Al-Qadisiyah Journal of Pure Science Vol.(25) Issue (2) (2020) pp. Math. 1–13

Al-Qadisiyah Journal of Pure Science

ISSN(Printed): 1997-2490 ISSN(Online): 2411-3514

DOI: /10.29350/jops.

http://qu.edu.iq/journalsc/index.php/JOPS

Bayesian Variable Selection for Semiparametric Logistic Regression

Authors Names

a. **Zainab Sami**

b. **Taha Alshaybawee**

Article History

Received on:
Revised on:
Accepted on:

Keywords:

Logistic regression
Bayesian inference
Single index model
Lasso
MCMC algorithm

DOI: https://doi.org/10.29350/jops.2020.25. 2.1066

ABSTRACT

Lasso variable selection is an attractive approach to improve the prediction accuracy. Bayesian lasso approach is suggested to estimate and select the important variables for single index logistic regression model. Laplace distribution is set as prior to the coefficients vector and prior to the unknown link function (Gaussian process). A hierarchical Bayesian lasso semiparametric logistic regression model is constructed and MCMC algorithm is adopted for posterior inference. To evaluate the performance of the proposed method BSLLR is through comparing it to three existing methods BLR, BPR and BBQR. Simulation examples and numerical data are to be considered. The results indicate that the proposed method get the smallest bias, SD, MSE and MAE in simulation and real data. The proposed method BSLLR performs better than other methods.

1. Introduction

Regression analysis methods are fundamental in analyzing the relevant data by describing the relationship between a set of independent variables and the dependent variable (Kerlinger & Pedhazur, 1973). However, it is unable to describe and explain the relationships between the covariates and the response variable if the latter has binary value, where the nature of the response variable is required to be a continuous quantity and not a classification (Lea, 1997)[1]. This is why the need has arisen for developing new statistical methods that have the power of linear regression in reaching the best equations and dealing with them. Quite often, the outcome variable is discrete; taking on two possible values, it can have only two possible outcomes which will be denoted as 1 and 0. A problem with the regression model is that the predicted probabilities will not be limited between 0 and 1. Two relevant binary regression models, logit (logistic) and probit regression when the dependent variable is a binary response and take two values: 0 and 1

[a] Department of Mathematics, College of Science, University of Al-Qadisiyah, Diwaniyah, Iraq, E-Mail: waggas.galib@qu.edu.iq

[b] Department of Mathematics, College of Education for Girls, University of Al-Kufa, Najaf, Iraq, E-Mail: alih.battoor@uokufa.edu.iq, abeerfarhan688@gmail.com

[c] Department of Mathematics and Computer Science, University of Oradea, 1,Universitatii str., 410087 Oradea, Romania, E-Mail: georgia_oros_ro@yahoo.co.uk

$$y = \begin{cases} 0 & if \quad no \\ 1 & if \quad yes \end{cases}$$

y is a response variable distributed as Bernoulli with probability of success p_i.
The binary regression model is defined as p_i= F (x$'\beta$) i=1, ……., N where β is a k X I vector of unknown parameters, x$'$= $(x_{i1}\ldots\ldots\ldots x_{ik})$ is a vector of known covariates, and F ( ) is a known cdf, linking the probabilities p, with the linear structure (x$'\beta$). The logit model is obtained if F is the logistic cdf . Whereas the probit model is obtained if F is the standard Gaussian cdf.

Sometimes the explanatory variables are non-linear, which led researchers to find another method that deals with the nonlinear effect of these variables or nonparametric regression. It was proposed by the researcher (Jacob) in 1942. Nonparametric regression suffers from some problems, including the problem of dimensions (the curse of dimensionality). Therefore, the attractive features of single index model have motivated the researchers to extend this model for modelling a binary data. Kong & Xia (2008)[2] suggest that the single-index model is one of the most general semiparametric models in econometrics. Single index models suppose that the response interest depends on a linear combination of covariates through an unknown link function (Hu, et al., 2013)[3].

Subset selection by regularization has attracted much interest recently (see for example, lasso by Tibshirani, 1996). Tibshirani, R. (1996)[4] proposed that lasso estimates will be taken as posterior mode estimates once the regression parameters are assigned independent and corresponding standard. Park and Casella (2008) [5] introduced the Bayesian lasso regression, using a conditional Laplace prior distribution represented as a scale mixture of normal with an exponential mixing distribution. Bayesian analysis method has become very widely applicable, as a result of its ability to benefit from all available information in the analysis. Bayesian variable selection is a flexible method for translating prior information into a selection of variables (Fridley, 2009)[6]. Several variable selection methods are used with a Bayesian framework.

In this paper the researchers also formulated the Bayesian lasso penalty approach for estimating and selecting variables in a single index logistic regression model. Nonetheless, to the best of our knowledge, no such research has been considered before.

