Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Bayesian lasso principal component regression with an application

Mohammed H. AL-Sharoot [1], Wael abd Al Hussein Kazem[2], Shatha Awwad Al-Fatlawi[3]

[1] College of Administration and Economics, Al-Qadisiyah University, Iraq. mohammed.lsharood@qu.edu.iq

[2] Babylon Education Directorate, Ministry of Education, Iraq . wael.a@s.uokerbala.edu.iq

[3] *College of Administration and Economics Al-Qadisiyah University, Iraq. shatha.awwad@qu.edu.iq* .

ARTICLE INFO

Article history:
Received: 17 /12/2023
Rrevised form: 27 /12/2023
Accepted : 29 /12/2023
Available online: 30 /12/2023

Keywords:

LASSO, Principal Component regression ,MCMC, Prior distribution .

ABSTRACT

The topic of variable selection in constructing statistical models has received wide attention in many applications, including medical, economic, engineering, and various other fields. Since, this method helps researchers save time and effort, by concentrating on crucial variables and eliminating unimportant variables from the statistical models under study. The famous variable selection technique known as the Least absolute Shrinkage and Selection Operator (LASSO) method works by reducing the number of variables with high explanatory power. This study will combine two methods for variable shrinkage, which minimize both simultaneously by removing variables with little effect and reducing the variance and bias of the estimated parameters. Two techniques were used to validate the results of our current study. The first approach uses simulation, where data is produced using predetermined hypotheses and models. The second approach uses data from actual sources .

MSC..

https:// 10.29304/jqcsm.2023.15.41357

1. Introduction

The statistical literature has long included methods, for variable selection, and model selection. Such as Akaike information criterion (AIC) (Akaike, 1973) and Bayesian information criterion(BIC)(Schwarz, 1978). But these techniques are regarded as classical approaches, and it frequently takes a long time to finish the process of selecting the best models. This is because the number of models that can be obtained is proportional to the number of independent variables included in the model. The potential number of statistical models is positively relationship with number of independent variables (2^p), p is number of independent variables(Murtaza Haider,2014). Modern approaches to variable and model selection, on the other hand, have developed swiftly. These approaches are distinguished by their ability to choose significant variables and eliminate unimportant variables more quickly. These techniques, which usually take place in a very short amount of time, estimate the parameters and choose variables all at once. Such as (LASSO) (Tibshirani, 1996), (SCAD) (Fan and Li, 2001), LARS (Efron et al., 2004), adaptive Lasso (Zou, 2006) , Bayesian Lasso (Park and Casella, 2008) and Bayesian adaptive Lasso and iterative adaptive Lasso (Sun et al., 2010) etc. The aforementioned strategies are all up-to-date techniques that simultaneously choose variables and estimate parameters, saving time and reducing the likelihood of estimation errors. These techniques seek to lessen

∗ Mohammed H. AL-Sharoot

Email addresses: *mohammed.lsharood@qu.edu.iq*

Communicated by 'sub etitor'

the inflated variance of the parameters linked to standard regression models; however, on occasion, these models experience multicollinearity because of a perfect or semi-perfect linear relationship between two or more of the independent variables(Frost, J. (2017)). The estimation techniques mentioned above can still be used even in the case of multicollinearity, but they lose their effectiveness in situations where the independent variables exhibit significant correlation. It is commonly known that principal component regression is a two-step process that can be used to address the problem of multicollinearity. In the first stage, the accompanying parameters for the principal components are estimated, and in the second stage, the landmarks of the original model are estimated through the inverse feedback of the estimators of the principal component parameters (Alabi,O,KAyinde,2008). In this paper, we propose a new methodology that combines the LASSO method with principal component regression. To successfully handle both the variable selection process and the multicollinearity issue, making sure both are taken care of. When the number of variables in the model is greater than the sample size, the suggested method can also estimate regression parameters effectively(n<p) . The benefits of our proposed method include reduced variance inflation of model parameters and reduced bias. This paper is divided into. In the second section, the focus will be on combining principal component regression with Lasso. In the third section, the emphasis will be on the Gibbs Sampler technique specific to our studied model. The fourth section will concentrate on using simulation methods to compare the performance of our proposed approach to other methods, In the fifth section, we also examine the performance of our proposed approach with other methods using real-world data, In the sixth section, we provide concise conclusions regarding our proposed approach.

