New Bayesian Estimation for Single Index Model

Rasha Majeed Abed Alghanemi
Department of Statistics, College of Administration and Economics, University of Al-Qadisyiah, IRAQ, admin.stat21.17@qu.edu.iq

Dr. Taha Hussain Ali Alshaybawee
Asst. Prof, Department of Statistics, College of Administration and Economics, University of Al-Qadisyiah, IRAQ, taha.alshaybawee@qu.edu.iq

Abstract

Generalization of the semi-parametric single index model to be more flexible than the general linear model by allowing non-linear relationships between the index function and the response variable. In this paper, new estimation and variable selection method through Bayesian approach is proposed. We have construct new hierarchical model based on the representation of scale mixture of normal distribution mixing Rayleigh density for the double exponential prior density of the parameters vector. Two simulation examples and real data are considered to evaluation our proposed method compare to some existing methods and we get some results.

Keywords: *Bayesian approach, scale mixture of Rayleigh, MCMC, Gaussian process, Single index model.*

INTRODUCTION

Generalization of the semi-parametric single index model (SSIM) to be more flexible than the general linear model by allowing non-linear relationships between the index function $(X_i'\beta)$ and the dependent variable (Y):-

$$Y = g(X_i'\beta) + u_i \quad \ldots\ldots (1)$$

Where y is response variable, β is a parameter vector (parametric part), g (.) is the nonparametric function (non- parametric part) and u is the error term $u \sim N(0, \sigma^2)$.

Among the regularization approaches, it is perhaps the most widely used approach in the recent literature for lasso variant selection. Imposing an $(\mathcal{L})$ penalty on the fitted vector modulus, the lasso performs a continuous contraction and an automatic variable selection at the same time. Lasso estimations can be achieved for processed transactions Lasso methods handle data maximization by reducing the value of the parameter (β) to the smallest possible by (Tibshirani,1996):-

$$\mathcal{L}(\beta, \gamma) = \ell_2(y - x'\beta)^2 + \gamma\|\beta\|_1 \quad \ldots\ldots (2)$$

Where $\ell_2(.)$ is the ℓ_2-Norm and $\boldsymbol{\gamma} \geq 0$ is the shrinkage parameter of β (Efron et al ,2004). The least absolute shrinkage and selection operator (lasso) has been established as a key workhorse of researchers in all domains working with high-dimensional regulation (Korobilis et al,2021) . Lasso estimates is the posterior mode when the prior distribution of the regression parameter distributed according to Laplace distribution (LD) . Considered the lasso method is multiuse function (multi-part) linear that defined on a series of(γ) and suggest that the solution path of (γ) following algorithm called Least Angle Regression (LAR) which suggests that the posterior distribution of regression on parameters is linearly for is series interval $\gamma[\gamma_z, \gamma_{z+1}]$ (Tibshirani,1996).

(Park & Casella, 2008) considered New Bayesian lasso through hierarchical model that represent the double exponential prior density as scale mixture of normal mixing (SMN) with Laplace distribution (LD). The full joint Bayesian posterior distribution of regression parameters for conditional Laplace distribution (LD) (Tibshirani,1996):-

$$\pi(\beta, \sigma^2|\bar{y}) \propto \pi(\sigma^2)(\sigma^2)^{-(n-1)/2} exp\left\{\frac{1}{2\sigma^2}\ (\bar{y} - x\beta)^T(\bar{y} - x\beta) - \lambda \sum_{j=1}^{p} |\beta_j|\right\} \quad ..(3)$$

A new family of scale normal mixtures (SNM) is identified , This class of distributions, the exponential family, has been used widely in robustness studies; it was introduced and popularized by (Box & Tiao ,1973) in the context of Bayesian modeling for robustness. However, The normal mixture property of the scale and the interesting relationship with the class of stable distributions are discussed so far. That exponential family distributions are scale normal mixtures (SNM). In this paper, It can be proven through the method construct new hierarchical model representation considering the double exponential prior density of the parameters as scale mixture of normal distribution mixing Rayleigh density (SMNR) (Flaih et al ,2020):-

$$\frac{1}{2\gamma} \exp\{-\frac{|\beta_j|}{\gamma}\} = \int_0^\infty \frac{1}{\sqrt{2\pi s}} e^{-\beta^2 j/2s} \frac{s}{\gamma} e^{-s^2/2\gamma} ds \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

2. Bayesian single index model proposed method and prior assumption:

Consider the following single index defined as (Ichamuri,1993)

$$\mathrm{Y} = \mathrm{g}(\, X_i'\beta) + u_i \quad , (\mathrm{i} = 1,2,3 \ldots \mathrm{n}) . \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Where $y_1, y_2 \ldots\ldots\ldots\ldots y_n$ are the dependent variables, $u_1, u_2 \ldots\ldots\ldots\ldots u_n$ are the errors and assumed to be (iid) normal distribution with (zero) mean and unknown $\sigma^2 \sim (0, \sigma^2{}_u)$, $\boldsymbol{X_i} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots \ldots \boldsymbol{x_p})'$ is p-dimensional predictive variables and g(.) is the unknown link function and β_j is the index parameters.