2-Single Index Logistic regression model and prior assumption:

Single-index model (SIM) introduce an efficient manner of handling high dimensional nonparametric estimation problems (Hardle et al., 1993; Yu and Ruppert, 2002)[7] and avert the 'curse of dimensionality' (Bellman et al., 1966)[8]. Nonparametric problems assume that the response is just associated with a single linear set of the covariates. It's one of the most common and necessary semiparametric models in statistics as well as applied sciences like econometrics and psychology due to its ability to reduce dimensions (Ichimura, 1993)[9]. The semiparametric single index regression model is:

$$y = m(X_i^T\beta) + \varepsilon_i \qquad (1)$$

where y is a response variable, $\beta$is a parameter vector (Parametric part), m: is an unknown link function (nonparametric part) and ε_i= errors are assumed to be iid.

The basic assumptions of logistic regression model are based on that the dependent variable (y) is binary take either of the two values (1, 0), with a success probability (P_i) and failure probability (1-P_i). Therefore, the response variable (y) is distributed as Bernoulli distribution and can be expressed as follows:

$$p(y) = {P_i}^{yi}(1 - P_i)^{1-yi} \qquad (2),$$

$$P_i = \frac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik}}}$$

$$1 - P_{i=}\frac{1}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik}}}$$

where y is a binary independent variable (1,0), $x_i, i = 1,2,\dots,k$ is covariates variables, P_i is the probability of success when y=1, $1 - P_i$ is the probability of failure when y=0, and $\beta_0, \beta_1, \dots, \beta_k$ is unknown coefficients vector of the logistic regression model.

In the linear regression model the researchers assume that an observation of the outcome variable may be expressed as $y = E(Y|x) + \varepsilon$. The most common assumption is that the error term ε follows a normal distribution with mean zero and some variance which is constant across levels of the independent variable. It follows that the conditional distribution of the outcome variable given x is normal with mean $E(Y|x)$, and a variance that is constant. This is not the case with a dichotomous outcome variable.

In this situation, we may express the value of the outcome variable given x as $y = p(x) + \varepsilon$. Here the quantity ε may assume one of two possible values. If $y = 1$ then $\varepsilon = 1 - p(x)$ with probability $p(x)$, and if $y = 0$ then $\varepsilon = -p(x)$ with probability $1 - p(x)$. Thus, ε has a distribution with mean zero and variance equal to $p(x)[1 - p(x)]$. That is, the conditional distribution of the outcome variable follows a binomial distribution with probability given by the conditional mean $p(x)$.

$$P_i = p(y = 1) = \frac{exp\ (m(x_i'\boldsymbol{\beta})}{1 + exp\ (m(x_i'\boldsymbol{\beta})}$$

P_i: represent the probability of the response that can be expressed:

$$1 - P_i = p(y = 0) = \frac{1}{1 + exp\ (m(x_i'\boldsymbol{\beta})}$$

$1 - P_i$: represent the probability of non-response that can be expressed.

The likelihood function is the probability density function of the data which is seen as a function of the parameter treating the observed data as fixed quantities. For a given sample size n, the likelihood function is given as:

$$L(y|m, \boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i)$$

$$L(y|m,\boldsymbol{\beta}) = \prod_{i=1}^{n} P_i^{\,y_i}(1-P_i)^{1-y_i}$$

Therefore, the likelihood function can be described as follows:

$$L(y|m,\boldsymbol{\beta}) = \prod_{i=1}^{n}\left(\frac{exp\ (m(x_i'\boldsymbol{\beta})}{1+exp\ (m(x_i'\boldsymbol{\beta})}\right)^{y_i}\left(\frac{1}{1+exp\ (m(x_i'\boldsymbol{\beta})}\right)^{1-y_i} \quad (3)$$

Following Choi et al. (2011)[10] and Gramacy and Lian (2012)[11], the researchers will set Gaussian process as the prior distribution for the unknown link function $m(.)$. Therefore, the distribution of $m(.)$ is a Gaussian process with zero mean and square exponential covariance function. It is written as follows:

$$m \sim GP\big(0, E(.,.)\big), \qquad where \qquad E(x,x') = \partial \frac{(x-x')^2}{w}$$

where ∂ and w are hyperparameters, so that this framework of single-index model can use the observed covariates, which can be shown as:

$$\pi(m_n|\beta,\partial) = \ det[E_n]^{-1/2} exp\left\{-\frac{m_n{}'\, E_n{}^{-1}\, m_n}{2}\right\}$$

where E_n is the covariance matrix with the dimension (n x n) and elements E (., .) as given in the equation:

$$E\big(x_i\,, x_j\,\big) = \partial\ exp\{-\big(x_i - x_j\big)'\beta\beta'\big(x_i - x_j\big)\}$$