1.1. Lasso principal component regression

The regression model is considered a statistical tool that focuses on estimating the relationship between a single dependent variable and one or more explanatory variables. This relationship is built using a regression model, as indicated by the following general formula.

$$y=X\beta+\varepsilon \qquad (1)$$

where y is dependent variable and $\underline{X}=(x_1,x_2,x_3\ldots x_P)$ is The matrix of explanatory variables (N*P) assumes that there is a certain level of correlation between two or more of these explanatory variables. ε It is the degree of random error that conforms to a normal distribution with a a mean of 0 and variance σ^2. One of the proposed methods to address multicollinearity is the principal component regression method. The principal component regression method focuses on transforming the original explanatory variables, which may have correlations, without omitting any variable. It transforms them into linear combinations called principal components. These principal components provide integrated information about the observations of the original explanatory variables. The original explanatory variables is $(x_1,x_2,x_3\ldots x_P)$ a linear combination of these variables can be formed for each principal component, as shown in the following equation.

$$F=XD \qquad (2)$$

where F_((n×p) matrix of principal component ,and D It is an orthogonal matrix of eigenvectors corresponding to the eigenvalues of the matrix, X^Tr X this matrix has rank P×P. The matrix D is the diagonal matrix , where ($DD^{Tr} = D^{Tr}D = I_P$) where,$FD^{Tr} = XDD^{Tr}$, $DD^{Tr} = I_P$ therefore $X = FD^{Tr}$.

By using dependent variable y, By considering the principal components as independent variables, we obtain the principal component regression model.

$$y = FD^{Tr}\beta + \varepsilon \ldots\ldots\ldots\ldots (3)$$

Let $D^{Tr}\beta = \gamma$: The principal component regression model takes the following form:

$$y=F\gamma+\varepsilon \qquad (4)$$

Therefore, the estimation of Lasso principal component regression can be defined according to the following formula:

$$\hat{\gamma}_{LPCR} = \|y - F\gamma\|^2 + \lambda\|\gamma\| \qquad (5)$$

where λ is shrinkage parameter, $\lambda \geq 0$. Equation (5) is not differentiable at zero, and γ is parameters of principal component regression, the estimation theses parameter by liner programming of But, in current paper, we will used

the Bayesian approach which is a useful technique for estimating the parameters because of a number of its advantages.

1.2. Bayesian Principal Component Regression with Lasso Approach

Using a Bayesian approach, the Laplace distribution must be used as a prior distribution in order to estimate the parameters of the model given in quation (5). To guarantee the procedure for choosing variables ,(Tibshirani, 1996) :

$$p\{\gamma_j|\lambda\} = \frac{\lambda}{2} e^{-\lambda|\gamma_j|} \qquad (6)$$

where λ is ship parameter and γ_j is scale parameter

Directly using the Laplace distribution as described in Equation (6) is considered a very challenging problem. Therefore, most researchers working in the field of variable selection opt for alternative formula (You and Hong 2012)(Alhamzawi 2013).

The alternative formula of Laplace distribution that proposed by(Andrews and Mallows 1974), which assumes that the Laplace distribution is Scale Mixture Normal. According to the alternative formula , the Laplace distribution can be expressed as a mixture distribution, where the first part is γ_j distributed according to normal distribution with mean 0 and variance s_j is a variance or estimated standard error, and second part is s_j which distributed as exponential distribution with a parameter λ_j.

$$p\{\gamma_j|s_1^2, s_2^2, \quad s_p^2, \sigma^2\} = \frac{1}{\sqrt{2\pi\sigma s_j}} e^{-\frac{\gamma_j^2}{2\sigma^2 s_j}}$$

where γ_j has Unknown variance s_j is a variance or estimated standard error is exponential distribution with ship parameter λ_j.

$$p\{s_j|\lambda_j\} = \frac{\lambda_j}{2} e^{\frac{\lambda_j s_j}{2}}$$

λ_j is the ship parameter responsible for controlling the degree of shrinkage towards zero, which are distributed according to *Gamma distribution*

$$p\{\lambda_j|\theta, k\} = \frac{\theta^k}{\Gamma(k)} \lambda_j^{k-1} e^{-\theta\lambda_j}$$

where θ,k are fix hyperparameters are takes very small values (θ=0.1,k=0.1),see (2010 (Li et al and (Hashem et al 2015).

