Pak. J. Agri. Sci., Vol. 58(1), 179-186; 2021
ISSN (Print) 0552-9034, ISSN (Online) 2076-0906
DOI: 10.21162/PAKJAS/21.9322
http://www.pakjas.com.pk

MODELLING OF WHEAT PRODUCTION IN PUNJAB THROUGH THE REGULARIZED REGRESSION APPROACH WHILE ADDRESSING MULTICOLLINEARITY

Nadia Idrees[1,2,*] and Shahid Kamal[3]

[1]Department of Mathematics and Statistics, University of Agriculture Faisalabad, Pakistan;[2]Collegeof Statistical and Actuarial Sciences, University of the Punjab Lahore, Pakistan;[3]Government College University Faisalabad, Pakistan.
[*]Corresponding author's e-mail: nadia.stat.uaf@gmail.com

Agriculture sector plays a significant role in the economy of Pakistan. Amongst the major crops, wheat is essential food of Pakistani people's. Wheat production is a function of a number of factors. The identificatsion of important variables that affects the production of wheat is of prime interest of researcher to meet the basic need of increasing population. With multiple predictors, statistical modeling becomes complicated mainly due to the collinearity within the predictors. Regularized regression approaches do variable selection and shrinkage at same time. In this study, a rigorous comparison of the predictive performance of seven regularized regression approaches was performed via simulation while considering different levels of multicollinearity and sparsity. The results showed that Smoothly Clipped Absolute Deviation (SCAD) and Minimax Concave Penalty (MCP) performed better under low and high variation when the true model is sparse for all sample sizes based on mean squared error prediction (MSEP). Moreover, wheat production forecast model for Punjab province was estimated; while, using regularized regression methods and the significant predictors were identified. Among six predictors, area under crop, average retail price of fertilizer and average maximum temperature are important parameters that affect production of wheat. It is recommended that the farmers of the Punjab must be care full about these factors while sowing the wheat and government should provide facilities to farmers regarding these factors.
Keywords: Concave penalty, multicollinearity, prediction, regularized regression, sparse.

INTRODUCTION

Agriculture is the mainstay of Pakistan economy as it adds 18.9% to the total gross domestic product (GDP) and providing work to more than 42.3% of the workforce. The population of Pakistan is increasing at the rate of 2.4% annually according to the Population and Housing Census of Pakistan, 2017. This increase in population is rising demand for products from agricultural sector. Amongst the main agriculture products, wheat is the necessary food of the people of Pakistan. Wheat production is a function of a number of factors like water availability, area under cultivation, prices of fertilizer etc. The identification of important variables that affects the production of wheat is of prime interest of researcher to increase the production of wheat to meet the basic need of increasing population.

In developing the relationship between response and predictors, both prediction accuracy and interpretability are key issues. When there are multiple predictors, the selection of best subset of important predictors is required which is possible by excluding the less important predictors. Moreover, shrinkage of variables is important mainly if the true model has a sparse representation. A good predictive method selects a useful list of predictors with high prediction accuracy. Therefore, the regression method that selects the relevant and important predictors is of prime interest to the researchers. The objectives associated to the selection of the significant predictors are mainly dual; one is to get unbiased estimates of parameters and other is the prediction accuracy. Prior information from the research literature is usually seen as the key basis for the inclusion or exclusion of variables, but generally, it is not available for all research problem (Greenland, 2008).

Ordinary least squares (OLS) is a traditional estimation procedure and frequently used for estimating the multiple linear regression models, but it performs poorly in case of large number of predictors. Because of multicollinearity, OLS estimates although are unbiased but have large variances, which affects the prediction accuracy. Subset regression and Ridge regression are alternates to improve the OLS estimators, but both have some drawbacks in their implementation (Breiman, 1995). Subset regression provides interpretable models by selecting only a subset of predictors but it is extremely variable due to its inherent discreteness (Breiman,1995). Second way is to compromise over unbiasedness for which Hoerl and Kennard (1970) proposed the Ridge regression to obtain the biased estimates. It is a continuous process, which shrinks the regression coefficients, but it does not provide interpretable model as it does not exclude the unimportant variables.