As same as Gramacy and Lian (2012)[11], when the Gaussian process is considered as a prior distribution to the unknown function, $\frac{\beta}{\sqrt{w}}$ is identifiable without the necessity for the constraint $\|\beta\| = 1$. Therefore, the researchers will, instead of $\frac{\beta}{\sqrt{w}}$ by β, reformulate the covariance function as follows:

$$E\big(x_i\,, x_j\,\big) = \partial\ \exp\left\{-\big(\,x_i{}'\boldsymbol{\beta} -\ x_j{}'\boldsymbol{\beta}\big)^2\right\} \quad (4)$$

The inverse gamma distribution is considered as prior, where it implies that $\partial \sim Inv.Gamma(a,b)$ where a and b are the hyperparameters. Following Park and Casella (2008) and Hu et al. (2013) conditional Laplace distribution is set as prior for the parameter vector and can be formed as:

$$f(\beta|\sigma\,,\lambda) = \prod_{j=1}^{p} \frac{\lambda}{2\sigma} \exp(-\lambda|\beta_j|\,/\sigma), \qquad \lambda > 0.$$

Where the prior distribution for $\sigma\,,\lambda$ are set as follows

$$\lambda \sim Gamma(c,d)\ \ , \sigma \sim Inv.Gamma(e,f)$$

3-Hierarchical model Posterior distribution

Bayesian hierarchical model for single index logistic regression model regularize by lasso is provided as follows:

$$\boldsymbol{f}(\boldsymbol{y}\,|\boldsymbol{m},\boldsymbol{\beta},\partial,\sigma\,,\lambda)=\prod_{i=1}^{n}\left(\frac{exp\,(m(x_i'\beta)}{1+\,exp(m(x_i'\beta)}\right)^{yi}\left(1-\frac{exp(m(x_i'\beta)}{1+\,exp(m(x_i'\beta)}\right)^{1-yi}$$

$$m|\beta,\partial\sim GP\big(0,E(.,.)\big),$$

$$\beta|\sigma\,,\lambda=\prod_{j=1}^{p}\frac{\lambda}{2\sigma}\exp(-\lambda|\beta_j|\,/\sigma) \qquad (5)$$

$$\partial\sim Inv.Gamma(a,b)$$

$$\sigma\sim Inv.Gamma(e,f)$$

$$\lambda\sim Gamma(c,d)$$

By using MCMC algorithm the researchers have found the conditional distribution for all parameters. The conditional posterior distribution for all parameters has been derived as follows:

- link function $\boldsymbol{m}|\,\boldsymbol{x_i},\boldsymbol{\beta},\partial,\sigma,\lambda\,,\boldsymbol{y}$ can be sample from the following conditional distribution:

$$\boldsymbol{\pi}(\boldsymbol{m}|\,\boldsymbol{x_i},\boldsymbol{\beta},\partial)$$
$$\propto\prod_{i=1}^{n}\left(\frac{exp\,(m(x_i'\beta)}{1+\,exp(m(x_i'\beta)}\right)^{yi}\left(1-\frac{exp(m(x_i'\beta)}{1+\,exp(m(x_i'\beta)}\right)^{1-yi}$$
$$\times\,det[E]^{-1/2}exp\left\{-\frac{m'\,E^{-1}\,m}{2}\right\}$$

- the conditional distribution of the parameter vector can be shown as:

$$\boldsymbol{\pi}(\boldsymbol{\beta}|\,\boldsymbol{x_i},\boldsymbol{m},\lambda,\sigma,\boldsymbol{y})$$
$$\propto\prod_{i=1}^{n}\left(\frac{exp\,(m(x_i'\beta)}{1+\,exp(m(x_i'\beta)}\right)^{yi}\left(1-\frac{exp(m(x_i'\beta)}{1+\,exp(m(x_i'\beta)}\right)^{1-yi}$$
$$\times\,det[E]^{-1/2}exp\left\{-\frac{m'\,E^{-1}\,m}{2}\right\}\times\prod_{j=1}^{p}\frac{\lambda}{2\sigma}\exp(-\lambda|\beta_j|\,/\sigma)$$

- The conditional distribution function of λ can be written as:

$$\boldsymbol{\pi}(\lambda|\,\boldsymbol{x_i},\boldsymbol{m},\boldsymbol{\beta},\partial,\sigma,\boldsymbol{y})\propto\prod_{j=1}^{p}\frac{\lambda}{2\sigma}\exp(-\lambda|\beta_j|\,/\sigma)\times\lambda^{c-1}\exp(-d\lambda)$$

Therefore, the conditional posterior of λ is Gamma distribution $(p + c, d + \sum \left|\beta_j\right|/\sigma)$.