The Bayesian hierarchical our proposed method given by.

$$\{y|\mu, F, \gamma, \sigma^2\} \sim N_P(\mu I_n + F\gamma, \sigma^2 I_n),$$

$$\{\gamma_j|\sigma^2, s_1 \ldots\ldots s_p\} \sim N_p(0_p, \sigma^2 D_s),$$

$$D_s = diag(s_1 \ldots\ldots s_p), \qquad (7)$$

$$(s_1 \ldots\ldots s_p) \sim Exp\left(\frac{\lambda_j}{2}\right),$$

$$\sigma^2 \sim \pi(\sigma^2)d\sigma^2, \sigma^2 \propto (\sigma^2)^{-a-1} e^{-b/\sigma^2}$$

$$p\{\lambda_j|\theta, k\} = \frac{\theta^k}{\Gamma(k)} \lambda_j^{k-1} e^{-\theta\lambda_j}$$

a,b and d are fix hyper parameters

From the above Bayesian hierarchical model ,shown in (7). The posterior distribution are :

- The full conditional distribution for γ is multivariate normal with $A^{-1}F^{T}y$,and variance $\sigma^{2}A^{-1}$. The value of A^{-1} is $A^{-1} = F^{T}F + D_{s}^{-1}$.

The full conditional distribution for σ^2 is inverse Gamma distribution as following

$$(\sigma^{2})^{-(n-1)(2-a-1)/2-p} \quad Exp\left\{\frac{-1}{\sigma^{2}}((y-F\,\gamma)^{T}(y-F\,\gamma))/2 + \gamma^{T}D_{s}^{-1}\gamma/2 + \text{b.}\right\}.$$

The Full conditional distribution for s_j, j=1,2……p is the inverse Gaussian with shape parameter $\sqrt{\frac{\lambda_j}{\gamma^2}}$

and scale parameter λ_j

The full conditional distribution for λ_j,j=1,2,……p is the is the gamma distribution with shape parameter θ+1 and scale parameter (k+s_j)/2

The distribution of the parameters of the original regression model can be obtained by utilizing the relationship between estimatorsβˆ and estimation of principal component regression.

$$D^{Tr}\beta = \hat{\gamma}$$

$(D^{Tr}\beta = \hat{\gamma}) * D$
wher
$DD^{Tr} = D^{Tr}D = I_P$
$\hat{\beta} = D\hat{\gamma}$

-

The full conditional distribution of βˆ is is multivariate normal with mean $DA^{-1}F^{T}y$ and variance $D^{2}\sigma^{2}A^{-1}$. Where $D^{2} = diag(\lambda_1^2, \dots \dots \lambda_p^2)$ .is $\lambda_1^2, \dots \dots \lambda_p^2$ is Egan value information matrix

Our proposed method will yield good posterior distributions. our algorithm will run for 12,000 iterations, discarding the first 2,000 iterations as burned in..

2. Simulation Approach

A simulation approach is required in order to more clearly demonstrate the superiority of our proposed method Bayesian lasso principal component (**BLPCReg)**) over other approaches in the context of principal component regression. By building an algorithm that is efficient and easy to use. Our proposed method will be compared with two other methods: The principal component regression(PCReg) method proposed by Alibuhtto et al. (2015), employing the PLS package, and the Bayesian principal component regression (BPCReg)method proposed by Junttila et al. (2017). Three sample size levels will be considered. The first is when the sample sizes are small (n = 10, n = 20), the second when the sample sizes are medium (n = 50, n = 60), and the third when the sample sizes are large (n = 100, n = 150). The independent variables are distributed normally and are highly correlated.

$corr\,(X_i, X_j) = (0.5)^{i+j} \qquad i = 1,2, \dots \dots n, \;\; j = 1,2, \dots . . p$

As for the model errors, they are generated according to four different types of errors: standard normal distribution ε∼N(0,1), normal distribution with mean 2 and variance 2 ε∼N(2,2), standard Laplace distribution with mean zero and variance 1 ε∼Lap(0,1)), and also a mixture Laplace distribution ε∼Lap(1,1)+Lap(2,2). Each simulation is

performed with different sample sizes, The first when the sample sizes are small (n=10, n=20), the second when the sample sizes are moderate (n=50, n=60), and the third when the sample sizes are large (n=200, n=250), All methods involved in the experiment are evaluated using the Median of Mean Absolute Deviation (MMAD) criterion, which is calculated using the following mathematical equation. $MMAD = \left|x_i^{tr}\hat{\beta} - x_i^{tr}\beta^{true}\right|$and the second is mean square error (MSE) In our current study, we will utilize two simulation methods.

