RESEARCH ARTICLE

Performance of LASSO and Elastic net estimators in Misspecified Linear Regression Model

M. Kayanan[1, 2,*] and P. Wijekoon[3]

[1]Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka
[2]Department of Physical Science, Vavuniya Campus of the University of Jaffna, Vavuniya, Sri Lanka
[3]Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka

Abstract: Ridge Estimator (RE) has been used as an alternative estimator for Ordinary Least Squared Estimator (OLSE) to handle multicollinearity problem in the linear regression model. However, it introduces heavy bias when the number of predictors is high, and it may shrink irrelevant regression coefficients, but they are still in the model. Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic net (Enet) estimator have been used to make the variable selection and shrinking the regression coefficients simultaneously. Further, the model misspecification due to excluding relevant explanatory variable in the linear regression model is considered as a severe problem in statistical research, and it will lead to bias and inconsistent parameter estimation. The performance of RE, LASSO and Enet estimators under the correctly specified regression model was well studied in the literature. This study intends to compare the performance of RE, LASSO and Enet estimators in the misspecified regression model using Root Mean Square Error (RMSE) criterion. A Monte-Carlo simulation study was used to study the performance of the estimators. In addition to that, a real-world example was employed to support the results. The analysis revealed that Enet outperformed RE and LASSO in both correctly specified model and misspecified regression model.

Keywords: LASSO, Elastic net, Misspecified model, Root Mean Square Error, Monte-Carlo simulation.

INTRODUCTION

Consider the linear regression model

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

where y is the $n \times 1$ vector of observations on the predictor variable, X is the $n \times p$ matrix of observations on p non stochastic regressor variables, β is a $p \times 1$ vectors of unknown parameters, ε is the $n \times 1$ normally distributed vector of disturbances, such that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \Sigma = \sigma^2 I$.

Usually, by minimising the Error Sum of Squares (ESS) $\varepsilon'\varepsilon$,

$$ESS = (y - X\beta)'(y - X\beta) \tag{2}$$

the Ordinary Least Squared Estimator (OLSE), which is the Best Linear Unbiased Estimator (BLUE) for β, is obtained as

$$\hat{\beta}_{OLSE} = (X'X)^{-1}X'y \tag{3}$$

It is well-known that OLSE is unstable and produces estimates having high variance when multicollinearity exists among explanatory variables, i.e., the columns of X are highly correlated. To handle this problem, Hoerl and Kennard (1970) proposed Ridge Estimator (RE) by minimising ESS subject to the constraint $\|\beta\|_2 \leq t$, where $t \geq 0$ is a turning parameter. Further, it can be defined as

$$\hat{\beta}_{RE} = \underset{\beta}{\operatorname{argmin}}\{(y - X\beta)'(y - X\beta) + k\|\beta\|_2\}$$

$$= (X'X + kI)^{-1}X'y \tag{4}$$

where $k > 0$ is the regularization parameter, and $\|\beta\|_2 = \sum_{j=1}^{p} \beta_j^2$ is known as the L2 norm. Note that for any k there exist an t equal to $\|\beta\|_2$. This is the dual form of the optimization problem.

RE helps to obtain the estimates with smaller variance by shrinking the regression coefficients towards zero. However, it has two significant issues in a high dimensional linear model. First issue is that it introduces heavy bias when the number of predictors is high, and secondly, it may shrink irrelevant regression coefficients, but they are still in the model. To overcome this problem, Tibshirani (1996) proposed the Least Absolute Shrinkage and Selection Operator (LASSO) by minimising ESS subject to the constraint $\|\beta\|_1 \leq t$, for variable selection and shrinking the regression coefficients. Further, it is defined as

$$\hat{\beta}_{LASSO} = \underset{\beta}{\operatorname{argmin}}\{(y - X\beta)'(y - X\beta) + \lambda\|\beta\|_1\} \tag{5}$$

where $\lambda > 0$ is the regularization parameter, and $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$ is known as the L1 norm. Similarly, for any λ there exist an t equal to $\|\beta\|_1$.