- The conditional distribution of ∂ is given as:

$$\boldsymbol{\pi}(\partial |\, \boldsymbol{x_i}, \boldsymbol{m}, \boldsymbol{\beta}, \lambda, \sigma, \boldsymbol{y}) \propto \; det[E]^{-1/2} exp\left\{-\frac{m'\,E^{-1}\,m}{2}\right\} \times \left(\frac{1}{\partial}\right)^{a_{\partial},+1} \exp\{\frac{b_{\partial}}{\partial}\}$$

- The conditional distribution σ is given as:

$$\boldsymbol{\pi}(\sigma |\, \boldsymbol{x_i}, \boldsymbol{m}, \boldsymbol{\beta}, \lambda, \partial, \boldsymbol{y}) \propto \prod_{j=1}^{p} \frac{\lambda}{2\sigma} \exp(-\lambda|\beta_j|\,/\sigma) \times \left(\frac{1}{\sigma}\right)^{e+1} \exp\{-\frac{f}{\sigma}\}$$

The posterior distribution of σ is Inverse Gamma $(p + e, f + \lambda \sum \left|\beta_j\right|)$

An efficient Gibbs sampler algorithm is considered to sample σ and λ, whereas a Metropolis-Hastings algorithm is used to sample $\boldsymbol{\beta_\tau}, \boldsymbol{m_n}$ and ∂. The researchers set the initial values for the hyperparameters a, b, c, d, e and f as (0.1).

4-Simulation study

Simulation examples are considered in this section to evaluation our proposed method Bayesian semiparametric lasso logistic regression (BSLLR). In this study, the researchers will compare our proposed methods (BSLLR) with some other existing methods, i.e., Bayesian Logistic Regression (BLR) and Bayesian Probit Regression (BPR). These methods are included in MCMC package. R package and Bayesian Binary Quantile Regression (BBQR) are included in Bayesian QR package. Two examples are reported in this study which are already used by many papers and researchers for instance (Hu et al. (2013), Alshaybawee et al. (2016)[12], Zhao and Lian (2015)[13], Alkenani and Yu (2013)[14], Lv et al. (2014)[15] and Kuruwita (2015))[16]. R code is constructed to implement MCMC algorithm and the algorithm is run 15000 iterations where the first 3000 remove as burn in.

4.1 Example One

The following regression model is considered to generate three samples size (n=50,150 and 250) each with 100 replication:

$$y_i^* = \rho(\boldsymbol{x_i'}\boldsymbol{\beta}) + \sqrt{\left(\sin(\boldsymbol{x_i'}\boldsymbol{\beta}) + 1\right)}\,\varepsilon, \quad , y_i = \begin{cases} 1 & if \quad y_i^* \geq 0 \\ 0 & otherwise \end{cases}$$

where $\rho(r) = 10\sin(0.75r)$, x_i $(i = 1,2,\ldots,6)$ are the explanatory variables from a normal distribution with $[0, (1/4)^2]$, $\boldsymbol{\beta} = \frac{1}{\sqrt{5}}(0,1,0,0,2,0)^T$, and ε is the error term distributed as standard normal.

In Table (1) the researchers summarize the bias to the parameters that are estimated by all the methods under study, i.e., the existing methods BLR, BPR and BBQR and the proposed method BSLLR. At the three samples size it can be seen that the proposed method, very clearly, get the smallest values of bias for all estimated parameters. This means that the estimated parameters are very close to the true parameters. On the other hand, it can be seen that the BBQR method gets the largest values of bias for most estimated parameters and at all samples size. For the other methods it can be seen that the BPR method gets bias values smaller than the BLR methods for most estimated parameters and at all samples size.

Table (1): The average bias of the parameter estimates of BSLLR, BLR, BPR and BBQR methods for the samples (Simulated Example 1)

N	Methods	$Bias.\beta_1$	$Bias.\beta_2$	$Bias.\beta_3$	$Bias.\beta_4$	$Bias.\beta_5$	$Bias.\beta_6$
50	**BSLLR**	0.1618972	0.0983581	0.0967632	0.163808	0.0672587	0.1618972
	BLR	0.6297775	6.0971639	4.6631283	3.818511	7.1292936	0.6297775
	BPR	0.3597775	5.6071639	4.0581283	3.438511	6.7729578	0.3597775
	BBQR	0.4255645	4.9956645	5.2095194	4.530049	7.6333901	0.4255645
150	**BSLLR**	0.1392425	0.1263955	0.0553966	0.1011218	0.4366313	0.1392425
	BLR	2.503743	7.2406796	0.1250081	2.2113468	7.4158539	2.503743
	BPR	2.123743	6.3630896	0.1182561	2.0513468	6.6336169	2.123743
	BBQR	1.7413032	7.8514459	0.0578579	1.9596144	6.8784795	1.7413032
250	**BSLLR**	0.2112708	0.0648566	0.0214108	0.1364848	0.4260109	0.2112708
	BLR	1.7674131	5.9690132	1.4424766	0.7821558	5.6917958	1.7674131
	BPR	1.5308931	5.1845732	1.2756936	0.6697778	4.8272193	1.5308931
	BBQR	2.0271176	6.7745163	1.7722985	1.4125118	6.0746302	2.0271176

