



Journal of Applied Economics and Business Studies (JAEBS)

Journal homepage: <https://pepri.edu.pk/jaeps>
ISSN (Print): 2523-2614
ISSN (Online) 2663-693X



Evaluation of Ridge, Elastic Net and Lasso Regression Methods in Precedence of Multicollinearity Problem: A Simulation Study

Shady I. Altelbany¹

¹. Faculty of Economic and administrative Sciences, Al Azhar University– Gaza, Palestine

Abstract

This study aims at performance evaluation of Ridge, Elastic Net and Lasso Regression Methods in handling different degrees of multicollinearity in a multiple regression analysis of independent variables using simulation data. The researcher simulated a collection of data with sample size $n=200, 1000, 10000, 50000$ and 100000 , independent variables $p=10$. The researcher compared the performances of the three methods using Mean Square Errors (MSE). The study found that Elastic Net method outperforms Ridge and Lasso methods to estimate the regression coefficients when a degree of multicollinearity is low, moderate and high for any sample size. While, Lasso method is the most accurate regression coefficients estimator when data containing severe multicollinearity at sample size less than 10000 observations.

Keywords

Ridge, Lasso, Elastic Net, Multicollinearity, Regression.

JEL

Classification
C02, C31, C63

1 Introduction

Multiple linear regression is frequently employed is appropriate in particular context to evaluate a model to predict the expected responses, or to explore the link between the dependent variable and the independent variables. The first goal, which is the design's prediction accuracy, is critical; however, the second goal, which is the model's complexity, is more important. Common linear regression procedures are popular for generally not carrying out well according to both prediction performance and model involvement (Doreswamy and Vastrad, 2013). There is a high number of hypotheses about the model in the regression analysis, specially, the most important one is (multicollinearity), in addition to (non-homogeneity of variance, autocorrelation and linearity). If one or more assumptions are broken, the model becomes unreliable,

and it is no longer suitable for estimating population parameters (Herawati et al., 2018).

When there is a close association or interaction between two or more independent variables, multicollinearity occurs in the study of multiple regression. Multicollinearity has the potential to produce inaccurate regression coefficient choices, increase regression coefficient standard errors, deflate partial t-tests for regression coefficients, produce wrong, non-significant p-values, and minimize model predictability. (Draper and Smith, 1998; Gujarati 1995).

The key issue with multicollinearity is that as the degree of collinearity rises, the coefficient estimates in the regression model become unsteady, and the standard errors for the coefficients become wildly maximized. Multicollinearity has two types; the first type is (full/perfect/exact multicollinearity and the second one is partially/less than perfect multicollinearity). The existence of the first type is when independent variables interrupted in a complete way. This means that no particular least squares solution to a multiple regression analysis can be computed under this condition. (Slinker, and Glantz 1985).

Since we know that multicollinearity is a serious problem when trying to make inferences or find predictive models, it's crucial to figure out the best way to deal with it. (Judge 1988). Multicollinearity can be detected using a variety of techniques and methods. Using pair-wise scatter plots of the independent variables, searching for near-perfect relationships, analyzing the correlation matrix for high interactions and the variance inflation factors (VIF), using eigenvalues of the correlation matrix of the independent variables, and testing the signs of the regression coefficients are just a few of the common approaches. (Montgomery and Peck, 1992; Kutner et al., 2005).

The reduce of variance at the cost of introducing a group of bias. The scholars call this method as “regularization or shrinkage methods” and is roughly beneficial for the predictive appearance of the model. In the study of current data, regularization is crucial. To overcome the shortcomings of ordinary least squares regression in terms of prediction precision, we introduced regularized regression methods for linear regression. Methods of regularization aid in the formalization of a unique solution to this well-posed problem. Any coefficients are reduced to zero using these techniques. This does not help with descriptor collection on its own, but it does reduce the gap at the expense of a small increase in bias. This form, on the other hand, improves the estimate's generalization. (Doreswamy and Vastrad, 2013). The least absolute shrinkage and selection operators Ridge, lasso, and Elastic Net are among the methods.

Using simulated data, this study examines three different regression methods to see which one works better for coping with multicollinearity obstacles.

2 Materials and Methods:

At first, we need to consider the basics of regression and what parameters of the equation they changed when using a distinct model. The relationship between a dependent variable and independent variables can be estimated using a multiple linear regression model and the Ordinary Least Square (OLS) method. If data comprises in n observations, $\{y_i, x_i\}_{i=1}^n$ each observation has a scalar response y_i and a vector of p independents x_{ij} for $j=1, \dots, p$, we can write a multiple linear regression model as:

$$Y = X \beta + \varepsilon \quad (1)$$

Where $Y_{n \times 1}$ is the dependent vector variable, $X_{n \times p}$ symbolizes the independent variables, $\beta_{p \times 1}$ is the set of regression coefficients that needs to be estimated, and $\varepsilon_{p \times 1}$ symbolizes the residuals.

$$\hat{\beta}^{ols} = (X'X)^{-1}X'Y \quad (2)$$

the regression coefficients are calculated using Ordinary Least Square by reducing the squared distances between the predicted dependent variable and the observed (Montgomery and Peck, 1992).