2.1. First Simulation Approach

In the first simulation method, a 'very sparse' case will be used, where the true parameters take the following form β=(1,0,0,0,0,0,0), where the model takes the following form:

$$y = x_{i1} + \varepsilon_i$$

After conducting 100 simulations, (this simulation not iteration) , in our current study, the results are summarized in Table 1.

Table 1 presents the values of the MMAD (Median of Mean Absolute Deviation) and MSE (Mean Squared Error) in the very sparse case.

Sample Size		**Comparison Methods**	$e_i \sim N(0,1))$	$e_i \sim N(2,2)$	$e_i \sim Lap(1,1) + Lap(2,2)$	$e_i \sim Lap\ (1.0)$
		PCReg	1.7645 (.08603)	1.6582 (0.8334)	1.4289 (0.8664)	1.5403 (0.7820)
	n=10	BPCReg	1.5469 (.08562)	1.4532 (0.7405)	1.7302 (0.6941)	1.7832 (0.8403)
		BLPCReg	**1.0576 (0.4563)**	**1.2218 (0.5324)**	**1.1383 (0.2062)**	**1.3241 (0.4520)**
Small sample		PCReg	1.3857 (.05020)	1.782 (.08439)	1.8734 (.07430)	1.7882 (.06740)
	n=20	BPCReg	1.8452 (0.7834)	1.8926 (0.8934)	1.5620 (0.7648)	1.3979 (0.7835)
		BLPCReg	**1.0723 (0.2306)**	**1.2384 (0.2372)**	**1.0563 (0.3222)**	**1.1263 (0.2408)**
		PCReg	1.5493 (.07830)	1.7806 (.08432)	1.6078 (0.6704)	1.6458 (.06747)
	n=50	BPCReg	1.7805 (.05693)	1.5896 (.06840)	1.5944 (.06894)	1.6033 (.06739)
Meddle		**BLPCReg**	**1.2306 (0.3542)**	**1.3832 (0.2354)**	**1.3103 (0.3006)**	**1.1099 (0.5625)**
Sample		PCReg	1.6743 (0.6731)	1.6870 (0.7833)	1.6789 (0.6947)	1.8932 (0.8393)
	n=60	BPCReg	1.6739 (0.8939)	1.4931 (0.7841)	1.8790 (0.6741)	1.9452 (0.4582)
		BLPCReg	**1.1247 (0.5646)**	**1.5682 (0.1042)**	**1.3530 (0.2307)**	**1.2066 (0.3005)**
		PCReg	1.7694	1.8564	1.8656 (0.7685)	1.8657

			(0.6758)	(0.8768)		(0.5645)
	n=100	BPCReg	(1.7837) (0.7832)	(1.5682) (0.6792)	(1.7839)(0.8621)	(1.7839) (0.6792)
		BLPCReg	**1.3468 (0.3401)**	**1.2672 (0.2899)**	**1.4521 (0.2361)**	**1.3461 (0.452)**
		PCReg	1.6822 (0.332)	1.7839 (0.282)	1.8722 (0.6726)	1.6878 (0.5629)
		BPCReg	(1.4572)	(1.7800)	(1.8399)(0.7923)	(1.8923)
Large	n=150		(0.3896)	(0.6437)		(0.5892)
Sample		**BLPCReg**	**1.3462 (0.0252)**	**1.6562 (0.1012)**	**1.4521 (0.3067)**	**1.1220 (0.3405)**

Note : the values within parentheses represent (MSE)

The estimated values of MMAD and MSE utilizing our proposed method are substantially lower than the principal component regression (PCR) and Bayesian principal component regression methods, according to the data shown in the above table. Consequently, when compared to the other two methods, our suggested approach performs well in feature estimation and variable selection. We observe that our proposed method exhibited superior performance across different different sample sizes and different distribution of random error .

2. 2 Second Simulation Approach

in the second simulation method, dense (all the initial parameters are not zero)case will be used, where the true parameters take the following form $\beta = (0.85,0.85,0.85,0.85,0.85,0.85,0.85)$ (because dense case), where the model takes the following form:

$$y = 0.85x_{i1} + 0.85x_{i2} + + 0.85x_{i3} + +0.85x_{i4} + +0.85x_{i5} + +0.85x_{i6} + +0.85x_{i7} + \varepsilon_i$$

After conducting 100 simulations, in our current study, the results are summarized in Table 2.