Since $\|\beta\|_1$ is not a differentiable function, there is no analytical solution to estimate LASSO. Therefore, the numerical methods have been used by researchers to find solutions to this problem. Tibshirani (1996) used the standard quadratic programming technique, Fu (1998) proposed the shooting algorithm, and Efron *et al*. (2004) proposed the Least Angle Regression (LARS)

*Corresponding Author's Email: mgayanan@vau.jfn.ac.lk

https://orcid.org/0000-0003-2662-4383

algorithm. The LASSO estimation method handles both the multicollinearity problem and best feature selection simultaneously in the high dimension linear regression model. However, according to Zou and Hastie (2005) LASSO estimation procedure is unstable when the number of predictors p is higher than the number of observations n. Further, the prediction performance of RE dominates LASSO if there exist high multicollinearity among predictors. To handle this problem, Zouand Hastie (2005) proposed Elastic net (Enet) estimator by combining RE and LASSO, and it is defined as

$$\widehat{\boldsymbol{\beta}}_{Enet} = \underset{\boldsymbol{\beta}}{\text{argmin}} \left\{ (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) + k\sum_{j=1}^{p}\beta_j^2 + \lambda\sum_{j=1}^{p}|\beta_j| \right\} \tag{6}$$

The LARS-EN algorithm, which is a modified version of the LARS algorithm, has been used to estimate the Enet.

The misspecification of the linear model is unavoidable in the practical situation when fitting regression models. It may occur due to including irrelevant explanatory variables or excluding relevant explanatory variables, or measurement errors in variables.

Note that model (1) can be written as

$$\boldsymbol{y} = \boldsymbol{X_1\beta_1} + \boldsymbol{X_2\beta_2} + \boldsymbol{\varepsilon} \tag{7}$$

where, $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are the $n \times q$ and $n \times r$ matrices $(p = q + r)$, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are the $q \times 1$ and $r \times 1$ vectors. Let us say that the researcher misspecified the regression model (7) by omitting $\boldsymbol{X_2}$, then the model (7) becomes

$$\boldsymbol{y} = \boldsymbol{X_1\beta_1} + \boldsymbol{u} \tag{8}$$

where $\boldsymbol{u} = \boldsymbol{X_2\beta_2} + \boldsymbol{\varepsilon}$.

Note that $E(\boldsymbol{u}) \neq \boldsymbol{0}$, and the omitted variables may be correlated with the variables in the model if $\boldsymbol{X}$ is multicollinear. Therefore, one or more assumptions of the linear regression model will be violated when the model is misspecified, and hence the estimators become biased and inconsistent.

FU (1998), Zou and Hastie (2005), and Oyeyemi *et al*. (2015) examined the performance of RE, LASSO and Enet estimators in the correctly specified linear regression model. In this research, we examined the performance of LASSO and Enet estimator in comparison to RE using Root Mean Square Error (RMSE) criterion when the regression model is misspecified due to the exclusion of some important variables. Rest of the article contains the following contents: estimators under the misspecified model, a common form to represent the RE, LASSO and Enet estimators, a Monte Carlo simulation study and a numerical example to discuss the performance of the estimators, and finally some concluding remarks.

MATERIALS AND METHODS

Now we write RE, LASSO and Enet estimator for the misspecified linear regression model (8) as follows

$$\widehat{\boldsymbol{\beta}}_{RE}^* = (\boldsymbol{X_1'X_1} + k\boldsymbol{I})^{-1}\boldsymbol{X_1'y} \tag{9}$$

$$\widehat{\boldsymbol{\beta}}_{LASSO}^* = \underset{\boldsymbol{\beta_1}}{\text{argmin}} \left\{ (\boldsymbol{y} - \boldsymbol{X_1\beta_1})'(\boldsymbol{y} - \boldsymbol{X_1\beta_1}) + \lambda\sum_{j=1}^{q}|\beta_j| \right\} \tag{10}$$

$$\widehat{\boldsymbol{\beta}}_{Enet}^* = \underset{\boldsymbol{\beta_1}}{\text{argmin}} \left\{ (\boldsymbol{y} - \boldsymbol{X_1\beta_1})'(\boldsymbol{y} - \boldsymbol{X_1\beta_1}) + k\sum_{j=1}^{q}\beta_j^2 + \lambda\sum_{j=1}^{q}|\beta_j| \right\} \tag{11}$$