To overcome such problems, Tibshirani (1996) developed a novel technique entitled as "Least absolute shrinkage and selection operator" (Lasso) for sparse model selection. Since that, various generalizations of Lasso method were proposed. The Lasso-type methods have become popular due to their property of (Albrecht *et al*., 2014) shrinkage. The objective function of such methods includes a penalty term. Different researchers suggest different assumptions regarding the penalty term, i.e. L_1 norm, L_2 norm or both L_1 and L_2 norm, which are termed as the tuning (regularized) parameters. Different shrinkage methods have been compared with OLS via Monte Carlo simulation in terms of model error and mean square error in literature. Adaptive elastic net and Bayesian model averaging revealed better stability and traditional estimates of regression coefficients and the standard errors as compared to basic stepwise methods (Morozova *et al.*, 2015). In a comparison of penalized regression methods, Oyeyemi *et al.* (2015) used Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for model selection. Their results showed that Lasso performs better at all three levels of muticollinearity and sample. Adaptive lasso performs best only for moderate muticollinearity and medium sample size.

A new approach named as raise regression as an alternate to the ordinary least squares was proposed by Gomez *et al.* (2020) to handle the problem of multicollinearity. They extended the theory of variance inflation factor after performing the raise regression. The application of this addition is to limit the raising factor. The also used mean square error for comparison as raise regression provides a biased estimator. The compared the raise estimator to ridge and Lasso. Their results showed that raise regression performed better than ridge regression in terms of mean square error. The regularized methods including Partial Least Squares, Principal Components Regression and Ridge Regression was compared and assessed by Goktas *et al.* (2020).They generate a number of different groups of datasets with six different levels of collinearities and sample sizes from standard normal distribution for 10000 replications. The mean squared error of the regression parameters had used for comparison purpose. Their findings showed that each method of prediction is affected by the number of predictors, sample size or level of collinearity. However, their results showed that Principal component regression had better results in terms of lower man squared error as compared to other two methods for any number of predictors.

An understanding of heredities for the improvement of current wheat varieties for mineral contents needs the investigation of genetic diversity. Ali *et al.* (2018) evaluated the wheat varieties via cluster analysis that have grains rich in zinc and iron. They estimate the hereditary diversity in grains for zinc and iron content along with agro-morphological traits 85 different wheat varieties. Average values for zinc and iron indicates the presence of genetic variability of all the varieties. Cluster analysis results in six clusters of germplasm. Mineral contents and most of the yield attributes were strongly correlated.

However, the earlier research work was based on a single moderate value of collinearity $\rho = 0.5$. In addition, most of the comparative studies did not consider the PLSR as a competitor to other shrinkage methods, which is also an important shrinkage method for high collinearity and non-sparse model. This study is helpful in deciding about the best regularized method in the presence or absence of collinearity, with large or small sample and with high or low variation, while the true model is sparse (not sparse).

MATERIALS AND METHODS

The usual multiple linear regression model is

$$Z = X\beta + \varepsilon \tag{1}$$

where $X_{n\times p}$ is the design matrix of p-predictors, vector of regression coefficients $\beta_{p\times 1}$, $\varepsilon_{n\times 1}$ is the independently and identically distributed normal noise with mean 0 and covariance $\sigma^2 I_n$, $Z_{n\times 1}$ is the random response vector. The OLS solution of (1) depends on the inverse of $X^T X$ and is

$$\boldsymbol{\hat{\beta}_{ols}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Z} \tag{2}$$

In case of multicollinearity, $X^T X$ is singular or nearly singular and hence creates problem in OLS solution. Ridge regression (Hoerl and Kennard, 1970) is a shrinkage method, an alternate method to improve the OLS estimators by allowing the biased estimators with smaller mean squared error. Calculation of shrinkage estimators is complicated than the ordinary least squares. The objective function for the shrinkage estimators is as follows:

$$\hat{\beta} = \arg\min_{\beta}\left\{\|Z - X\beta\|_2^2 + \lambda_n \sum_{j=1}^{p} |\beta_j|^{\vartheta}\right\} \tag{3}$$

where λ_n is the tuning parameter of the shrinkage method, $\lambda > 0$, and ϑ is the researcher's specified value of the norm. The first part of (3) is just the least squares objective function while second term is the penalty factor. The ridge estimator by solving the penalized least squares from (3) is

$$\hat{\beta}_{Ridge} = \arg\min_{\beta}\left\{\|Z - X\beta\|_2^2 + \lambda\|\beta\|_2^2\right\} \tag{4}$$

where $\|Z - X\beta\|_2^2$ is the usual residual sum of squares. The term $\|\beta\|_2^2 = \sum_{j=1}^{p}\beta_j^2$ is the L_2-norm penalty on β. Here strength of the penalty is controlled by the regularization parameter λ. The large values of λ will result in better shrinkage.