Table 2 presents the values of the MMAD (Median of Mean Absolute Deviation) and MSE (Mean Squared Error) in the dense case.

Sample Size		Comparison Methods	$e_i \sim N(0,1))$	$e_i \sim N(2,2)$	$e_i \sim Lap(1,1) + Lap(2,2)$	$e_i \sim Lap$ (1.0)
		PCReg	1.4543 (.03460)	1.4547 (0.2401)	1.7340 (0.5634)	1.4645 (0.3420)
Small sample	n=10	BPCReg	1.2748 (.03416)	1.4532 (0.3426)	1.4533 (0.3068)	1.3461 (0.3405)
		BLPCReg	**1.0052 (0.0461)**	**1.1194 (0.0733)**	**1.3875 (0.0673)**	**1.2302 (0.1242)**

		PCReg	1.4531 (.00343)	1.5630 (.00532)	1.4530 (.01295)	1.4231 (.02305)
	n=20	BPCReg	1.1584 (0.4521)	1.6734 (0.5473)	1.2352 (0.6450)	1.5634 (0.4561)
		BLPCReg	**1.0067 (0.2306)**	**1.2343 (0.2372)**	**1.1834 (0.2406)**	**1.1273 (0.1855)**
		PCReg	1.4536 (.05434)	1.5634 (.06734)	1.8657 (0.4532)	1.67451 (.07404)
	n=50	BPCReg	1.5637 (.04530)	1.6732 (.05431)	1.4638 (.05604)	1.4583 (.05746)
		BLPCReg	**1.1547 (0.2267)**	**1.3832 (0.4537)**	**1.2404 (0.2064)**	**1.4530 (0.6574)**
		PCReg	1.5634 (0.2351)	1.4682 (0.2323)	1.5646 (0.2273)	1.5648 (0.3429)
Meddle Sample	n=60	BPCReg	1.4541 (0.45351)	1.5603 (0.3421)	1.3495 (0.3595)	1.9452 (0.4582)
		BLPCReg	**1.1247 (0.5646)**	**1.2682 (0.1042)**	**1.0533 (0.1183)**	**1.3486 (.00745)**
		PCReg	1.5364 (0.7632)	1.6732 (0.6892)	1.7923 (0.5620)	1.5637 (0.4793)
Large Sample	n=100	BPCReg	(1.5642) (0.5643)	(1.4585) (0.4394)	(1.6742)(0.5649)	(1.4535) (0.45734)
		BLPCReg	**1.2350 (0.2745)**	**1.3956 (0.1856)**	**1.3495 (0.2729)**	**1.2845 (0.3420)**
		PCReg	1.5649 (0.2345)	1.3432 (0.6573)	1.7869 (0.5674)	1.6751 (0.5643)
	n=150	BPCReg	(1.5643) (0.8767)	(1.6754) (0.5695)	(1.6750)(0.6573)	(1.5649) (0.2348)
		BLPCReg	**1.2363 (0.0453)**	**1.3460 (0.2320)**	**1.5640 (0.4632)**	**1.1267 (.01415)**

Note : the values within parentheses represent (MSE)

The estimated values of MMAD and MSE utilizing our proposed method (BLPCReg)are substantially lower than the principal component regression (PCR) and Bayesian principal component regression methods, according to the data

shown in the above table. Consequently, when compared to the other two methods (PCReg and BPCReg),, our suggested approach performs well in feature estimation and variable selection. We observe that our proposed method exhibited superior performance across different sample sizes and different distribution of random error .

After obtaining the results from the simulation method, we find that our proposed method performs well, and the behavior of the estimated features can be observed through Figures (1) and (2).

