Tanzania Journal of Science 47(3): 988-998, 2021
ISSN 0856-1761, e-ISSN 2507-7961

Determination of Quantile Range of Optimal Hyperparameters Using Bayesian Estimation

Ayoade I. Adewole[1*] and Olusoga A. Fasoranbaku[2]
[1]Department of Mathematics, Tai Solarin University of Education Ijagun, P.M.B. 2118, Ijebu-Ode Ogun State Nigeria. E-mail: hayorhade2005@gmail.com
[2]Department of Statistics, Federal University of Technology Akure, P.M.B. 204, Akure, Ondo State Nigeria. E-mail: oafasoranbaku@yahoo.com
**Corresponding author*
Received 1 Feb 2021, Revised 10 July 2021, Accepted 20 Jul 2021, Published Aug 2021
DOI: https://dx.doi.org/10.4314/tjs.v47i3.10

Abstract

Bayesian estimations have the advantages of taking into account the uncertainty of all parameter estimates which allows virtually the use of vague priors. 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. A Monte Carlo simulation approach was used to generate a sample size of 200 data-set. Observation precisions and posterior precisions were estimated from the regression output to determine the posterior means estimate for each model to derive the new dependent variables. The variances were divided into 10 equal parts to obtain the hyperparameters of the prior distribution. Average absolute deviation for model selection was used to validate the adequacy of each model. The study revealed the optimal hyperparameters located at 5th and 7th deciles. The research simplified the process of selecting the hyperparameters of prior distribution from the data with vague information in empirical Bayesian inferences.

Keywords: Optimal Hyperparameters, Quantile Ranges, Bayesian Estimation, Vague prior.

Introduction

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 structures are 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, Bayesian approach provides a mathematical rule explaining how existing beliefs can be changed in the light of new evidence. In other words, it allows scientists to combine new data with their existing knowledge or prior. Researchers consider 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 was 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. Inferences drawn from 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. Model uncertainty can be ignored when inferences are conditioned on a single selected model, and this can lead to underestimation of conclusions drawn about quantities of interest.

This research work 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 applications 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. Authors Bergstra and Bengio (2012), Hutter et al. (2013), and Holder et al. (2021) worked on speeding up automatic selection of hyperparameter value for neural work. Liseo and Macaro (2013) and Petrone et al. (2014) considered the problem of deriving objective priors for the causal/ stationary autoregressive model of order p. Consonni et al. (2018) provided review of prior distributions for objective Bayesian analysis. Olubusoye and Okonkwo (2012) explored the application of Bayes theory to Normal Linear regression model in choosing prior distributions for the parameters of interest in the selection of variables in the case of reduced models. However, not much work has been done on determining the range of obtaining the optimal hyperparameter in Bayesian estimation of linear regression models. Furthermore, this study opts to determine the optimal quantile range of the prior parameters and determine the prior parameters from ordinary least squares (OLS) model confidence intervals.

Materials and Methods

Baye's Theorem for the regression model

Baye's theorem is constantly summarized by

$$\text{Posterior} \propto \text{prior} \times \text{likelihood} \qquad (1)$$

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

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

where;

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

ss_x denotes sum of squares of the independent variable

and

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

The 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 normal independent prior for each parameter, the joint prior of the two parameters is the product of the two individual priors.

$$g(a_x, \beta) = g(a_x) \times g(\beta) \qquad (5)$$

The joint prior follows a normal distribution.

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

$$g(a_x, \ \beta/data) \propto g(a_x, \ \beta) \times \text{likelihood sample } (a_x, \beta) \qquad (6)$$

Where the dataset is the set of ordered pair $(x_i, y_i), \ldots, (x_n, y_n)$

Regression analysis was run for a data-set, the standard errors obtained from the linear regression 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 the analysis using chi-square with (n-1) degree of freedom.

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

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

$$\bar{\beta}_s \pm t_{\propto/2,\text{n-2}} \times SE \quad (8)$$

The difference between lower and upper intervals obtained above are 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.

$$\beta \sim \text{N}(m_{\beta,}\ s^2_\beta) \quad (9)$$

$m_{\beta,}\ s^2_\beta$ denotes the hyperparameters of the normally distributed priors.

The likelihood function is of the form:

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

where $\bar{X}$ is the average of $X_i, \ldots, X_{n.}$

The likelihood of the regression parameters $(\beta_0 - \beta_4)$ were estimated, for the intercept (β_0), MSE is the mean square error from the regression analysis and the sample size n were used, while SSX_1 , SSX_2, SSX_3 and SSX_4 are sum of squares of the independent variables $(X_s), s = 1, \ldots, 4,$ respectively and the mean square error obtained were used to determine the likelihood of $\beta_1, \beta_{2,}\ \beta_{3,}$ and β_4.