Based on 100 replications shown in Figure (1) the standard division for the parameters estimate by all proposed and existing methods. It can be seen the SD for the proposed method BSLLR are the smallest compared to the other methods and over all samples size. BBQR method gets small values compared to the other two methods BLR and BPR when the sample size 50, whereas these values are increase and exceed that for BLR and BPR at (n=150 and 250). The SD values for BLR method are larger than SD values for BPR method at all cases.

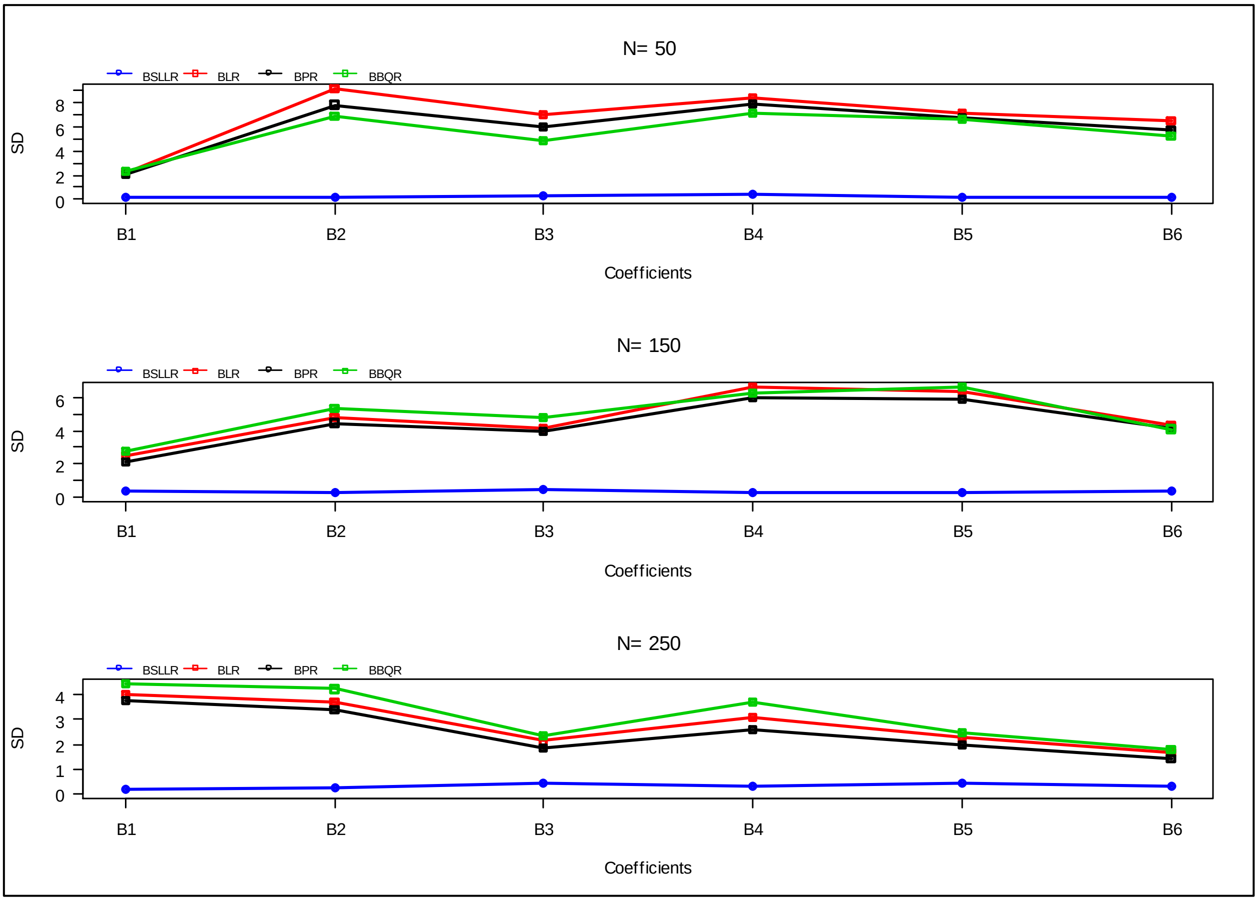

Figure 1. show the SD values for BSLLR, BLR, BPR and BBQR methods at three samples size (Example 1).