The likelihood function for error $\boldsymbol{u_i} = \boldsymbol{y_i} - \mathbf{g}(\, \boldsymbol{X_i'}\boldsymbol{\beta})$.(i=1,2,......,n) .Where (n) is the sample size can be shown as:-

$$\pi(u/\sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(u_1)^2\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

$$\pi(u/\mathrm{g}, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\ \frac{y_i - \mathrm{g}(\, X_i'\beta)^2}{\sigma^2}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

$$\pi(u/\mathrm{g}, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\ \sum_{i=1}^{n} (y_i - \mathrm{g}(\, X_i'\beta)^2\right\} \ldots\ldots\ldots\ldots (8)$$

The Gaussian process distribution set as prior for the unknown link function $\mathbf{g}(.)$

In other word , $\mathbf{g}(.)$ function is a Gaussian process with mean (zero) and square exponential covariance furcation for more detail (see Chio etal, 2011, and Gramcy &Lian 2012) , it on be shown as :-

$$\mathrm{g}(.) \sim GP\ (0, E(.,.)) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

$$E(X_i, X_j) = \mathcal{S}\ \ exp\left\{-\frac{(X_i - X_j)^2}{d}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

Where $\boldsymbol{\mathcal{S}}$ and $\mathbf{d}$ hyper parameters, we can write the prior distribution for the unknown link function .

$$\pi(g/,\beta,\sigma^2,\mathcal{S}) = det\,[E_n]^{-\frac{1}{2}} exp\left\{-\frac{g_n' E_n g_n}{2}\right\} \ldots\ldots\ldots\ldots\ldots\ldots(11)$$

$\boldsymbol{E_n}$ with dimension (n×n) is denoted the covariance matrix and given:-

$$E\left(X_i\beta, X_j'\beta\right) = \mathcal{S} exp\left\{-\left(X_i - X_j\right)'\beta\beta'(X_i - X_j)\right\} \ldots\ldots\ldots\ldots\ldots\ldots(12)$$

Gramacy and Lion ,2012 mention that the identifiable can be satisfy $\frac{\beta}{\sqrt{d}}$ with necessity for the constraint $\|\beta\| = 1$. When use Gaussian process as prior for non parametric link function , so that (β) will be use instead of $\frac{\beta}{\sqrt{d}}$ in the covariance function .

$$E\left(X_i'\beta, X_j'\beta\right) = \mathcal{S}\ exp\left\{-\left(X'_i\beta - X'_j\beta\right)^2\right\} \ldots\ldots\ldots\ldots\ldots\ldots(13)$$

Inverse Gamma will be set as a prior distribution , $\boldsymbol{\mathcal{S}} \sim IG(a_{\mathcal{S}}, b_{\mathcal{S}})$, where $a_{\mathcal{S}}$ and $b_{\mathcal{S}}$ are the hyper parameters.

Laplace distribution put as a prior distribution for the parameters index see (Tibshirani,1996. Park &casella ,2008):-

$$\pi(\beta/\sigma^2) = \prod_{j=1}^{p} \frac{\gamma}{2(\sigma^2)^{\frac{1}{2}}} \exp\left\{-\frac{\gamma}{\sqrt{\sigma^2}}\|\beta_j\|\right\} \ldots\ldots\ldots\ldots\ldots\ldots(14)$$

Where $(\gamma > 0)$ is penalty parameter .

In the paper the researchers suggested , new scale mixture of normal mixing with Rayleigh density (Flaih et al ,2020) as in equation (4).

3. Scale mixture Rayleigh distribution (SMRD)

Follow (Flaih et al,2020) new scale mixture of normal distribution mixing Rayleigh density for Laplace distribution we can rewrite the prior distribution of the index parameter as follow .Let $\boldsymbol{\gamma} = \frac{\sqrt{\boldsymbol{\sigma^2}}}{\boldsymbol{\lambda}}$

$$\pi(\beta/\gamma) = \prod \int_0^\infty \frac{1}{2\gamma} \exp\left\{-\frac{|\beta_j|}{\gamma}\right\} = \prod_{j=1}^{p} \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} \exp\left\{-\frac{\beta_j^2}{2s_j}\right\} . \frac{s_j}{\gamma} \exp\left\{-\frac{s_j}{2\gamma}\right\} ds_j \ldots\ldots\ldots(15)$$