Now, let $\alpha = \frac{\lambda}{k+\lambda}$ then $(1 - \alpha) = \frac{k}{k+\lambda}$. Zou and Hastie (2005) have shown that solving $\boldsymbol{\beta_1}$ in equation (11) is equivalent to minimising ESS subject to the constraint $(1 - \alpha)\|\boldsymbol{\beta_1}\|_2 + \alpha\|\boldsymbol{\beta_1}\|_1 \leq t,$, where $0 < \alpha < 1$ and α is referred to as the mixing percentage of regularization parameters. Further, it can be defined as

$$\widehat{\boldsymbol{\beta}}_{Enet}^* = \underset{\boldsymbol{\beta_1}}{\text{argmin}} \left\{ (\boldsymbol{y} - \boldsymbol{X_1\beta_1})'(\boldsymbol{y} - \boldsymbol{X_1\beta_1}) + (k + \lambda)\sum_{j=1}^{q}\left[(1 - \alpha)\beta_j^2 + \alpha|\beta_j|\right] \right\} \tag{12}$$

Note that $\alpha = 0$ implies $\lambda = 0$, and then Enet estimator in equation (12) is equivalent to RE. Similarly, $\alpha = 1$ implies $k = 0$, and then Enet estimator in equation (12) is equivalent to LASSO. Hence, we can write the RE, LASSO and Enet estimator in a common form in the misspecified regression model as below:

$$\widehat{\boldsymbol{\beta}}^*(\alpha) = \underset{\boldsymbol{\beta_1}}{\text{argmin}} \left\{ (\boldsymbol{y} - \boldsymbol{X_1\beta_1})'(\boldsymbol{y} - \boldsymbol{X_1\beta_1}) + (k + \lambda)\sum_{j=1}^{q}\left[(1 - \alpha)\beta_j^2 + \alpha|\beta_j|\right] \right\} \tag{13}$$

where $\widehat{\boldsymbol{\beta}}^*(\alpha) = \begin{cases} \widehat{\boldsymbol{\beta}}_{RE}^*, & \alpha = 0 \\ \widehat{\boldsymbol{\beta}}_{Enet}^*, & 0 < \alpha < 1 \\ \widehat{\boldsymbol{\beta}}_{LASSO}^*, & \alpha = 1 \end{cases}$

The RMSE, which is the expected prediction error of the estimators, is given by

$$RMSE\left(\widehat{\boldsymbol{\beta}}^*(\alpha)\right) = \sqrt{\frac{1}{n}\left(\boldsymbol{y}_{new} - \boldsymbol{X}_{new}\widehat{\boldsymbol{\beta}}^*(\alpha)\right)'\left(\boldsymbol{y}_{new} - \boldsymbol{X}_{new}\widehat{\boldsymbol{\beta}}^*(\alpha)\right)} \tag{14}$$

where $(\boldsymbol{y}_{new}, \boldsymbol{X}_{new})$ includes new observations that are not used to obtain the coefficient estimates $\widehat{\boldsymbol{\beta}}^*(\alpha)$.

The "glmnet" package available in R programming language was used to estimate RE, LASSO and Enet solutions. A Monte Carlo simulation study was conducted to examine the performance of the RE, LASSO and Enet estimator. Further, a numerical example is employed to supports the results. For the simulation study, α of the Enet estimator was selected as $(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9)$. The minimum RMSE and the suitable value of regularization parameters for the particular problem were calculated using K-fold cross-validation method as suggested by Tibshirani (1996) and FU (1998).