Table (2) shows MSE and MAE values for all methods in this study. The proposed method BSLLR gets the smallest values of MSE and MAE compared to the other methods. The existing method BBQR gets the largest values of MSE and MAE compared to the other two methods BLR and BPR when the samples size (150 and 250) but it is smaller than these methods when the sample size is 50. BPR method gets MSE and MAE values smaller than BLR method at all cases.

Table (2). The values of MSE and MAE of BSLLR, BLR, BPR and BBQR methods for each sample (Simulated Example 1)

N	Methods	MSE	MAE
50	**BSLLR**	0.4993051	0.5153148
	BLR	1.0330221	0.8555687
	BPR	0.9283775	0.8063997
	BBQR	0.7734473	0.7924135
150	**BSLLR**	0.4864899	0.4976007
	BLR	0.8336368	0.6803427
	BPR	0.7721769	0.6334271
	BBQR	0.8651809	0.7119312
250	**BSLLR**	0.5066259	0.5217550
	BLR	0.6785633	0.6574913
	BPR	0.6347043	0.6027816
	BBQR	0.8846936	0.7447513

4.2 Example Two

Three samples size (n=50, 150 and 250) with 100 replications are generated from the following regression model:

$$y_i^* = g(\boldsymbol{x}_i^T\beta) + \varepsilon, \quad y_i = \begin{cases} 1 & if \quad y_i^* \geq 0 \\ 0 & otherwise \end{cases},$$

where $g(t) = \exp(t)$, x_i $(i = 1,2, \ldots, 5)$, where the independent variables are generated from $N(0,1)$, $\beta = (\beta_1, \beta_2, \ldots, \beta_5)^T = \frac{1}{\sqrt{3}}(1,1,0,0,1)^T$and the error term generated from standard normal distribution.

Table (3) show the bias values to all parameters that estimate by the methods in this study. The proposed method gets the smallest values of bias compared to the other methods for all the estimated parameters at all samples size. BBQR methods gets the largest values of bias compared to the other methods. The bias for the BPR is smaller than that for the BLR method for most of the parameters estimated and at all samples size. Figure (2) summarizes the SD values for the proposed and existing methods. This figure shows that the proposed method gets the smallest values of SD compared to the other methods in all samples size. In the other side, the BBQR method gets the largest values of SD for most of the parameters estimated and at the samples size. The BPR method gets small values of SD compared to the BLR method.

Table (3): The average bias of the parameter estimates of BSLLR, BLR, BPR and BBQR methods for the samples (Simulated Example 2)

N	Methods	$Bias.\boldsymbol{\beta_1}$	$Bias.\boldsymbol{\beta_2}$	$Bias.\boldsymbol{\beta_3}$	$Bias.\boldsymbol{\beta_4}$	$Bias.\boldsymbol{\beta_5}$
50	**BSLLR**	0.5898498	0.1879269	0.0340038	0.1570391	0.5723522
	BLR	1.4213798	1.3231986	0.0825936	0.8771994	1.5418044
	BPR	1.1513798	1.1331986	0.0775936	0.6971994	1.1818044
	BBQR	1.793648	1.3554607	0.1146276	0.2682978	1.9055461
150	**BSLLR**	0.5888237	0.2537129	0.1471941	0.211417	0.3486755
	BLR	0.390086	0.3838365	2.6758617	0.6275007	0.5400257
	BPR	0.310086	0.2938365	2.2758617	0.4675007	0.4700257
	BBQR	1.1218885	0.3948856	2.4482466	1.5275868	2.2879032
250	**BSLLR**	0.4253592	0.2702062	0.0708861	0.1805457	0.4818558
	BLR	0.9265998	0.4810002	1.0117244	0.5187261	0.4665055
	BPR	0.8065998	0.4410002	0.8817244	0.4387261	0.4019555
	BBQR	0.5430176	0.2440631	2.6227755	1.3646287	0.0445077