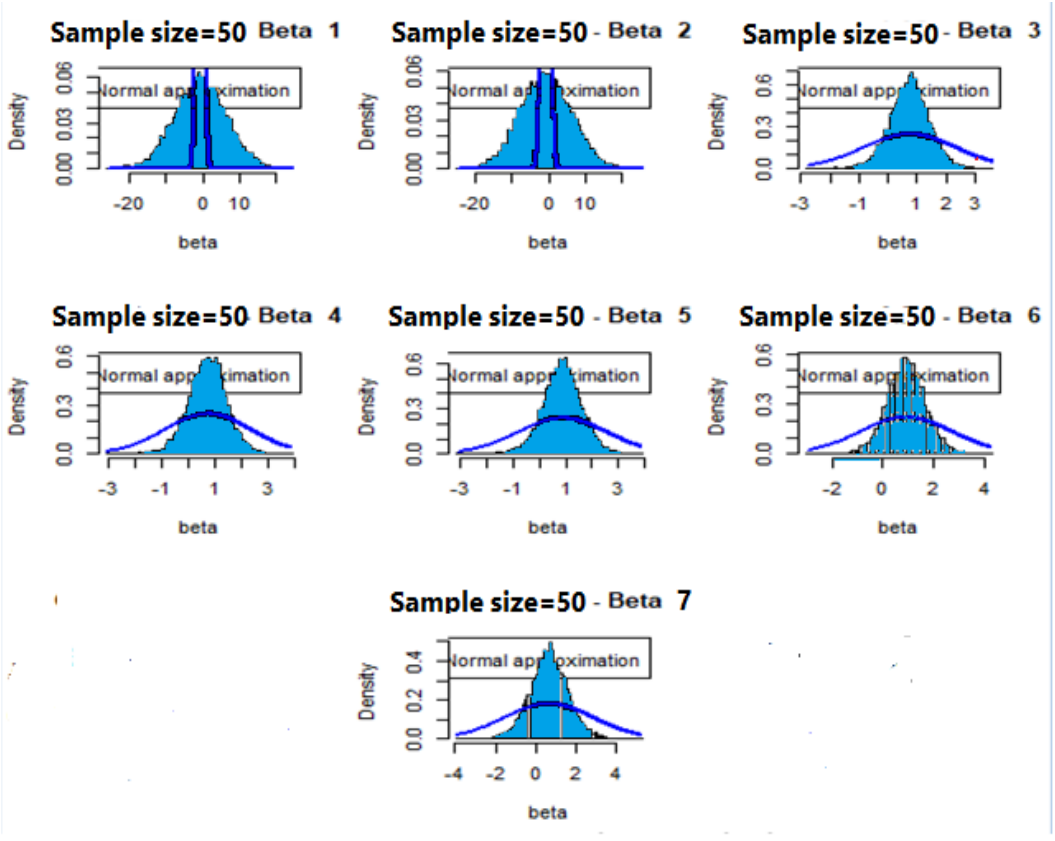

Figure (1) show the histograms of BLPCReg in density case and sample size is 50

From the above figure, which Is the shape is histogram of the estimated parameters in the density case at a sample size of 50, we can clearly observe that the estimated parameters follow a normal distribution and match the underlying distribution of the model's parameters perfectly. This figure demonstrates that the estimated parameters align perfectly with the true distribution of the model's parameters.