RESULTS AND DISCUSSIONS

Simulation study

According to McDonald and Galarneau (1975), now we generate the regressor variables as follows:

$$X_{ij} = \sqrt{(1-\rho^2)}z_{ij} + \rho z_{i,p+1} \quad ;$$

$$i = 1, 2, \ldots, n. \; j = 1, 2, \ldots, p. \quad (15)$$

where z_{ij} is an independent standard normal pseudo random number, and ρ is specified so that the theoretical correlation between any two explanatory variables is given by ρ^2. We used the linear regression model of 100 observations and 20 regressors, and the dependent variable is generated by using the following equation

$$y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{20} X_{i20} + \varepsilon_i \quad ;$$

$$i = 1, 2, \ldots, 100, \quad (16)$$

Where ε_i is a normal pseudo random variable with mean zero and variance one. Also, we choose $\boldsymbol{\beta'} = (\beta_1, \beta_2, \ldots, \beta_{20})$ as the normalized eigenvector corresponding to the largest eigenvalue of XX' for which $\boldsymbol{\beta'\beta} = 1.$.

To investigate the effects of different degrees of multicollinearity on the estimators, we choose $\rho = (0.7, 0.8, 0.9)$ which represent low, moderate and high multicollinearity, respectively. To study the effect of misspecification, we choose $X_I = (X_1, X_2, \ldots, X_{14})$ and $X_2 = (X_{15}, X_{16}, \ldots, X_{20})$, where X_2 was assumed as the regressor matrix related to omitted variables in the misspecified regression model. Since the execution time for the simulation of the algorithm is too long, the simulation was repeated 50 times according to Tibshirani (1996) and FU (1998).

The estimated RMSE values of the Ridge, LASSO and Enet estimators versus regularization parameter when $\rho = 0.7$, $\rho = 0.8$ and $\rho = 0.9$ are displayed in Figures 1–3, respectively. The average minimum simulated RMSE values of the estimators and the optimal value of regularization parameters are summarized in Table 1.