PLSR is used particularly when the predictors are highly correlated and is alternate to ridge regression. It was originally developed by Wold (1975) and thereafter Frank and

Friedman (1993) and Goutis (1996) proved its properties. After that PLSR have been applied in many fields and compared with other shrinkage methods (Phatak and Jong, 1997; Braak and Jong, 1998). PLSR attempts to find the linear decomposition of X and Z such that $X = RU^T + A$ and $Z = SV^T + B,$ where R and S are the score matrices of X and Z respectively, U and V are loading matrices and A and B are the matrices of residuals. The PLSR estimator obtained by regressing Z not on X itself but on the scores is

$$\hat{\beta}_{PLSR} = UV \tag{5}$$

The Lasso estimator of regression coefficients uses the L_1 penalty that enables the Lasso to regularize the least squares fit and shrinks some of beta coefficients to zero simultaneously for some chosen value of tuning parameter. The Lasso estimator obtained from (3) is

$$\hat{\beta}_{Lasso} = \arg\min_{\beta} \left\{ \|Z - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \tag{6}$$

where $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$ is the L_1 norm penalty on β and $\lambda \geq 0$ is the regularization parameter.

Fan and Li (2001) showed that Lasso turn out in biased estimates for large coefficient, hence, not attain high prediction accuracy. They suggested a new method for variable selection called SCAD and proved that it satisfies the properties of an oracle procedure. Fan and Li (2001) proposed the SCAD penalty, and Zhang (2010) proposed the MCP penalty. Both are the penalized methods with nonconvex penalties and are solutions to

$$\arg\min_{\beta} \left\{ \|Z - X\beta\|^2 + P_{\lambda,\gamma}(\beta) \right\} \tag{7}$$

where for some $\gamma > 2, t > 0,$ and $\lambda \geq 0$

$$P_{\lambda,\gamma}^{T}(t) = \lambda \left\{ I(t < \lambda) + \frac{(\gamma\lambda - t)_+}{(\gamma - 1)\lambda} I(t > \lambda) \right\} \tag{8}$$

for SCAD and

$$P_{\lambda,\gamma}^{T}(t) = \left(\lambda - \frac{t}{\gamma} \right)_+ \tag{9}$$

for MCP.

Another generalization of Lasso was introduced by Zou and Hastie (2005), called the "Elastic net". Elastic net estimator is

$$\hat{\beta}_{Elnet} = \arg\min_{\beta} \left\{ \|Z - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \right\} \tag{10}$$

where λ_1 and λ_2 are tuning parameters usually selected by cross validation. The penalty term of elastic net is the combination of Ridge penalty and the penalty term in Lasso. Zou (2006) suggested the use of adaptive weights for penalizing different coefficients, and provided the evidence regarding the inconsistency of the Lasso. The Adaptive Lasso estimator is

$$\hat{\beta}_{AdLasso} = \arg\min_{\beta} \left\{ \|Z - X\beta\|_2^2 + \lambda \sum_{j=1}^{p} \hat{w}_j \|\beta_j\| \right\} \tag{11}$$

where λ is the regularization parameter, $\hat{w} = 1 / |\hat{\beta}|^{\gamma}$, are the adaptive weights with $\gamma > 0$ (possible values suggested by Zou (2006) are 0.5, 1, and 2) for the adjustment of adaptive weights, $\hat{\beta}$ is any consistent initial estimator of $\beta,$ it may be $\hat{\beta}_{OLS}$ or $\hat{\beta}_{Ridge}$.

Simulation study: To evaluate the performance of above explained seven regularized regression methods for each combination of sample size, variation level and correlation $(n, \sigma,$ and $\rho),$ we conduct a simulation study using R-software. The R-packages "GLMNET" (Friedman et al., 2009), "NCVREG" (Breheny and Breheny, 2020), and "PLS" (Wehrens and Mevik, 2007) have been used to estimate the regularized regression models. The OLS coefficients have been used to compute the adaptive weights for Adaptive Lasso as suggested by Zou (2006). An important concern during the model estimation is the selection of regularized parameter. Various approaches like Mallow's Cp, leave one out cross validation, k-fold cross validation, and generalized cross validation are used for this purpose. In this article, k-fold (k=10) cross validation methodology was adopted as this approach has been frequently and successfully adopted in diverse statistical approaches. In addition, it is directly associated to the predictive performance. The multi-variables data sets were simulated from the model

$$Z = X\beta + \sigma\varepsilon \tag{12}$$

where β is the vector of parameters (two true parameter vectors having different sparsity structures were considered), ε is a random error from standard normal distribution, and x_i (columns of X matrix) were simulated from $N(0, \Sigma),$ where the (i, j) element of Σ (positive definite covariance matrix) is $\rho^{|i-j|}$ for all i and j (i=1, 2, …, p, j= 1, 2, ..., p, and $i \neq j$). To study the effect of different levels of collinearity among predictors, the correlation between x_i and x_j was considered as