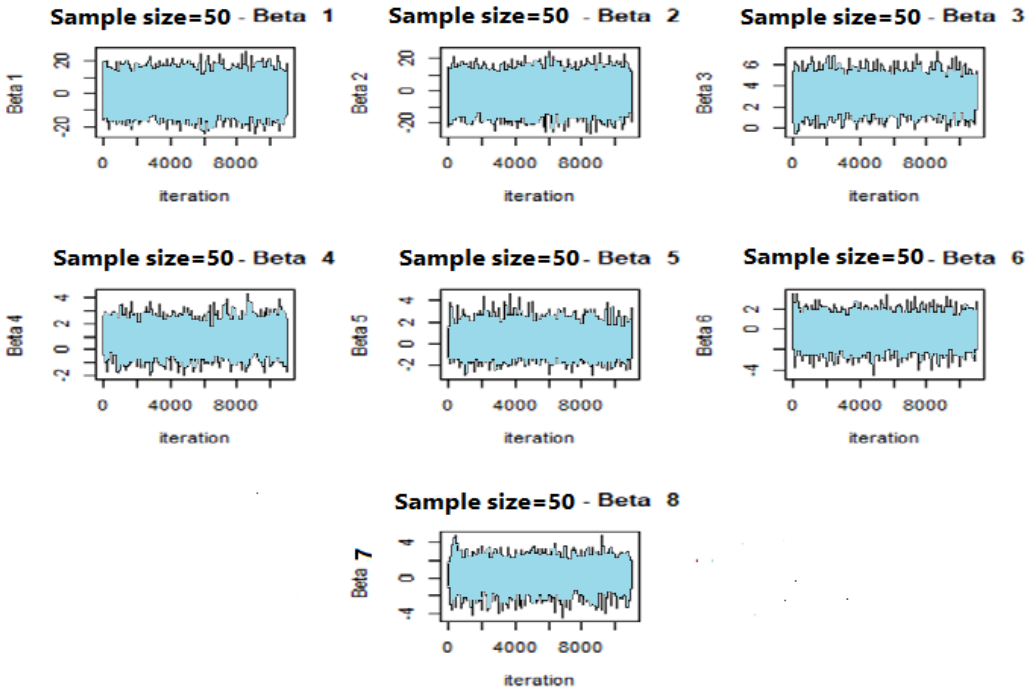

Figure (2) show the trace plot of BLPCReg in density case and sample size is 50

From the above figure, which represents trace plots of the estimated parameters in the density case at a sample size of 50, we can clearly observe that the estimated features exhibit convergence and stability across all iterations. This indicates that the MCMC samples for the posterior distribution of the estimated parameters are consistently stable throughout the iterations. This implies that our proposed algorithm is stable, easy to implement, and efficient.

3. Real Data

After demonstrating the efficiency of our proposed method through simulation compared to a set of methods in the same field, it is essential to test the effectiveness of our proposed method with real-world data. Medical data was collected from the Obstetrics and Gynecology Hospital in Al-Hilla city. The data under study consists of a dependent variable, y_i, representing 'neonatal jaundice,' and 15 independent variables that directly or indirectly affect neonatal jaundice. These variables are as follows:

1- x_1 : Maternal weight during pregnancy.
2- x_2: Number of hours of sleep for the mother during pregnancy.
3- x_3: Level of bilirubin in the mother's blood during pregnancy
4- x_4 :blood sugar level in the mother during her pregnancy.
5- x_5: Number of working hours for the mother during pregnancy
6- x_6: Blood pressure level during pregnancy.
7- x_7: Number of visits to prenatal care clinics.
8- x_8 : Hemoglobin level in the mother's blood during pregnancy.
9- x_9: Whether the mother is a smoker during pregnancy.

10-x_{10} : Age of the mother during pregnancy.
11-x_{11}: Birth order of the current child.
12-x_{12}: Blood group of the mother.
13-x_{13}: Type of delivery (natural birth, cesarean section).
14- x_{14}: Blood type of the child.

15-x_{15}: Level of blood sugar in the child's blood.

After obtaining the data, a regression model was built consisting of one dependent variable and 15 independent variables. The performance of our proposed method will be tested compared to the PCReg (Principal Component Regression) and BPCReg (Bayesian Principal Component Regression) methods. The evaluation criteria used will be Mean Squared Error (MSE) and Standard Error (S.E).

1. Table 3 -present the values of Mean Square Error (RMSE) and Standard Error (S.E).for the real-world data.

Methods	MSE	S E
PCReg	11.4239	8.5654
BPCReg	9.4825	5.0067
BLPCReg	3.4239	1.7832

Based on the results displayed in Table 3, we find that the values of the evaluation criteria, namely Mean Square Error (MSE) and Standard Error (S.E), obtained using our proposed method (BLPCReg) are significantly lower than the values obtained using the comparison methods (PCReg and BPCReg). This result indicates that our proposed method is superior to the comparison methods in terms of variable selection and parameters estimation.

From the figure, we can observe that variables x_2, x_5, and x_9 are independent variables that do not significantly influence the dependent variable (neonatal jaundice). On the other hand, the remaining independent variables have a clear impact on the dependent variable (neonatal jaundice.(This implies that variables x_2, x_5, and x_9 may not be important predictors of neonatal jaundice, while the other independent variables have a significant effect on neonatal jaundice.