So that hierarchical model for single index based on (Flaih et al,2020) proposed method can be formed as follows :-

$$y/\beta, \mathbf{g}, \sigma^2 \sim N(\mathbf{g}(X_i'\boldsymbol{\beta}), \sigma^2)$$
$$\mathbf{g}/X, \beta,\ \mathcal{S} \sim GP(0, E(.,.))$$
$$\beta, s_j \big/ \gamma \sim \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi s_j}} \exp\left\{-\frac{\beta_j^2}{2s_j}\right\} . \frac{s_j}{\gamma} \exp\left\{-\frac{s_j}{2\gamma}\right\}$$
$$\boldsymbol{\mathcal{S}} \sim IG\ (a_{\mathcal{S}}, b_{\mathcal{S}})$$
$$\gamma \sim Ga\ (\ a_\gamma, b_\gamma)$$
$$\sigma^2 \sim Ga(a_{\sigma^2}, b_{\sigma^2}) \ldots\ldots\ldots\ldots\ldots\ldots(16)$$

Where $a_{\mathcal{S}}, b_{\mathcal{S}},\ a_\gamma, b_\gamma, a_{\sigma^2}, b_{\sigma^2}$ hyper parameters

4. The full conditional posterior distribution

For all the parameter posterior can be given back of as follows :-

$$P(\mathbf{g_n}, \boldsymbol{\beta}, \boldsymbol{s}, \boldsymbol{\gamma}, \boldsymbol{\mathcal{S}}, \boldsymbol{\sigma^2}, \boldsymbol{y}) \propto \left\{det[D]^{-1} exp\left\{-\frac{(y-\mathbf{g_n})'\boldsymbol{D^{-1}}(y-\mathbf{g_n})}{\mathbf{2}}\right\}\right\} \times\ [\boldsymbol{E_n}]' \boldsymbol{exp}\ \left\{-\frac{\mathbf{g}_n' E_n \mathbf{g}_n}{\mathbf{2}}\right\} \times \prod_{j=1}^{p} \int_0^\infty \frac{1}{\sqrt{2\pi s_j}} exp\left\{-\frac{\beta_j^2}{2s_j}\right\} \times \frac{s_j}{\gamma} exp\left\{-\frac{s_j}{2\gamma}\right\} \times \prod_{i=1}^{n} \frac{1}{\sigma^2} exp\left(-\frac{s_j}{\sigma^2}\right) \times \prod_{j=1}^{p} (\gamma)^{a_\gamma - 1} \exp\left(-b_\gamma \gamma\right) \times (\boldsymbol{\mathcal{S}})^{-a_{\mathcal{S}}-1} \times \exp\left(-\frac{b_{\mathcal{S}}}{\boldsymbol{\mathcal{S}}}\right) \ldots(17)$$

As for the Gibbs sampler the full conditional distribution of (β) is a normal distribution and those of (σ^2) and $(\boldsymbol{\gamma})$are Gamma distribution and full conditional distribution of $(\boldsymbol{\mathcal{S}})$ and $(\ s_j)$ are generalized inverse Gaussian distribution (Jorgensen, 2012)

4.1 The full conditional distribution for sampling $\mathbf{g}_n$

$$\pi(\mathbf{g}_n, \beta, \boldsymbol{S}, \sigma^2, s, y) \propto \pi(y/\mathbf{g}_n, \beta, \sigma^2) \times \pi(\mathbf{g}_n/\boldsymbol{\beta}, \mathcal{S})$$

$$\propto [\det(D)]^{-\frac{1}{2}} exp\left\{-\frac{(y-\mathrm{g}_n)' D^{-1}(y-\mathrm{g}_n)}{2}\right\} \times \det(E_n)^{-\frac{1}{2}} exp\left\{-\frac{\mathrm{g}_n' E_n \mathrm{g}_n}{2}\right\}$$

The full conditional distribution for g_n is normal distribution with mean

$A = D(D+E_n)^{-1}y$, and variance $B = D(D+E_n)^{-1}E_n$

4.2 The full conditional distribution for sampling β:

$$\pi(\beta/\mathrm{g}_n, \mathcal{S}, s, \gamma, \sigma^2, y) \propto \pi(y/\mathrm{g}_n, \beta, \sigma^2) \times \pi(\mathrm{g}_n/\beta, \mathcal{S}) \times \pi(\beta/s)$$

$$\propto exp\left\{-\frac{(\boldsymbol{y}-\mathbf{g}_n)'\boldsymbol{D}^{-1}(\boldsymbol{y}-\mathbf{g}_n)}{\mathbf{2}}\right\}[\det(D)]^{-\frac{1}{2}} \times \det(\boldsymbol{E}_n)^{-\frac{1}{2}} exp\left\{-\frac{\mathbf{g}_n'\boldsymbol{E}_n\mathbf{g}_n}{\mathbf{2}}\right\} \times \prod_{j=1}^{p} \boldsymbol{exp}\left\{-\frac{\beta_j^2}{2s_j}\right\}$$

Metropolis algorithm will be considered to sample β.