From Figure 1-3, we can observe that Enet outperforms RE and LASSO in both correctly specified and misspecified regression model.
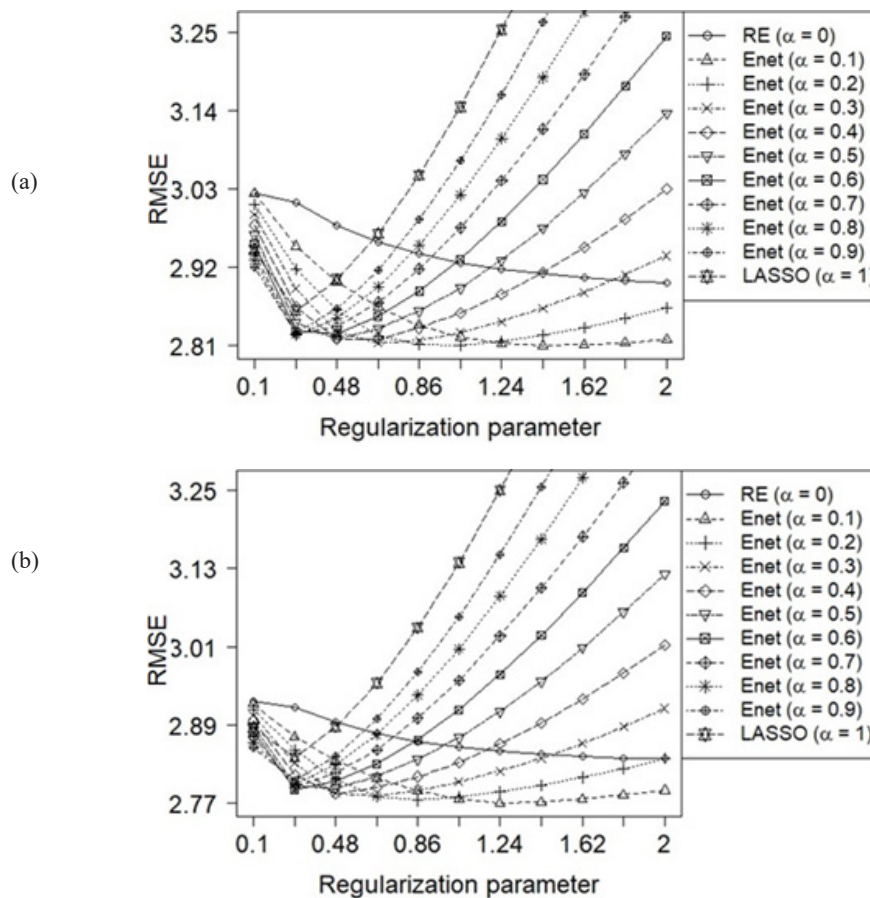
Figure 1: Estimated RMSE values of the estimators when ρ= 0.7; (a) under correctly specified model (b) under misspecified model.
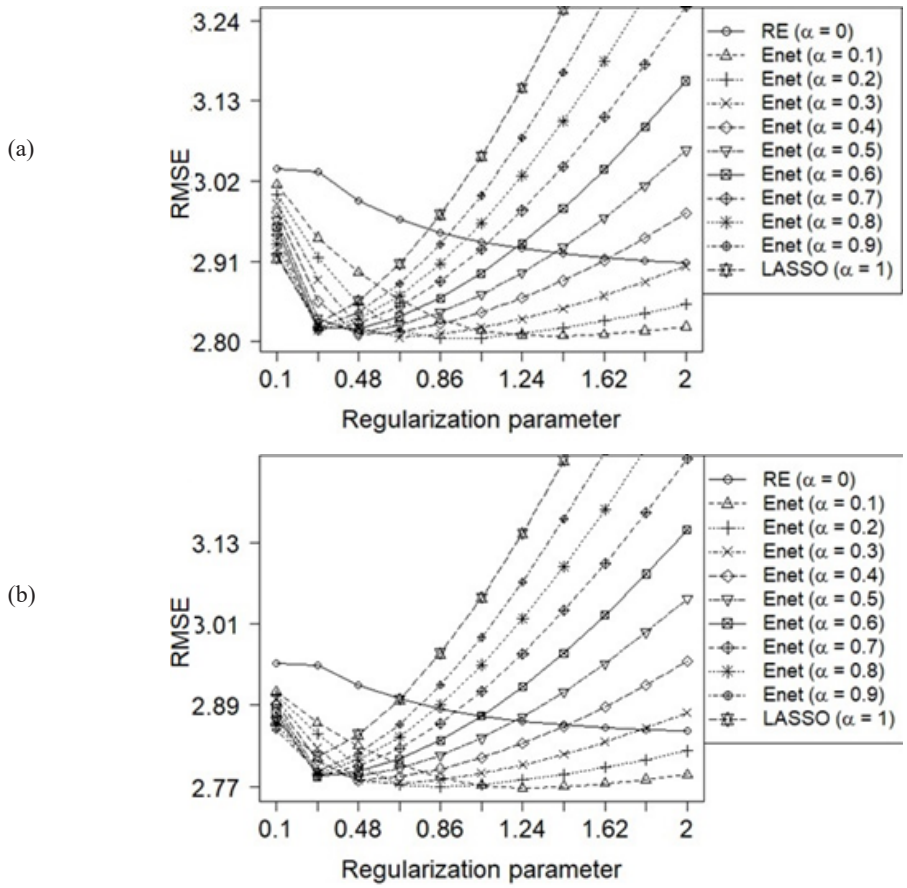
(a)

(b)