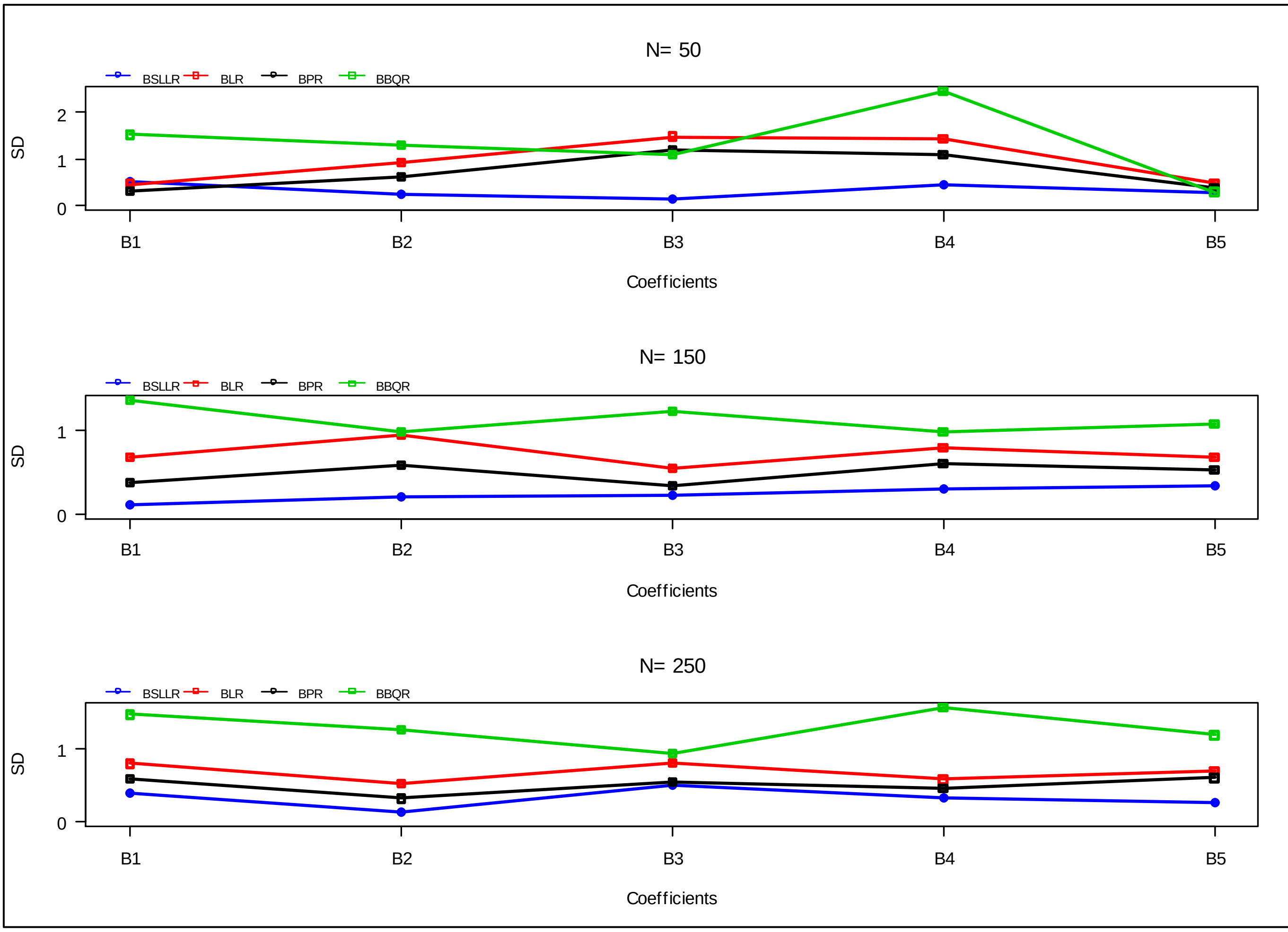

Figure 2. show the SD values for BSLLR, BLR, BPR and BBQR methods at three samples size (Example 2).

Table (4) shows the MSE and MAE values for the model that is estimated by the proposed and existing method. It is clear that the proposed method BSLLR gets the smallest values of MSE and MAE compared to the other methods. The existing method BBQR gets the largest values of MSE and MAE compared to all the other methods in this study. The MSE and MAE values for the BLR method are bigger than that values for the BPR method at all samples size.

Table (4). The values of MSE and MAE of BSLLR, BLR, BPR and BBQR methods for each sample (Simulated Example 2)

N	Methods	MSE	MAE
50	**BSLLR**	0.5597303	0.7260386
	BLR	6.6902721	2.5322350
	BPR	6.0944381	2.1389660
	BBQR	7.4845307	2.6566942
150	**BSLLR**	0.2845815	0.5161062
	BLR	2.0601757	1.1752145
	BPR	1.8675441	1.0045329
	BBQR	4.0915452	1.6901830
250	**BSLLR**	0.3556246	0.5836441
	BLR	0.2050050	0.3706665

BPR	0.1845571	0.2976875
BBQR	4.3777821	1.8898055

5-Real Data Example

The real data that is considered in this study is (churn) data. This dataset is included in bayesQR package in R. This dataset describes a random sample of the active customers at the end of June 2006 of a EFS company. This dataset consists of four explanatory variables ‘gender’: which shows the gender of customer (female=0, male=1), ‘Social-Class-Score’: which shows the social class of customer. ‘lor’: which shows the length of relationship with the customer, ‘recency’: which shows the number of days since last purchase, whereas the independent variables are ‘churn’: churn (yes/no). This data is constructed with 400 observations. The proposed and existing methods are employed to modelling this dataset. The results are reported as follows:

Table (5): The parameter estimates of BSLLR, BLR, BPR and BBQR methods for the real data example.