When building a regression model, the model becomes more complicated as the number of data and variables grows, and major optimization issues arise (Zou and Hastie, 2005). Furthermore, classical regression analysis fails when assumptions such as constant variance, multi-collinearity, and normality are not met (Ogutu et al., 2012). As a result, high coefficients in the model must be corrected, or penalized.

Regularized regression is a form of regression in which the coefficient estimates are constrained to zero. It penalizes the magnitude of the error term as well as the magnitude of the coefficients. Complex models are discouraged, mainly to prevent over-fitting. A typical least squares model has some flaws, such as the fact that it does not generalize well to data sets other than its training data. Regularization greatly reduces the model's variance while having little effect on its bias. The effect on bias and variance is dominated by the tuning parameter λ used in the regularization systems mentioned. As the value of λ increases, the rate of coefficients decreases, lowering the variance. To some extent, this increase in λ is advantageous because it only reduces variance (thus preventing over-fitting) while losing no significant properties in the results, but after a certain value, the model loses significant properties, resulting in bias and under fitting. Accordingly, the value of λ they should select it carefully (Biswas, 2019). There are three kinds of regularization systems called the Ridge, Lasso, and Elastic Net.

Ridge regression corrections are made with squared values, while Lasso regression corrections are made with absolute values. The ridge and Lasso biased estimation regression methods are combined in Elastic Net regression (Zou and Hastie 2003).

2.1 Ridge Regression:

It is obvious that Ordinary Least Square (OLS) is unsteady and presents estimates having a lot of variance when multicollinearity appears among independent variables, e.g. the columns of X are strikingly correlated. Hoerl et al (1975) develop ridge Regression and this approach is the adjustment of the least squares method, which allows for biased regression coefficient estimators. (Myers, 1986).

Ridge regression approach depends on adding a ridge parameter to the diagonal of $(X'X)$ matrix resulting a new matrix $(X'X + \lambda I)$. Since the diagonal of systems in the correlation matrix can be interpreted as a ridge, we call it ridge regression (Hoerl and Kennard, 2000). The ridge formula to find the coefficients is:

$$\hat{\beta}_{ridge} = (X'X + \lambda I)^{-1} X'Y, \lambda \geq 0 \quad (3)$$

When λ equal zero, the ridge estimator appears as the Ordinary Least Square (OLS). If they all λ 's are like each other, the estimators that resulted are called the ordinary ridge estimators (Hoerl, 1962; Hoerl et al., 1975). It is usually acceptable to edit ridge regression in Lagrangian form:

$$\hat{\beta}_{ridge} = \arg \min_{\beta} \left\{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \right\} \quad (4)$$

Where $\|y - X\beta\|_2^2 = \sum_{i=1}^n (y_i - x_i^T \beta)^2$ is the L_2 -norm (quadratic) loss function (i.e.

residual sum of squares), x_i^T is the row of X , $\|\beta\|_2^2 = \sum_{j=1}^p \beta_j^2$ is the L_2 - norm penalty

on β and $\lambda \geq 0$ is the tuning parameter (penalty, regularization) that controls the power of the penalty (linear shrinkage) by selecting the relative importance of the data-dependent practical error and the penalty term. The vaster the value of, the greater is the amount of shrinkage. Since the value of is reliant on the data, we can find it out using data-possessed techniques, includes; cross-validation (Doreswamy and Vastrad, 2013).

By constraining the coefficient estimates, Ridge regression can overpower this multicollinearity, as a result, it may decrease the estimator's variance while also introducing bias. (James et al., 2013).

2.2 Lasso Regression:

They broadly used lasso regression approaches in handling with big databases, such as those used in drug discovery, where efficient and quick algorithms are required (Hastie and Friedman, 2010) the Lasso estimator is also recognized as basis pursuit (Chen et al., 1998). Still, Because there are steep correlations between descriptors, Lasso will choose one and ignore the others, and when all descriptors are

the same, it will decrease. The Lasso penalty looks for many coefficients that are similar to zero, with only a small subset of them being the best (and not equal zero). To get a sparse solution to the following expansion problem, the Lasso estimator uses the L1 penalized least squares basis (Tibshirani, 1996).

$$\hat{\beta}_{Lasso} = \arg \min_{\beta} \left\{ \|y - X \beta\|_2^2 + \lambda \|\beta\|_1 \right\} \quad (5)$$

Where $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$ is the L₁-norm penalty on β , that causes the solution to become sparse, and $\lambda \geq 0$ is a parameter for fine tuning. Penalizing the absolute values of the coefficients introduces shrinkage towards zero, likewise ridge regression. In contrast, unlike ridge regression, It reduces certain coefficients to zero; such solutions have a large number of identically zero values. The penalty acts as a continuous variable selection tool (Herawati, 2018). The Lasso estimation method handles both the multicollinearity issue and best feature selection together in the high dimension linear regression model. Nonetheless, according to Hastie and Zou (2005) Lasso estimation procedure is unstable if the amount of predictors is greater than the amount of observations. Further, the prediction performance of RE dominates Lasso if there is high multicollinearity among predictors.