4.3 The full conditional distribution for sampling $\boldsymbol{\mathcal{S}}$:

$$\pi(\mathcal{S}/\mathrm{g}_n, \beta, s, \gamma, \sigma^2, y) \propto \pi(y/\mathrm{g}_n, \beta, \sigma^2) \times \pi(\mathrm{g}_n/\beta, \mathcal{S}) \times \pi(\mathcal{S})$$

$$\propto exp\left\{-\frac{(\boldsymbol{y}-\mathbf{g}_n)'\boldsymbol{D}^{-1}(\boldsymbol{y}-\mathbf{g}_n)}{\mathbf{2}}\right\}[\det(D)]^{-\frac{1}{2}} \times \det(\boldsymbol{E}_n)^{-\frac{1}{2}} exp\left\{-\frac{\mathbf{g}_n'\boldsymbol{E}_n\mathbf{g}_n}{\mathbf{2}}\right\} \times \left(\frac{\mathbf{1}}{\boldsymbol{\mathcal{S}}}\right)^{\boldsymbol{a}_{\mathcal{S}}+\mathbf{1}} \boldsymbol{exp}\left\{-\frac{\boldsymbol{b}_{\mathcal{S}}}{\boldsymbol{\mathcal{S}}}\right\}$$

Metropolis algorithm will be considered to sample $\boldsymbol{\mathcal{S}}$.

4.4 The full conditional distribution for sampling $\boldsymbol{s_j}$

$$\pi(s_j/\mathrm{g}_n, \beta, \sigma^2, \gamma, \mathcal{S}, y) \propto \pi(\beta/s_j) \times \pi(s_j)$$

$$\propto \frac{1}{\sqrt{2\pi s^2}} exp\left\{-\frac{\beta_j^2}{2{s_j}^2}\right\}\frac{s_j}{\gamma} exp\left\{-\frac{{s_j}^2}{2\gamma}\right\}$$

$$\propto exp\left\{-\frac{1}{2}\left({\beta_j}^2({s^2}_j)^{-1} + \frac{1}{\gamma}{s^2}_j\right)\right\} \times \frac{\mathbf{1}}{\boldsymbol{\gamma}\sqrt{\mathbf{2\pi s^2}}}$$

The full conditional distribution is Generalized Inverse Gaussian (GIG) distribution

4.5 The full conditional distribution for sampling $\boldsymbol{\gamma}$

$$\pi(\gamma/\mathrm{g}_n, \beta, \gamma, \mathcal{S}, \sigma^2, y) \propto \pi(s/\gamma) \times \pi(\gamma)$$

$$\propto \prod_{j=1}^{p} \frac{\mathrm{s_j}}{\upgamma} \exp\left\{-\frac{{\mathrm{s_j}}^2}{2\upgamma}\right\} \times \upgamma^{\mathrm{a}_\upgamma - 1}\exp\{-\mathrm{b}_\upgamma \upgamma\}$$

$$\propto \prod_{j=1}^{p} \frac{\mathrm{s_j}}{\upgamma^{2-\mathrm{a}_\upgamma}} \exp\left\{-\left(\frac{{\mathrm{s_j}}^2}{2\upgamma} - \mathrm{b}_\upgamma\upgamma\right)\right\}$$

Metropolis algorithm used to sample γ.

4.6 The full conditional distribution for sampling $\boldsymbol{\sigma^2}$

$$\pi((\sigma^2/\mathrm{g}_n, \beta, \gamma, \mathcal{S}, s, y) \propto \pi(y/\mathrm{g}_n, \beta, \sigma^2) \times \pi(\sigma^2)$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} exp\left\{-\frac{1}{2\sigma^2}(y_i - \mathbf{g}_n)^2\right\} \times \left(\frac{1}{\sigma^2}\right)^{a_{\sigma^2}+1} exp\left\{\frac{b_{\sigma^2}}{\sigma^2}\right\}$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+a_{\sigma^2}+1} \exp\left\{-\frac{1}{\sigma^2}\left[\frac{1}{2}(y_i - \mathbf{g}_n)^2 + b_{\sigma^2}\right]\right\}$$

The conditional distribution for σ^2 is Inverse Gamma distribution .

5. Simulation study

Simulation approach are used to investigating the performance of our proposed method (BSIReg) compared with Bayesian single index quantile regression with quantile τ=0.50

($BQSIM_{\tau=0.50}$) and non-Bayesian single index (SIReg) methods. Two simulation examples are considered, for each example, random errors are generated from exponential distribution with shape parameter (0.5). For each simulation examples, we run one hundred replications. All methods under study are investigated based on standard division of the parameter estimates and the median of mean absolute deviations denoted as MMAD, and mean absolute error referred to as "MAE".