Figure 2: Estimated RMSE values of the estimators when $\rho = 0.8$; (a) under correctly specified model (b) under misspecified model.
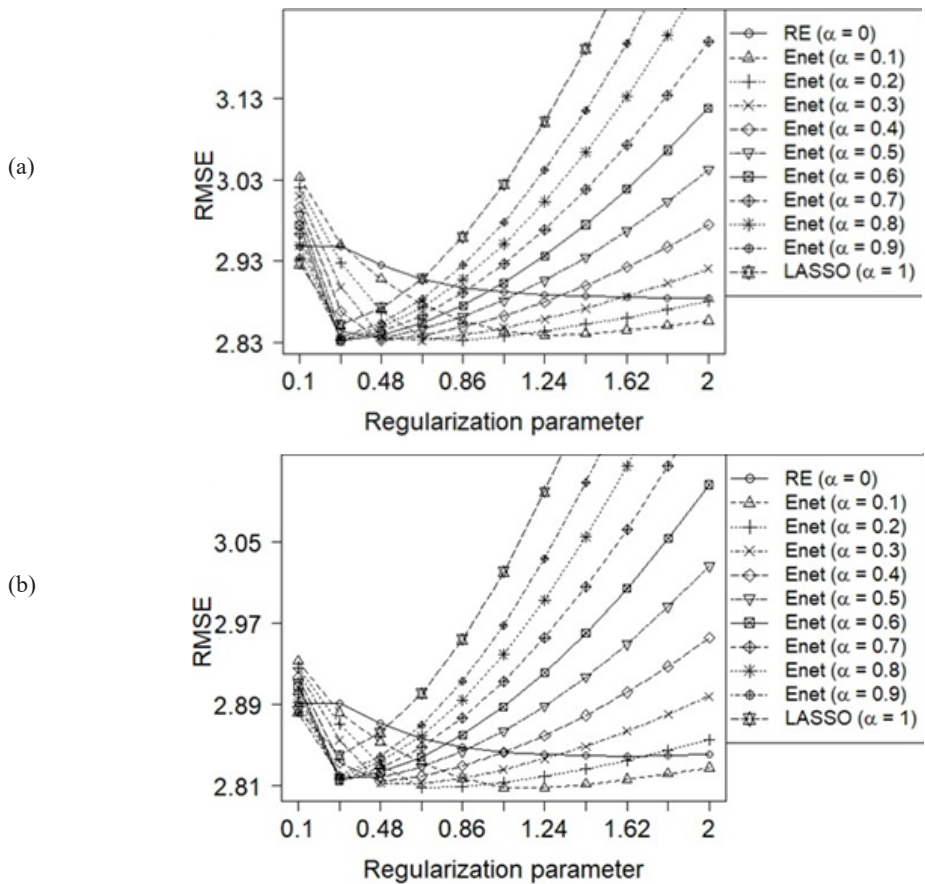
(a)

(b)

Figure 3: Estimated RMSE values the estimators when $\rho = 0.9$; (a) under the correctly specified model (b) under misspecified model.

Table 1: Optimal Regularization parameter values and estimated minimum RMSE of the simulation study.

		Correctly specified model			Misspecified model		
		Regularization parameter	RMSE	Number of selected variables	Regularization parameter	RMSE	Number of selected variables
$\rho = 0.7$	RE	2.00 $(\alpha = 0)$	2.898	20	2.00 $(\alpha = 0)$	2.838	14
	LASSO	0.29 $(\alpha = 1)$	2.858	9	0.29 $(\alpha = 1)$	2.840	8
	Enet	1.43 $(\alpha = 0.1)$	**2.809**	13	1.24 $(\alpha = 0.1)$	**2.770**	9
$\rho = 0.8$	RE	2.00 $(\alpha = 0)$	2.908	20	2.00 $(\alpha = 0)$	2.852	14
	LASSO	0.29 $(\alpha = 1)$	2.825	9	0.29 $(\alpha = 1)$	2.815	8
	Enet	0.86 $(\alpha = 0.2)$	**2.804**	10	1.24 $(\alpha = 0.1)$	**2.767**	9
$\rho = 0.9$	RE	2.00 $(\alpha = 0)$	2.885	20	1.62 $(\alpha = 0)$	2.839	14
	LASSO	0.29 $(\alpha = 1)$	2.851	6	0.29 $(\alpha = 1)$	2.839	6
	Enet	0.48 $(\alpha = 0.4)$	**2.832**	9	1.24 $(\alpha = 0.1)$	**2.808**	9