$$\rho = 0.0, 0.05, 0.15, 0.25, ..., 0.95, 0.99$$

Additionally, different levels of variation $(\sigma = 1, 3)$ were also considered in order to check the effect of variation on prediction. The data sets were simulated separately for training set as well as for a test set. We estimate the models by using training data while test data were used to compute mean squared error prediction (MSEP). Sample sizes for training set (test set) were taken as: 20(40), 60(120), 100(200), 200(400), 500(1000). The MSEP is defined as

$$MSEP = \frac{1}{n^*}\sum_{i=1}^{n^*}\left(Z_i - \hat{Z}_i\right)^2 \qquad (13)$$

where Z_i and $\hat{Z}_i$ are the observed and predicted response in test data, and the sum is over all observations in the test data. For all the shrinkage methods and each data set, the median of mean squared error prediction over 500 simulations was used for comparison purpose.

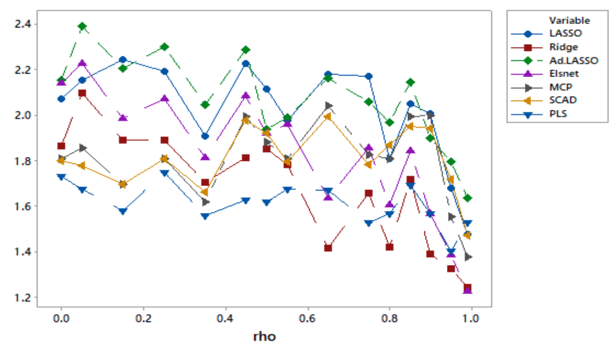
Model-1: a sparse true model with eight predictors, $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$

Model-2: true model is not sparse with eight predictors,

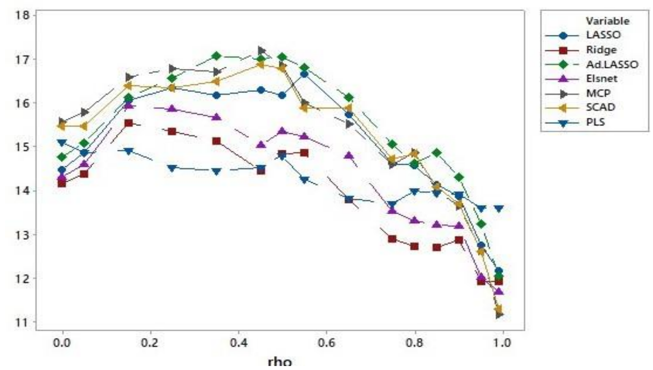
$\beta = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)$

RESULTS

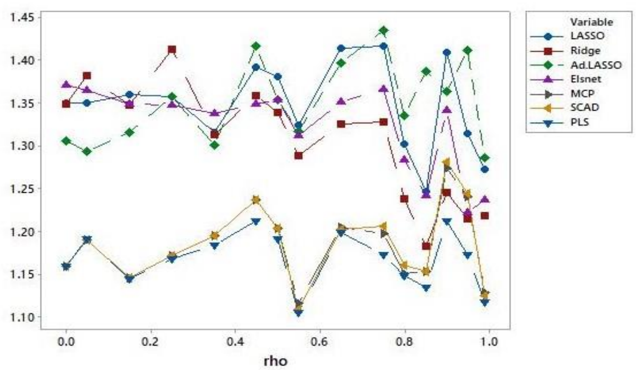
In this section, a graphical comparison of the regularized regression approaches has been presented. The simulation results for sample sizes 200(400) and larger showed that there were non-significant changes occurred in the selection of best shrinkage method. Therefore, the results of only six combinations (I-VI) are shown in Figures 1 & 2 due to the scarcity of space. We considered the following cases:



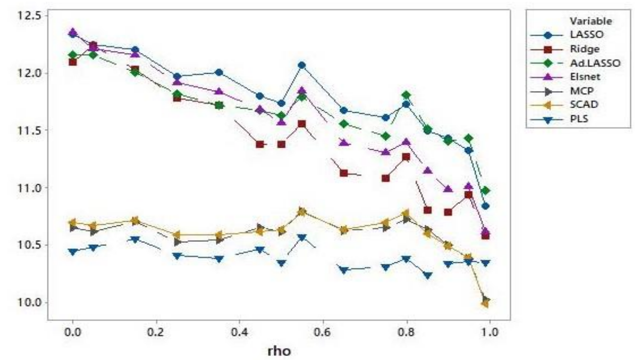
(a): n=20(40) and σ=1



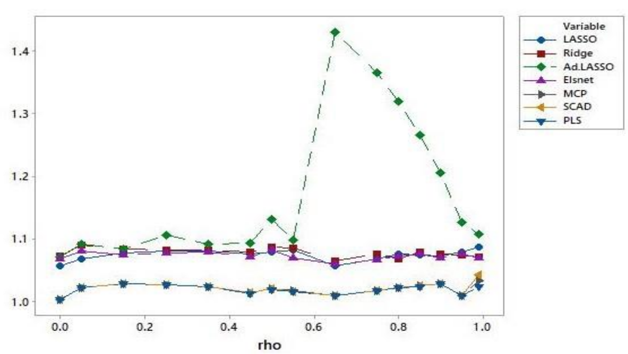
(d): n=20(40) and σ=3



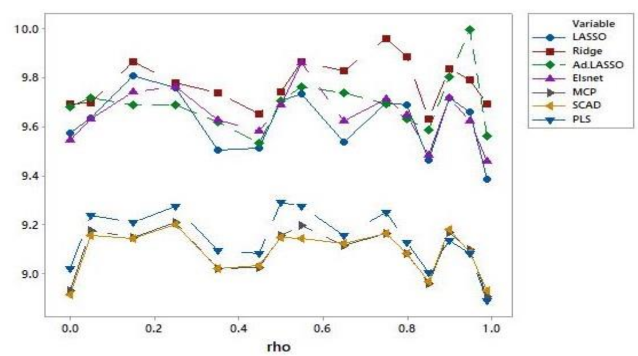
(b): n=60(120) and σ=1



(e): n=60(120) and σ=3

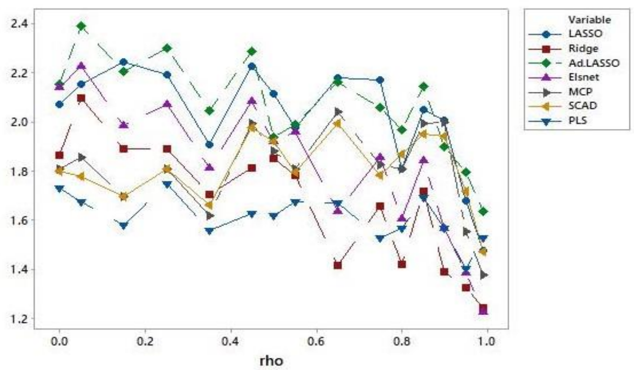


(c): n=500(1000) and σ=1

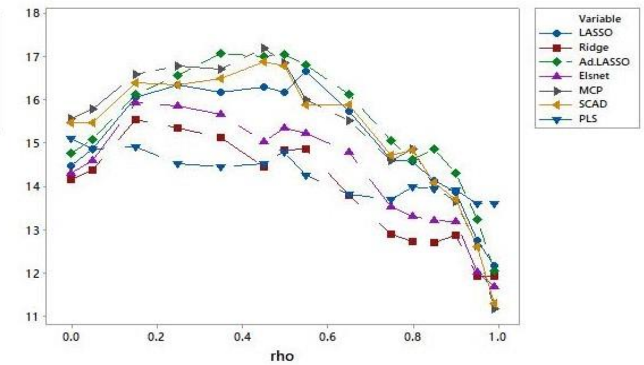


(f): n=500(1000) and σ=3

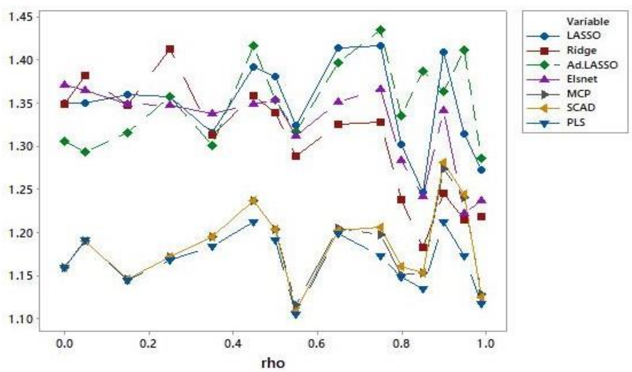
Figure 1. Median (MSEP) vs rho for model-1