5.1 Example One

In this simulation example , We demonstrate how the proposed approach performed for true model that used to simulate the data:

$$y = g(x_i^T \beta) + \epsilon \quad where\ g(t) = 5\cos(t) + \exp(-t^2)$$

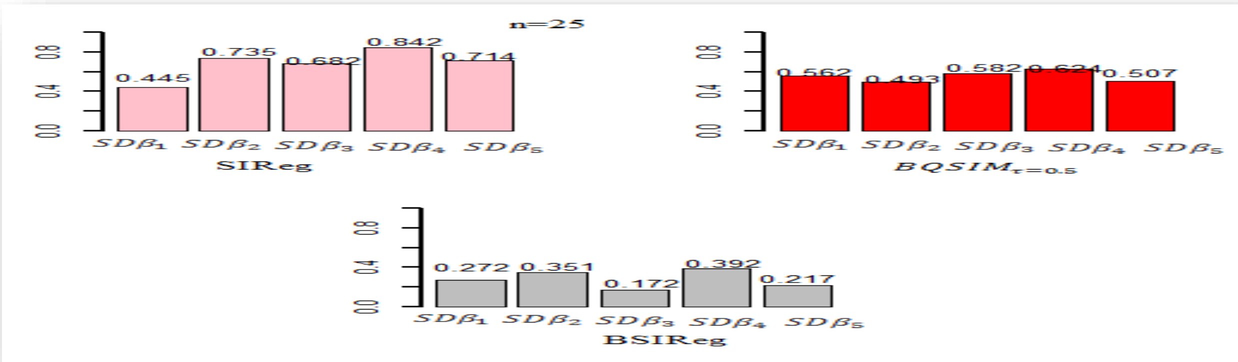
where the explanatory variables $x_i, (i = 1,2,3,4,5,)$ are identical independent distribution are simulated from normal distributions with mean 0 and variance $(0.25)^2$, $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = \frac{1}{\sqrt{5}}(2,0,1,0,0)$. In this simulation , we use different sample sizes (25,50,100,150,200 and 250). The results for simulation listed in Table 1 , and Figure 1

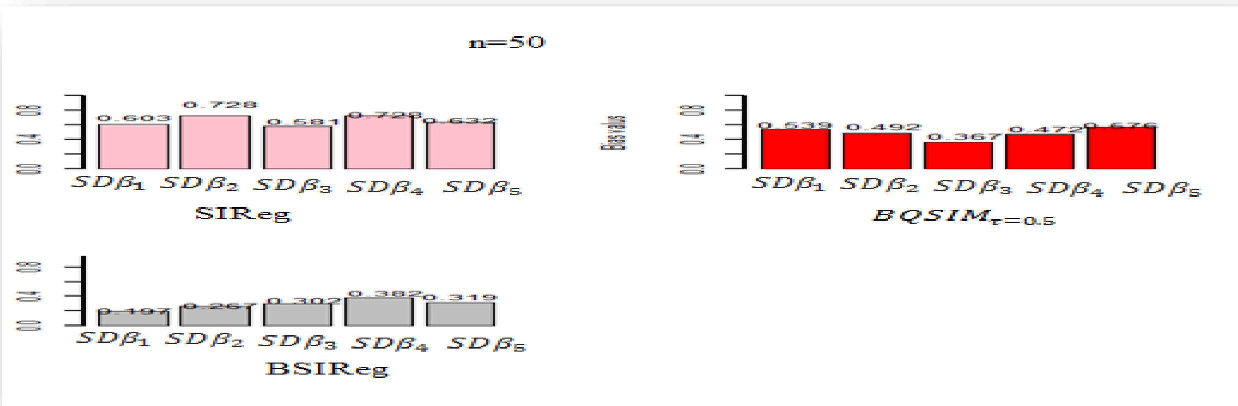
Table (1): comparison SD of the parameter estimates for SIReg, $BQSIM_{\tau=0.50}$ and BSIReg based on 100 replications for Simulated 1.