According to Table 1, we can observe that Enet estimator produces minimum RMSE in all cases considered in this study. Further, we observe different performances of RE, LASSO and Enet estimators when the model is correctly specified and misspecified.

Numerical example

The US crime dataset was considered to analyse the performance of Ridge, LASSO and Enet estimators. This data was used by Venables and Ripley (1999) to examine the effect of punishment regimes on crime rates, and it contains 16 variable with 47 observation. The data set is attached to the MASS library in R package, and it includes the following variables: M (percentage of males aged 14--24), So (indicator variable for a Southern state), Ed (mean years of schooling), $Po1$ (police expenditure in 1960), $Po2$ (police expenditure in 1959), LF (labor force participation rate), $M.F$ (number of males per 1000 females), Pop (state population), NW (number of non-whites per 1000 people), $U1$ (unemployment rate of urban males 14--24), $U2$ (unemployment rate of urban males 35--39), GDP (gross domestic product per head), $Ineq$ (income inequality), $Prob$ (probability of imprisonment), $Time$ (average time served in state prisons), y (rate of crimes in a particular category per head of population).

For the model fitting, y is considered as a dependent variable, the variable So is ignored because it is categorical. According to ANOVA of UScrime dataset, the p-values for the predictor variables

$M, Ed, Po1, Po2, LF, M.F, Pop, NW, U1, U2, GDP, Ineq, Prob$ and $Time$ are 0.26, 0.00, 0.00, 0.09, 0.47, 0.12, 0.33, 0.12, 0.89, 0.01, 0.35, 0.00, 0.024 and 0.62, respectively. Based on the ANOVA, we assume that the variables $U2, GDP, Ineq, Prob$ and $Time$ are missing and therefore the regression model is misspecified by omitting these variables. The Variance Inflation Factor (VIF) values of the regressor variables of the dataset are 2.86, 5.05, 104.58, 113.028, 2.88, 3.69, 2.53, 3.84, 5.19, 4.83, 9.97, 7.43, 2.75, and 2.66, which shows a moderated multicollinearity on the dataset.

The estimated RMSE values of the RE, LASSO and Enet estimators versus regularization parameter are displayed in Figure 4. The minimum RMSE values of the estimators and the optimal value of regularization parameter are summarized in Table 2.

From Figure 4, we can observe that Enet estimator outperforms other two estimators in both correctly specified and misspecified regression model. It is evident that the RMSE of RE, LASSO and Enet estimators show a significant difference performance when the model is correctly specified and misspecified.