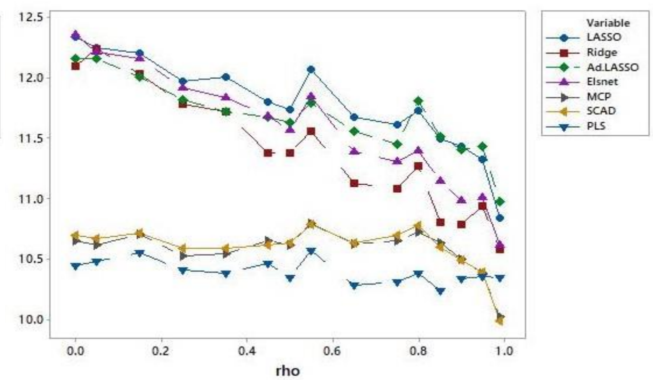
(a): n=20(40) and σ=1



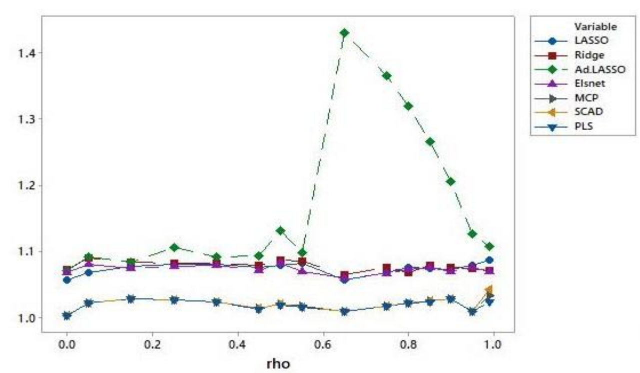
(d): n=20(40) and σ=3



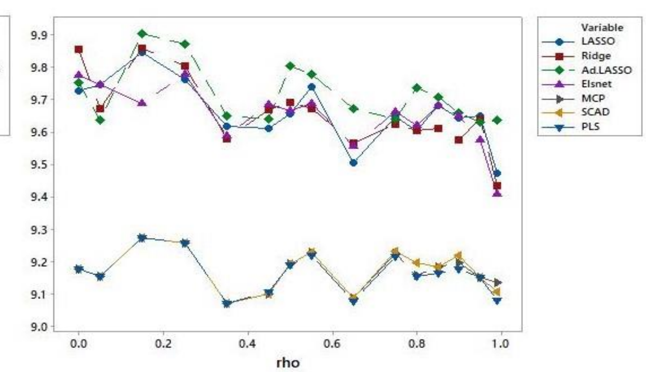
(b): n=60(120) and σ=1



(e): n=60(120) and σ=3



(c): n=500(1000) and σ=1



(f): n=500(1000) and σ=3

Figure 2. Median(MSEP) vs rho for model-2

Case I: n=20(40), σ=1	Case IV: n=20(40), σ=3
Case II: n=60(120), σ=1	Case V: n=60(120), σ=3
Case III: n=500(1000), σ=1	Case VI: n=500(1000), σ=3

Figure 1, present the results for model-1. In this case the true model is a sparse one. The method of SCAD and MCP outperforms for low and high variation for low to moderate collinearity. Furthermore, for model-1, PLSR shows low MSEP for small samples and high multicollinearity. Increase in the variation (from 1 to 3) also increases the mean squared error prediction for all methods. However, performance of Ridge regression and Adaptive lasso is of poorer quality in all scenarios.

Figure 2 presents the results of model-2. The model is not sparse, having all small non-zero coefficients. In case of not sparse model, some different observations have been made on the performance of regularized regression methods. PLSR outperforms for small and moderate sample while performs equally as MCP and SCAD for large samples. Performance of Ridge regression is also quite well for small samples and high correlation.