Likelihood of $\beta_0 = \frac{MSE}{n}$ (11)

Likelihood of $\beta_1 = \frac{MSE}{SSX_1}$ (12)

Likelihood of $\beta_2 = \frac{MSE}{SSX_2}$ (13)

Likelihood of $\beta_3 = \frac{MSE}{SSX_3}$ (14)

Likelihood of $\beta_4 = \frac{MSE}{SSX_4}$ (15)

The posterior distribution:

$$\beta\ /\ X \sim N(m^{\iota}_{\beta,}\ \left(S^{\iota}_{\beta}\right)^2) \quad (16)$$

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

For $\beta_{0,}$ $\frac{1}{\left(S^{\iota}_{\beta_0}\right)^2} = \frac{1}{S^2_{\beta_0}} + \frac{n}{\sigma^2}$ (17)

For $\beta_{1,}$ $\frac{1}{\left(S^{\iota}_{\beta_1}\right)^2} = \frac{1}{S^2_{\beta_1}} + \frac{SSX_1}{\sigma^2}$ (18)

For $\beta_{2,}$ $\frac{1}{\left(S^{\iota}_{\beta_1}\right)^2} = \frac{1}{S^2_{\beta_2}} + \frac{SSX_2}{\sigma^2}$ (19)

For $\beta_{3,}$ $\frac{1}{\left(S^{\iota}_{\beta_1}\right)^2} = \frac{1}{S^2_{\beta_3}} + \frac{SSX_3}{\sigma^2}$ (20)

For $\beta_{4,}$ $\frac{1}{\left(S^{\iota}_{\beta_1}\right)^2} = \frac{1}{S^2_{\beta_4}} + \frac{SSX_4}{\sigma^2}$ (21)

The Bayes estimates of $\widehat{\beta_0} - \widehat{\beta_4}$ were obtained using the following equations:

$$\frac{Prior\ precision}{Posterior\ precision} \times Prior\ mean + \frac{Observation\ precision}{Prosterior\ precision} \times\ OLS\ estimate \quad (22)$$

For $\beta_{0,}$, $\frac{\frac{1}{S^2{\beta_0}}}{\frac{1}{S^2{\beta_0}} + \frac{n}{\sigma^2}} \times PM_{\beta_0} + \frac{\frac{n}{\sigma^2}}{\frac{1}{S^2{\beta_0}} + \frac{n}{\sigma^2}} \times \beta_0\ OLS\ estimate$ (23)

For $\beta_{1,}$, $\frac{\frac{1}{S^2{\beta_1}}}{\frac{1}{S^2_{\beta_1}} + \frac{SSX_1}{\sigma^2}} \times PM_{\beta_1} + \frac{\frac{SSX_1}{\sigma^2}}{\frac{1}{S^2_{\beta_1}} + \frac{SSX_1}{\sigma^2}} \times \beta_1\ OLS\ estimate.$ (24)

For $\beta_{2,}$, $\frac{\frac{1}{S^2{\beta_2}}}{\frac{1}{S^2_{\beta_2}} + \frac{SSX_2}{\sigma^2}} \times PM_{\beta_2} + \frac{\frac{SSX_2}{\sigma^2}}{\frac{1}{S^2_{\beta_2}} + \frac{SSX_2}{\sigma^2}} \times \beta_2\ OLS\ estimate.$ (25)

For $\beta_{3,}$, $\frac{\frac{1}{S^2\ \beta_3}}{\frac{1}{S^2_{\beta_3}} + \frac{SSX_3}{\sigma^2}} \times PM_{\beta_3} + \frac{\frac{SSX_3}{\sigma^2}}{\frac{1}{S^2_{\beta_3}} + \frac{SSX_3}{\sigma^2}} \times \beta_3\ OLS\ estimate.$ (26)

For $\beta_{4,}$, $\frac{\frac{1}{S^2{\beta_4}}}{\frac{1}{S^2_{\beta_4}} + \frac{SSX_4}{\sigma^2}} \times PM_{\beta_4} + \frac{\frac{SSX_4}{\sigma^2}}{\frac{1}{S^2_{\beta_4}} + \frac{SSX_4}{\sigma^2}} \times \beta_4\ OLS\ estimate$ (27)

Simulation study

Data were generated using the Monte Carlo approach for the analysis to determine the optimal hyperparameters and the optimal quantile range.

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

for I = 1…, n.

The general multiple 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 replicas of random data with sample size 200 were simulated for the four independent variables with y generated from a linear regression model with normally distributed error term. The parameters used for the simulation were chosen arbitrarily, given 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 $\varepsilon \sim N(0, 1)$,

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 determination of the optimal hyperparameter. The average absolute deviation for the posterior estimates for each of 9 grid points of the models were computed and the model with the least AAD is chosen 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 deviated was computed as

$$\frac{1}{N} \sum_{i,j,k=1}^{N,S,T} \left| Y_{i,,j,k} - \hat{Y}_{i,j,k} \right| \quad (29)$$

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 and $\hat{Y}_i$ 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 are presented in Table 1.

