theory and Computing, World Scientific, 1-2, ISBN9789812834119.