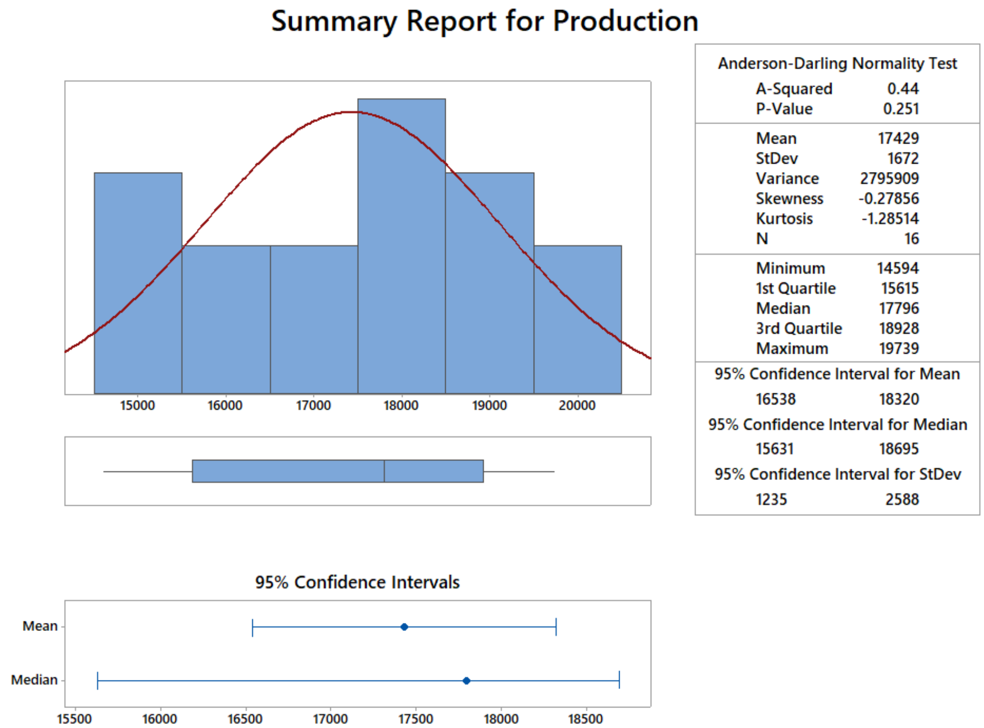


Figure 3. Summary report for the wheat production

Application to real data: For real data example, the data on wheat production was taken from Agriculture statistics, Pakistan Bureau of Statistics (2018) for the years 2000-2016.The data set contains the wheat production, average rainfall (AR), area under cultivation, average retail price of fertilizer (ARPF), distribution of seed (DOS), number of tube wells (TW), and average maximum temperature (AMT).The dependency of production of wheat on different factors affecting the production was investigated using regularized regression methods. Moreover, a sparse model which has maximum shrinkage and minimum prediction error have to fit. The Anderson-Darling Normality test was performed for testing the normality of wheat production, Figure 3 shows the normality of response (A^2 = 0.44, p-value > 0.05). A regression model estimated by OLS had been used to assess if there is any multicollinearity effect present. The variance inflation factor (VIF) obtained for TW was large (>5) and thus, there is an evidence of the presence of multicollinearity. Because of multicollinearity, the combine effect of all the predictors on wheat production was highly significant (F-ratio=14.35, p-value=0.000) while they were non-significant individually. Examination of correlation matrix (Table 1) also confirms the presence of collinearity. Hence, fitting regression model by using ordinary least squares was not appropriate and the regression approaches based on regularized methods were used to establish the relationship between production of wheat and 6 candidate predictors.

Table 1. Correlation matrix of all variables

	Prod	**Area**	**ARPF**	**DOS**	**TW**	**ARF**	**AMT**
Prod	1.00	0.88	0.82	-0.54	0.90	0.34	-0.37
Area	0.88	1.00	0.70	-0.50	0.84	0.28	-0.19
ARPF	0.82	0.70	1.00	-0.28	0.81	0.32	-0.25
DOS	-0.54	-0.50	-0.28	1.00	-0.60	-0.29	0.25
TW	0.90	0.84	0.81	-0.60	1.00	0.37	-0.32
ARF	0.34	0.28	0.32	-0.29	0.37	1.00	-0.40
AMT	-0.37	-0.19	-0.25	0.25	-0.32	-0.40	1.00

Table 2 represents the number of variables selected from seven regression methods. Two methods i.e. Ridge regression and PLS regression do not set any regression coefficient equal to zero because these are the elementary methods of shrinkage without the ability of variable selection.

Table 2. Number of variables selected by each regularized regression method

Total Predictors	**6**
Ridge Regression	6
Lasso	3
Adaptive Lasso	1
Elastic net	5
MCP	6
SCAD	5
PLSR	6

Table 3. Regression coefficients from different regularized regression approaches

	Ridge	Lasso	Ad Lasso	EL.Net	MCP	SCAD	PLSR
(Intercept)	11366.950	2222.050	-7013.890	6670.830	5378.720	4784.330	11145.390
Area	1.270	1.801	3.738	1.497	2.581	2.592	0.737
ARPF	0.533	0.388	*	0.536	0.776	0.721	1.129
DOS	-0.018	*	*	-0.008	-0.012	-0.005	-0.025
Tubewells	0.002	0.003	*	0.003	0.002	0.003	0.004
ARF	0.592	*	*	*	*	*	0.014
AMT	-132.736	*	*	-47.938	-222.058	-225.750	-0.001

The Lasso and Adaptive Lasso result in the greatest reduction. MCP and SCAD, however, do not reduce the model size significantly. This is an indication that the model is sparse with few large coefficients and others are zero. Based on the results of simulation study (model-1), only MCP, SCAD and Lasso methods have been used to fit the wheat production forecast model. The most effecting predictors to the production of wheat, identified by the Lasso were; area, ARPF and TW. These predictors were also selected by all other methods except that adaptive lasso. In addition, MCP and SCAD also identified two other variables (AMT and DOS) as significantly affecting the production of wheat.