According to Table 2, we can observe the superiority of the Enet estimator in both correctly specified and misspecified regression model. Further, we can note that LASSO and Enet select different number of variables.
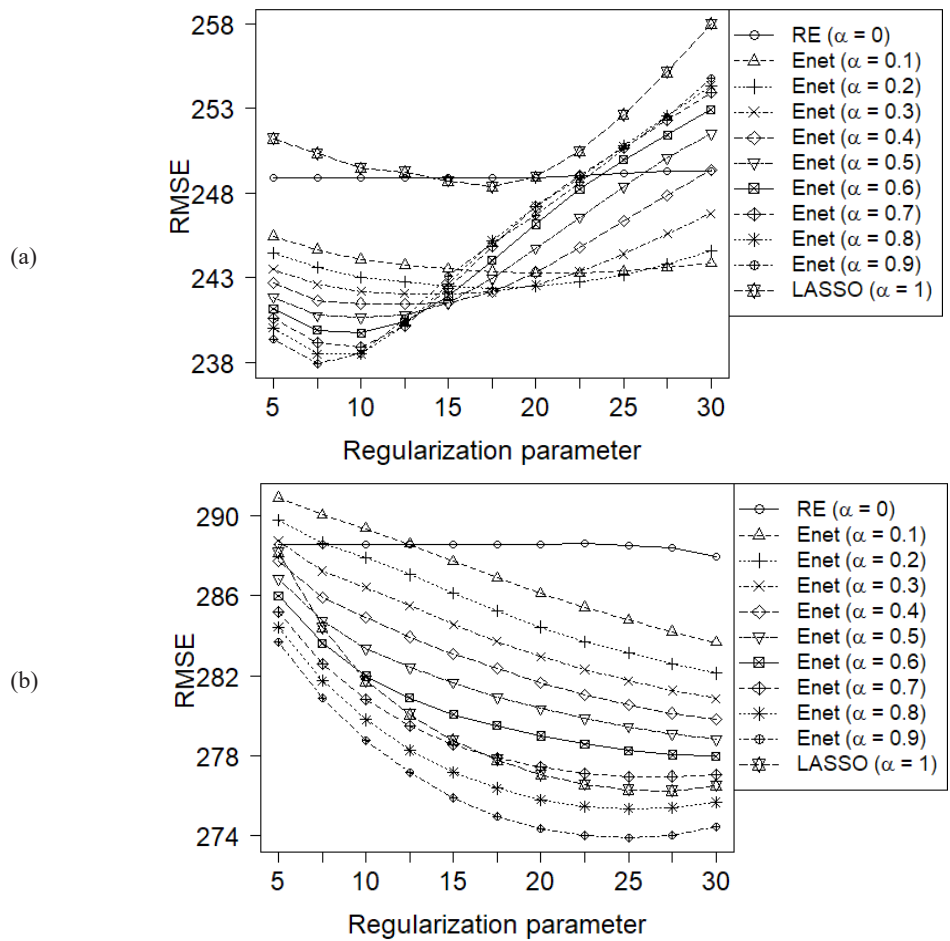
Figure 4: Estimated RMSE values of the estimators of US crime dataset; (a) under correctly specified model (b) under misspecified model.

Table 2: Optimal Regularization parameter values and estimated minimum RMSE of US crime dataset.

	Correctly specified model			Misspecified model		
	Regularization parameter	RMSE	Number of selected predictors	Regularization parameter	RMSE	Number of selected predictors
RE	20 ($\alpha = 0$)	248.93	14	30 ($\alpha = 0$)	287.96	9
LASSO	17.5 ($\alpha = 1$)	248.37	8	27.5 ($\alpha = 1$)	276.24	5
Enet	7.5 ($\alpha = 0.9$)	**237.90**	11	25 ($\alpha = 0.9$)	**273.88**	6

CONCLUSIONS

When comparing Enet with RE and LASSO it outperforms both in correctly specified and misspecified regression model when multicollinearity exists among the regressor variables. This study clearly shows the consequences of the performance of RE, LASSO and Enet estimators when the regression model is misspecified by excluding relevant explanatory variables.

REFERENCES

Efron, B., Hastie, T., Johnstone, B., and Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, **32**(2): 407-499.

Fu, W.J. (1998). Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics,* **7**(3): 397-416.

Hoerl, A., & Kennard, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics,* **12**: 55-67.

Oyeyemi, G.M., Ogunjobi, E.O., and Folorunsho, A.I. (2015). On Performance of Shrinkage Methods – A Monte Carlo Study. *International Journal of Statistics and Applications,* **5**(2): 72-76.

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

Venables, W.N., and Ripley, B.D. (1999). *Modern Applied Statistics with S-PLUS.* Springer.

Zou, H., and Hastie, T. (2005). Regression shrinkage and selection via the elastic net, with applications to microarrays. *Journal of the Royal Statistical Society Series B,* **67**: 301-320.