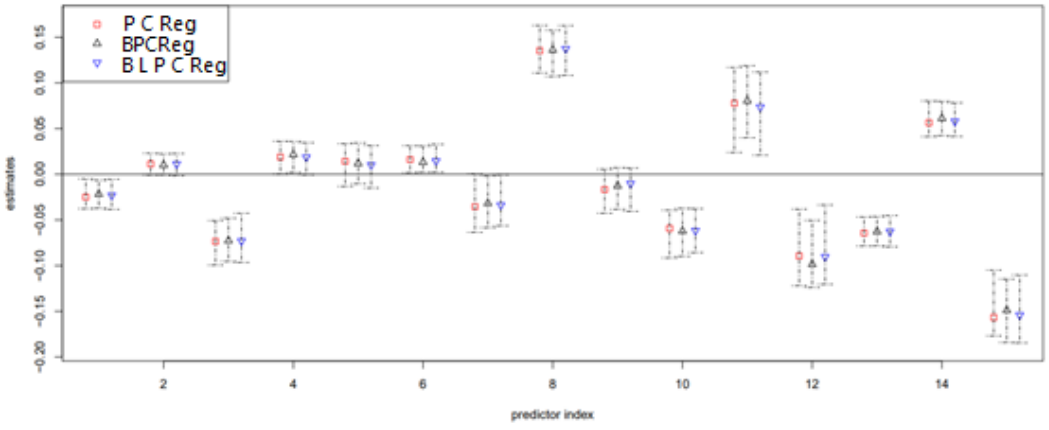

Figure -4- Shown confidence interval for independent variables according to three methods

4. Conclusions and Recommendations

4.1 Conclusions

Through simulation and real data analysis, we can conclude that our proposed method performs well in estimating the parameters and selecting variables compared to other methods. This is evident from the calculated values of the comparison criteria used in our current study, where the values of MMAD (Mean Magnitude of Absolute Differences), MSE (Mean Squared Error), and SE (Standard Error) obtained from the BLPCReg method are significantly smaller than the values obtained from the PCReg and BPCReg methods. This result clearly gives an advantage to our proposed method.

In the analysis of real data, we found that three independent variables had a non-significant effect on the response variable (neonatal jaundice). However, the remaining twelve variables showed a clear impact on the response variable (neonatal jaundice).

4.2 Recommendations

Using our proposed method to estimate regression models that suffer from multicollinearity is beneficial because our method addresses the issue of multicollinearity in two stages. The first stage involves employing principal component regression, and the second stage involves utilizing the Lasso method. We can expand our current study to other regression models due to their good properties in parameter estimation and variable selection. This study can be extended to binary regression models, tobit regression models, quantile regression models, and so on.

Acknowledgements

Acknowledgements and Reference heading should be left justified, bold, with the first letter capitalized but have no numbers. Text below continues as normal.

REFERENCES

[1] Akaike, H. (1973). Maximum likelihood identification of Normal autoregressive moving average models. Biometrika, 60(2), 255-265. distribution of birth outcomes. Empirical Economics, 26:247{257.

[2] Alhamzawi, R., 2013, "*Tobit quantile regression with adaptive lasso penalty*", The 4th International Scientific Conference of Arab Statistics 450 ISSN, pp. 1681-6870).

[3] Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. Journal of the Royal Statistical Society, Series B 36, 99–102. 7.

[4] Alibuhtto, M. C., & Peiris, T. S. G. (2015). Principal component regression for solving multicollinearity problem.

[5] Alabi, O. O., Ayinde, K., & Olatayo, T. O. (2008). Effect of multicollinearity on power rates of the ordinary least squares estimators. *Journal of Mathematics and Statistics*, *4*(2), 75.

[6] Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, *32*(2), 407-499.

[7] Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.

[8]Frost, J. (2017). Multicollinearity in regression analysis: problems, detection, and solutions. *Statistics by Jim*, *2*.

[9] Hashem, H., Vinciotti, V., Alhamzawi, R., and Yu, K. (2015). "Quantile regression with group lasso for classification." Advances in Data Analysis and Classification, 1–16.

[10] Junttila, V., & Laine, M. (2017). Bayesian principal component regression model with spatial effects for forest inventory variables under small field sample size. *Remote Sensing of Environment*, *192*, 45-57.

[11] Li, Q., Xi, R., & Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, *5*(3), 533-556.

[12]Murtaza Haider (2014). Model Selection and Akaike's Information Criterion (AIC) in R Journal of Statistical Software, Volume 36, Issue 3

[13] Park, T. and Casella, G. (2008). The Bayesian lasso. Journal of the American Statistical Association 103, 681–686.

[14] Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics ,461–464.4

[15] Sun, W., J. G. Ibrahim, and F. Zou (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. Genetics 185, 349–359.

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

[17] Yue, Y. R., & Hong, H. G. (2012). Bayesian Tobit Q Reg model for medical expenditure panel survey data. *Statistical Modelling*, *12*(4), 323-346.