DISCUSSION

The results of simulation suggested that the PLSR is a worth competitor to concave penalty methods if true model is not sparse and sample size is small or when multicollinearity is very high. PLSR produces lower MSEP even than Lasso, Elastic net and Adaptive lasso. Furthermore, with sparse true model, SCAD and MCP performs best. Adaptive lasso and Elastic net, however, not result well in case of reducing prediction error significantly. However, due to different levels of collinearity, dimensions and sparsity structures in data sets, no single shrinkage method is robust in all situations. On the other hand, while fitting wheat production model, area under crop, average retail price of fertilizer and average maximum temperature are important parameters that affect production of wheat. It is further recommended that the farmers of the Punjab must be care full about these factors while sowing the wheat.

REFERENCES

Albrecht, T., H.J. Auinger, V. Wimmer, J.O. Ogutu, C. Knaak, M. Ouzunova, H.P. Piepho and C.C. Schön. 2014. Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theo. Appl. Gen. 127:1375-1386.

Ali, A., H.A. Sadaqat, M. Kashif and M.A. Wahid. 2018. Exploration of breeding potential for genetic biofortification and yield in spring wheat (*Triticum aestivum* L.). Pak. J. Agr. Sci. 55: 793-799.

Braak, C.J.F.T. and S.D. Jong. 1998. The objective function of partial least squares regression. J. Chemo. 12:41-54.

Breiman, L. 1995. Better subset regression using the nonnegative garrote. Technometrics 37:373-384.

Breheny, P. and M.P. Breheny. 2020. R-Package ‘ncvreg’. https://cran.rproject.org/web/packages/ncvreg/index.html

Fan, J. and R. Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am.Stat. Assoc. 96:1348-1360

Frank, L. E. and J.H. Friedman. 1993. A statistical view of some chemometrics regression tools. Technometrics 35:109-135.

Friedman, J., T. Hastie and R. Tibshirani. 2009. glmnet: Lasso and elastic-net regularized generalized linear models. R package version 1. https://cran.r-project.org/web/packages/glmnet/readme/README.html.

Gomez, R.S., A.R. Sanchez, C.G. Garcia and J.G Perez. 2020. The VIF and MSE in Raise Regression. Mathematics 8:605-633.

Goktas, A. and O. Akkus. 2020. Comparison of partial least squares with other prediction methods via generated data. J. Stat. Comp. Sim. 90:1-16.

Goutis, C. 1996. Partial least squares algorithm yields shrinkage estimators. The ann. Stat. 24:816-824.

Greenland, S. 2008. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. Am. J. Epid. 167:523-529.

Hoerl, A.E. and R.W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12:55-67.

Morozova, O., O. Levina, A. Uusküla and R. Heimer. 2015. Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in Russia. BMC Med. Res. Methodol. 30:15-71.

Oyeyemi, G.M., E.O. Ogunjobiand A.I. Folorunsho. 2015. On performance of shrinkage methods–a Monte Carlo Study. Int. J. Stat. Appl. 5:72-76.

Phatak, A. and D.S. Jong. 1997. The geometry of partial least squares. J. Chemo. 11:311-338.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B. (Method). 58:267-288.
Wehrens, R. and B.H. Mevik. 2007. The pls package: principal component and partial least squares regression in R. R package. https://cran.r-project.org/web/packages/pls/pls.
Wold, H. 1975. Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. J. Appl. Prob. 12:117-142.
Zhang, C.H. 2010. Nearly unbiased variable selection under minimax concave penalty. The Ann. Stat. 38:894-942.
Zou, H. 2006. The adaptive lasso and its oracle properties. J. Am. Stat. Assoc. 101:1418-1429.
Zou, H. and T. Hastie. 2005. Regularization and variable selection via the elastic net. J. R.Stat. Soc. Ser. B (Method). 67:301-320.

[Received 17 July 2019; Accepted 16 Sept 2020; Published (online) 11 Jan 2021]