Table 1: Summary of linear regression outputs

Regression statistics						
Multiple R			0.6523			
R square			0.8836			
Adjusted R square			0.7106			
Standard error			0.8529			
Observation			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 estimation of the prior variances and prior precisions for the regression parameters $\beta_0 - \beta_4$

The standard errors of the regression parameters $\beta_0 - \beta_4$ from the regression analysis 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 and the incremental values added to the lower limit of each of the prior variance (LL) to obtain D_i and where i = 1 to 9. The prior precision obtained is the reciprocal of the prior variance. Tables 2 to 6 show the results of prior variances, prior precisions for the parameter estimates for the 9 grid points.

Table 2: The prior variance and precision for the parameter **e**stimate $\boldsymbol{\beta_0}$

		$\frac{(n-1)s^2}{\chi^2_{\alpha/2,,n-1}}$	$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,,n-1}}$	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 β_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 obtained from the mean of the independent variables SSX_1, SSX_2, SSX_3 and SSX_4 from the simulated data were 32.19, 24.09, 414. 61 and

291.52, respectively. The observation precisions for $\beta_{0,}\ \beta_{1,}\ \beta_{2,}\ \beta_{3,}\ and\ \ \beta_4$ 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 presents the posterior precisions of the 9 grid points.

Table 7: The posterior precisions of the parameter estimates

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 $\beta_0, \ldots, \beta_4$, were obtained from the linear regression analysis statistics in Table 1, and the critical values of the parameters were also obtained from the Student's 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 in Table 8 to Table 12.

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 β_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 β_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 β_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 estimate 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 are presented in Tables 13 and 14, respectively.

Table 13: The Bayes estimates of the parameters

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 data in Tables 13 and 14, the posterior means produced at the average quantile level in Table 13 highlighted in bold form were the same as the ordinary least squares estimates of the parameters in Table 14 which justifies the accuracy of the empirical Bayesian analysis.

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 variable $(\hat{Y})$. For each data set 9 average absolute deviations were derived from which least average absolute deviation is chosen. Table 15 presents 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, the least average absolute deviation was located at 5th quartiles of the simulated dependent variables.

Table 16 revealed the frequencies 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 the 5th and 7thquantiles.

Table 16: Frequencies 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

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 study revealed that the best model from 10000 exhaustive trials were within 5th and 7th grid points when the confidence intervals were divided into 10 quantiles.

The study also 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 hyperparameters 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. The research simplified the process of selecting the hyperparameters of prior distribution from the data with vague information in empirical Bayesian inferences, the Bayesian approach allows direct probability statements about the parameters which are much more useful than the confidence statements allowed by frequentist statistics. The results justified the work of Victoria and Moraga (2021) that worked on automatic tuning of hyperparameters using Bayesian optimization that concluded Bayesian method can clearly obtain optimized values of the hyperparameters and improve model selection performances.

References

Adankon M and Cheietl M 2009 Least squares support vector machines-World scientific. *Proc. 2009 Int. Joint Con. 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 AC, Chaloner K, Juritz J and Herzberg AM 1993 Optimum experimental designs for properties of a computational model. *Biometrics* 49: 325-337.

Bergstra J and Bengio Y 2012 Random search for hyperparameter optimization. *J. Mach. Learn. Res.* 13: 281-305.

Consonni G, Fouskakis D, Liseo B and Ntzoufras I 2018 Prior distribution for objective Bayesian analysis. *Bayesian Anal.* 13(2): 627-679.

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

Holder JP, Nguyen V and Roberts SJ 2021 Provably efficient online hyperparameters optimization with popul-based bandits. *34th Conference on Neural Information Processing System (NUL IPS 2020)* Vancouver Canada.

Liseo B and Macaro C 2013 Objective priors for causal AR(P) with partial autocorrelations. *J. Stat. Comput., Simulat.* 83(9): 1613-1628.

Lunn D, Barreti J, Sweeting M and Thompson S 2013 Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *J. Roy. Stat. Soc.* 62(4): 551-572.

Olubusoye OE and Okonkwo OC 2012 Choice of priors and variable selection in Bayesian regression. *J. Asian Sci. Res.* 2(7):354-377.

Petrone S, Rizzelli S, Rousseau J and Scricciolo C 2014 Empirical Bayes methods in classical and Bayesian Inference. *Metron* 72(2): 201-215.

Raftery AE, Madigan D and Hoeting JA 1997 Bayesian model averaging for linear regression models. *J. Am. Stat. Assoc.* 92(439): 179-191.

Victoria H and Moraga G 2021 Automatic turning of hyperparameters using Bayesian approaches. *Evolving Systems* 12: 217-221.