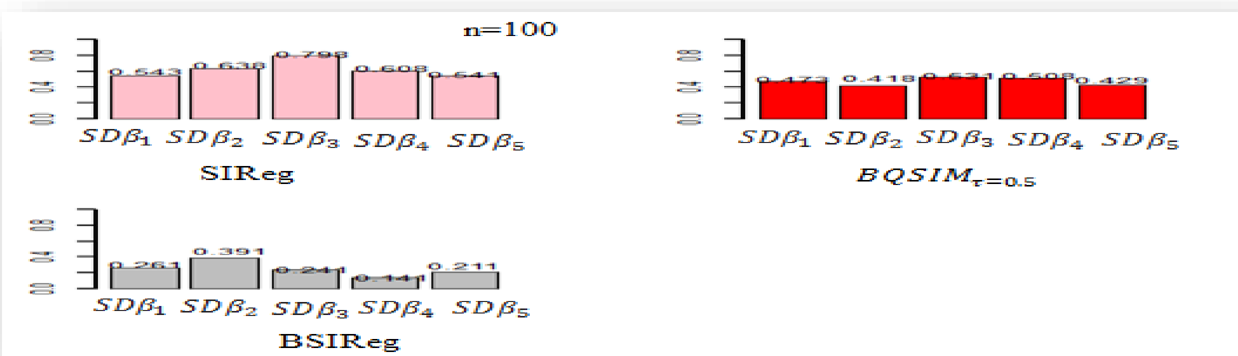
N	Methods	$SD\beta_1$	$SD\beta_2$	$SD\beta_3$	$SD\beta_4$	$SD\beta_5$
N=25	SIReg	.0445	.0735	.0682	.0842	.0714
	$BQSIM_{\tau=0.50}$	.0562	.0493	.0582	.0624	.0507
	BSIReg	**.0272**	**.0351**	**.0172**	**.0392**	**.0217**
N=50	SIReg	.0603	.0728	.0581	.0728	.0632
	$BQSIM_{\tau=0.50}$	.0539	.0492	.0367	.0472	.0576
	BSIReg	**.0197**	**.0267**	**.0302**	**.0382**	**.0319**
N=100	SIReg	0.543	0.638	0.798	0.608	0.541
	$BQSIM_{\tau=0.50}$	0.473	0.418	0.531	0.508	0.429
	BSIReg	**0.261**	**0.391**	**0.241**	**0.141**	**0.211**
N=150	SIReg	0.618	0.531	0.481	0.592	0.445
	$BQSIM_{\tau=0.50}$	0.677	0.505	0.537	0.675	0.472
	BSIReg	**0.132**	**0.226**	**0.126**	**0.111**	**0.104**
N=200	SIReg	0.627	0.538	0.439	0.475	0.734
	$BQSIM_{\tau=0.50}$	0.527	0.519	0.407	0.568	0.636
	BSIReg	**0.240**	**0.184**	**0.118**	**0.294**	**0.184**
N=250	SIReg	0.429	0.492	0.379	0.477	0.423
	$BQSIM_{\tau=0.50}$	0.395	0.311	0.301	0.292	0.246
	BSIReg	0.119	0.128	0.096	0.078	0.098

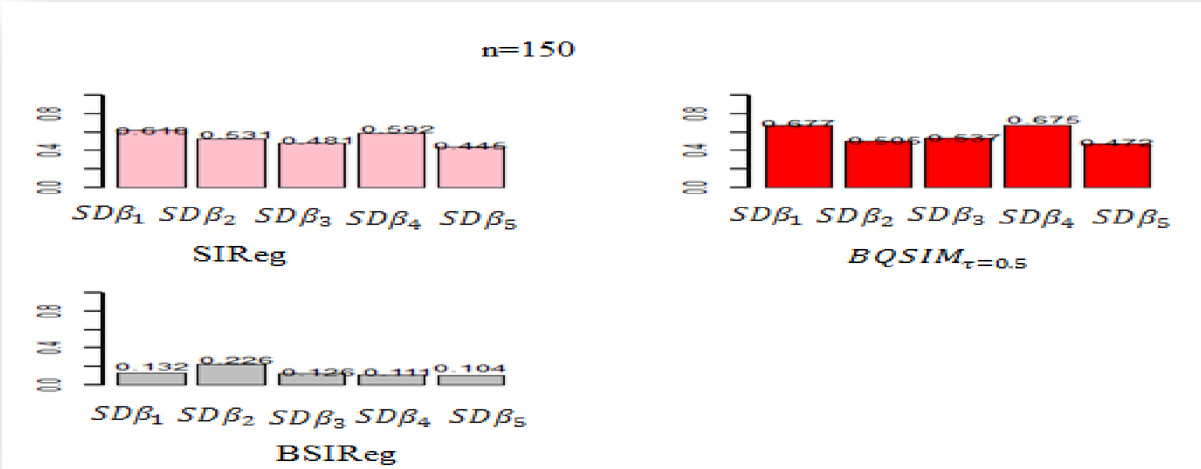
The results listed in Table 1 present the standard deviation (SD) of the parameters estimates for three methods under study. We can see clearly the SD values for our proposed method much smaller than that values for the other two methods. Therefore, our proposed method is better than the other methods. Figure 1 shows standard deviation for these three methods via different samples size.

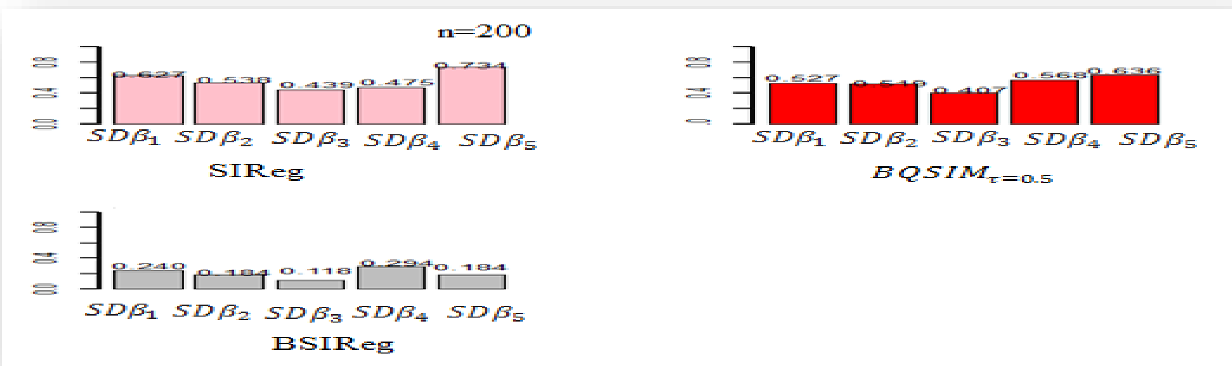
Figure (1) display standard deviation of parameters estimation for the samples size (0.25,50,100,150,200 and 250)