Methods	β_1	β_2	β_3	β_4
BSLLR	- 0.01398	0.001937	-0.21524	0.57435
BLR	-0.02847	0.022106	-0.61732	0.56903
BPR	-0.02098	0.016402	-0.37823	0.34526
BBQR	-0.03578	0.012224	-0.81171	0.75725

Table (5) shows the parameters estimates by the methods in this study. This table shows that the parameter estimate for the second independent variable is so small compared to the mean. This variable is not important. While in the proposed method, the parameter is closer to the zero.

Table (6). The values of MSE and MAE of BSLLR, BLR, BPR and BBQR methods for the real data example.

Methods	MSE	MAE
BSLLR	0.5458904	0.6239342
BLR	0.9924065	0.7866441
BPR	0.7384185	0.7081966
BBQR	1.0645598	0.8367619

In Table (6) the MSE and MAE values are summarized. It can be seen that the proposed method gets the smallest values of MSE and MAE compared to other methods. The largest values of MSE and MAE are for BBQR method. As same as the results in simulation examples, BLR method gets MSE and MAE values larger than BPR.

1- Conclusion

In this paper, Bayesian estimation and variable selection approach are suggested to estimate the parameters and link function and select the important variables for single index logistic regression model. Laplace distribution is set as prior to the coefficients vector and prior to the unknown link function (Gaussian process). A hierarchical Bayesian lasso semiparametric logistic regression model is constructed and MCMC algorithm is adopted for posterior inference.

Three existing methods BLR, BPR and BBQR are considered to be compared with the proposed method BSLLR. Real data and two simulation examples are used to compare the performance of BSLLR with the existing methods BLR, BPR and BBQR. The results indicate that the proposed method gets the smallest bias, SD, MSE and MAE in simulation and real data. In most cases BBQR method gets the largest values of bias compared to the other existing methods, i.e., SD, MSE and MAE. In addition, it can be seen that the existing method is doing better than BLR method and gets small values of bias. The researchers conclude that the proposed method BSLLR performs better than the other methods, i.e., SD, MSE and MAE.

References

[1] Alshaybawee, T., Midi, H., & Alhamzawi, R. (2017). Bayesian elastic net single index quantile regression. Journal of Applied Statistics, 44(5), 853-871.

[2] Alkenani, A., & Yu, K. (2013). Penalized single-index quantile regression. International Journal of Statistics and Probability, 2(3), 12

[3] Bellman, R., Kalaba, R., & Sridhar, R. (1966). Adaptive control via quasi-linearization and differential approximation. *Computing*, *1*(1), 8-17.

[4] Choi , T., Q. Shi, J. & Wang, B. (2011). A Gaussian process regression approach to a single-index model. In *Journal of Nonparametric Statistics*. Vol, 23, no 1, pp. 21-36

[5] Fridley, B. L. (2009). Bayesian variable and model selection methods for genetic association studies. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, *33*(1), 27-37.

[6] Hu, Y., Gramacy, R. B., & Lian, H. (2013). Bayesian quantile regression for single-index models. Statistics and Computing, 23(4), 437-4542010
[9] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of econometrics*, *58*(1-2), 71-120.

[7] Gramacy, R. B., & Lian, H. (2012). Gaussian process single-index models as emulators for computer experiments. *Technometrics*, *54*(1), 30-41

[8] Greenberg, E. (2008). *Introduction to Bayesian Econometrics*. Cambridge University Press, New York Haggag,

[9]Kuruwita, C. N. (2016). Non-iterative Estimation and Variable Selection in the Single-index Quantile Regression Model. *Communications in Statistics-Simulation and Computation*, *45*(10), 3615-3628.

[10] Lea, S. (1997). " Multivariate Analysis II: Manifest variables analysis.Topic 4: Logistic Regression and Discriminant Analysis ", University of EXETER, Department of Psychology. Available at:
www.exeter.ac.uk/~SEGLea/multivar2/diclogi.html

[11] Lv, Y., Zhang, R., Zhao, W., & Liu, J. (2014). Quantile regression and variable selection for the single-index model. *Journal of Applied Statistics*, *41*(7), 1565-1577

[12] Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, no.*103*(482), pp.681-686.

[13] Richardson, S., Leblond, L. and Jaussent, I. and Green, P. (2002). Mixture models in measurement error problems, with reference to epidemiological studies. In *J. R. Statist. Soc*. vol.165, Part 3, pp. 549-566

[14] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288

[15] Zhao, K., & Lian, H. (2015). Bayesian Tobit quantile regression with single-index models. Journal of Statistical Computation and Simulation, 85(6), 1247-1263