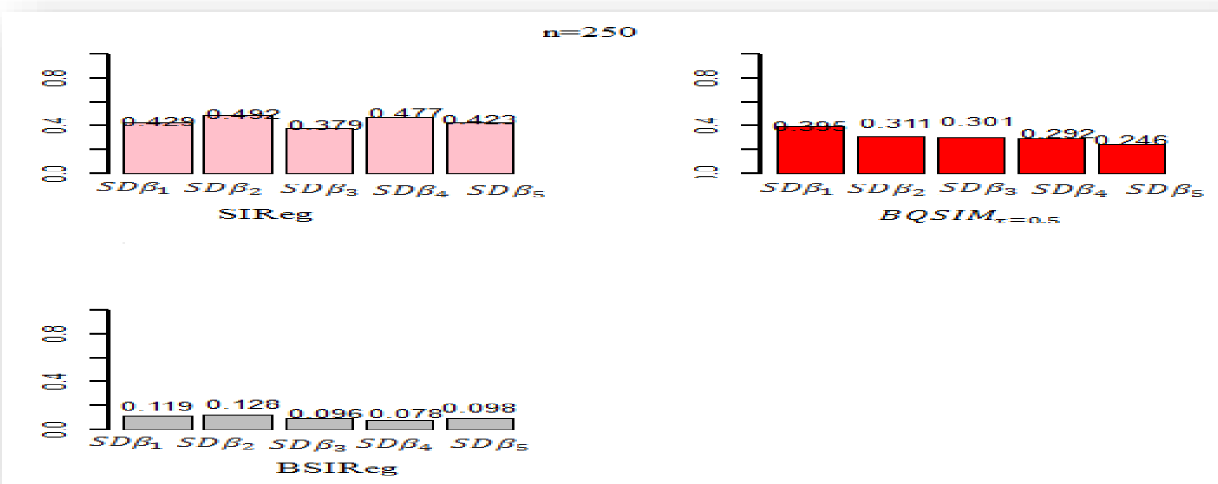







n=150
SIReg
BQSIM$_{\tau=0.5}$
BSIReg
0.132
0.226
0.104
0.675
0.531
0.481
0.592


n=200
SIReg
BQSIM$_{\tau=0.5}$
BSIReg
0.538
0.439
0.475
0.118
0.184


n=250
SIReg
BQSIM$_{\tau=0.5}$
BSIReg
0.429
0.492
0.379
0.477
0.423
0.311
0.301
0.292
0.246
0.119
0.128
0.096
0.078
0.098

5.2 Example Two

In this example, we generated the data from the following model

$$y = g(x_i^T \beta) + 0.1\epsilon \quad , \quad g(t) = sin\left\{\frac{\pi(t-D)}{(E-D)}\right\}$$

where y is the dependent variable , x_i, $i = 1,2,3,4,5,6,7,8,9,10$ are the independent variables, these variables are simulated from uniform distributed with interval $[0,1]$, and $D = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $E = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$. The parameters vector $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9 \beta_{10})$ $= \frac{1}{\sqrt{3}}(1,1,0,0,1,0,0,0,0,0)$. The performance of these three methods are illustrate via "MMAD" "MAE".

Table -2- MMADs and MAE for simulation 2, The results are averaged over 100 simulations.

N	Methods	MAE	MMAD
N=25	SIReg	0.536 (0.397)	0.541 (.0384)
	$BQSIM_{\tau=0.50}$	0.417 (.0278)	0.391 (.0217)
	BSIReg	**0.185 (.0087)**	**0.157 (.0079)**
N=50	SIReg	0.527 (.0383)	0.543 (.0385)
	$BQSIM_{\tau=0.50}$	0.374 (.0185)	0.343 (.0186)
	BSIReg	**0.183 (.0064)**	**0.156 (.0084)**
N=100	SIReg	0.569 (.0383)	0.529 (.0255)
	$BQSIM_{\tau=0.50}$	0.329 (.0215)	0.330 (.0227)
	BSIReg	**0.163 (.0087)**	**0.142 (.0074)**
N=150	SIReg	0.438(.0271)	0.426 (.0252)
	$BQSIM_{\tau=0.50}$	0.392 (0.237)	0.384 (.0216)
	BSIReg	**0.126 (.0077)**	**0.139 (.0037)**
N=200	SIReg	0.413 (0.296)	0.428 (.0268)
	$BQSIM_{\tau=0.50}$	0.276 (.0145)	0.431 (.0284)
	BSIReg	**0.137 (0.057)**	**0.122 (.0049)**
N=250	SIReg	0.373 (.0235)	0.328 (.0283)
	$BQSIM_{\tau=0.50}$	0.284 (.0162)	0.276 (.0134)
	BSIReg	**0.110 (.0056)**	**0.112 (.0061)**

Note: In the parentheses are standard deviation (S.D)

From results of MMADs and MAE that listed in Table 2, It is readily observed that for all samples size under considerations, the proposed method (BSIReg) achieves the smallest values of MMADs and MAE ,which shows that (BSIReg) estimates more accurately. Also, the standard deviation (SD) computed for the coefficients of our proposed method (BSIReg) is much smaller than the (SD) values that computed for the coefficients that estimated by the other two methods for all samples size under consideration. Therefore, our proposed method (BSIReg) is more accurately compared with other methods.

6. Real Dataset

In this section, we used Air Pollution dataset which belong to Public Roads Administration in Norway. We used these dataset to investigated the performance the methods under considerations. These dataset consist from one dependent variables and seven independent variables, The dependent variable is represented log (concentration of NO2per hour), and the seven independent variables are: x_1 represented (log (number of cars per hour)), x_2 represented temperature , x_3 represented wind speed in meters per second, x_4 represented the temperature difference , x_5 represented wind direction, x_6 represented time of day in hours and x_7 represented day number.

Similar to simulation section, we compare three methods (SIReg)and (BQSIMτ=0.50) and our proposed method (BSIReg). These methods under considerations are investigated based on the median of mean absolute deviations (MMAD) , and mean squared error (MSE). The results are listed in the table below.

Table (3): MMAD and MSE for Air Pollution dataset

Methods	MMAD	MSE
SIReg	1.518(1.273)	0.981(0.764)
$BQSIM_{\tau=0.50}$	1.382(1.086)	0.859(0.697)
BSIReg	0.872(0.648)	0.792(0.577)

Note: In the parentheses are standard deviation (S.D)

From the results of MMAD , MSE and SD are listed in Table (3), we can see that our proposed method (BSIReg) tends to get lower values of MMAD , MSE and SD compare to the other two methods. Therefore, based on these result reported in this table we can conclude that the performance of the proposed method is better than others. We can see coefficients estimation of air Pollution dataset via our proposed method and other two method as following:

Table (4): Parameters estimates of air Pollution dataset

Independent variables	Parameters	SIReg	$BQSIM_{\tau=0.50}$	BSIReg
x_1	β_1	0.261	0.341	0.513
x_2	β_2	0.471	0.136	0.008
x_3	β_3	0.006	0.004	0.004
x_4	β_4	-0.241	-0.175	-0.023

x_5	β_5	0.156	0.251	0.734
x_6	β_6	0.007	0.002	-0.005
x_7	β_7	-0.035	0.004	0.0009

In general, some of the parameter estimates by our proposed method is very closed from zero compared with other two methods. The independent variable (x_4) temperature difference has negative affecting through three methods. But independent variables (x_1,x_2,x_3 and x_5) have positive affecting through three methods.

7. Conclusion

In this paper, we proposed new estimation and variable selection method through Bayesian approach for single index model. We have construct new hierarchical model based on the representation of scale mixture of normal distribution mixing Rayleigh density for the double exponential prior density of the parameters vector. This model shrinkage the dimensionality while retention nonparametric flexibility.

Two simulation examples and real data are considered to evaluation the performance of the proposed method compare to the other two existing methods. The results that we reported in simulation and real data tables this paper demonstrated the superiority of our proposed method compare to the other competitor methods.

Reference

Jorgensen, B. (2012)." Statistical properties of the generalized inverse Gaussian distribution", (Vol. 9). Springer Science & Business Media.

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

Flaih, A. N., Alshaybawee, T., & Alhusseini, F. H. H. (2020). "Sparsity via new Bayesian Lasso". Periodicals of Engineering and Natural Sciences, 8(1), 345-359.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). "Least angle regression". The Annals of Statistics 2004, Vol. 32, No. 2, 407–499.

Korobilis, D. (2021). "High-dimensional macroeconomic forecasting using message passing algorithms". Journal of Business & Economic Statistics, 39(2), 493-504.

Park, T., & Casella, G. (2008). "The bayesian lasso". Journal of the American Statistical Association, 103(482), 681-686.

Tiao, G. C., & Box, G. E. ,1973,"Some comments on “Bayes” estimators". The American Statistician, 27(1), 12-14.

Ichimura, H. (1993). "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models". Journal of econometrics, 58(1-2), 71-120.

Choi, T., Shi, J. Q., & Wang, B. (2011). "A Gaussian process regression approach to a single-index model". Journal of Nonparametric Statistics, 23(1), 21-36.

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