2.3 Elastic Net Regression:

According to (Friedman et al., 2010) This is a continuity of the Lasso that is robust to the strongest correlations among the predictor variables.. In order to prevent the imbalance of the Lasso solution paths when predictor variables are strongly correlated, they projected the Elastic Net for assessing high-dimensional data. Zou and Hastie (2005) recommended Elastic Net estimator by using a mixture of ridge and lasso, and it is:

$$\hat{\beta}_{Elastic\ Net} = \arg \min_{\beta} \left\{ \|y - X \beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \right\} \quad (6)$$

The regularization parameter λ is the sum of two nonnegative penalties $\lambda = \lambda_1 + \lambda_2$, Now, let $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ then $1 - \alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2}$, where $0 < \alpha < 1$. Further, it can be defined as:

$$\hat{\beta}_{Elastic\ Net} = \arg \min_{\beta} \left\{ \|y - X \beta\|_2^2 + \lambda \left[\alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|_2^2 \right] \right\} \quad (7)$$

Note that $\alpha = 0$, and then Elastic Net estimator in equation (7) is equivalent to Ridge. Similarly, $\alpha = 1$, and then Elastic Net estimator in equation (7) is equivalent to Lasso. If $\lambda = 0$, so using this method, the elastic net method, decreases to ordinary least squares regression.

Hence, we can write the Ridge, Lasso and Elastic Net estimator in a common form in the mis specified regression model as below:

$$\hat{\beta}_{(\alpha)} = \arg \min_{\beta} \left\{ \|y - X \beta\|_2^2 + \lambda \left[\alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|_2^2 \right] \right\} \quad (8)$$

Where

$$\hat{\beta}_{(\alpha)} = \begin{cases} \hat{\beta}_{Ridge} & , \alpha = 0 \\ \hat{\beta}_{Elastic\ Net} & , 0 < \alpha < 1 \\ \hat{\beta}_{Lasso} & , \alpha = 1 \end{cases}$$

The MSE, which is the scheduled prediction error of the estimators is given by:

$$MSE \left(\hat{\beta}_{(\alpha)} \right) = \frac{1}{n} \left(y_{new} - X_{new} \hat{\beta}_{(\alpha)} \right)' \left(y_{new} - X_{new} \hat{\beta}_{(\alpha)} \right) \quad (9)$$

where (y_{new}, X_{new}) includes new observations that are not used to obtain the coefficient estimates $\hat{\beta}_{(\alpha)}$.

In brief, the following are some salient distinctions between Lasso, Ridge and Elastic Net (Hastie et al., 2001):

- Lasso has a sparse selection, unlike Ridge which does not have.
- Ridge regression shrinks the two coefficients towards one another if we have extremely correlated variables. Furthermore, Lasso is neutral and picks one over the other. In terms of context, no one would know which variable was chosen. Elastic Net is an adjustment between the two which attempts to shrink and do a sparse selection at the same time.
- Ridge estimators are neutral to multiplicative scaling of the data. If constants multiplied both X and Y, the coefficients of the fit do not change for λ parameter. However, for Lasso, the fit is not separate from the scaling. In fact, the multiplier must scale the λ parameter up to get the same result. It is more complicated for Elastic Net.
- In a comparison with Lasso, Ridge penalizes the largest β 's rather than it penalizing the small ones (as it square them in the penalty term). Lasso penalizes the small ones more consistently. Sometimes, This is of no consequence. When faced with a forecasting issue involving a strong predictor, the Ridge shrinks the predictor's effectiveness as compared to the Lasso.

3 Results: Simulation Study

Using R package, we simulate the linear regression model for number of data $n = 200, 1000, 10000, 50000, 100000$ observations and 10 independent variables. To explore the implements of different grades of multicollinearity on the estimators, we

choose $\rho = (0.70, 0.80, 0.90, 0.99)$ which represent low, moderate, high multicollinearity and severe multicollinearity. (McDonald and Galarneau, 1975) generate the independent variables:

$$x_{ij} = (1 - \rho^2)^{1/2} u_{ij} + \rho u_{ij}, i = 1, 2, \dots, n \ \& \ j = 1, 2, \dots, p \quad (10)$$

Where u_{ij} are independent, standard normal pseudo-random numbers and ρ is fixed, so that any two independent variables' theoretical correlation is defined by ρ^2 .

Performance Assessment of ridge, lasso and elastic net approaches are compared on the basis of MSE value. Cross validation is a technique for determining a value for the λ value for Ridge, Lasso and Elastic Net, and are displayed in Tables 1-4.

The estimated MSE values of the ridge, lasso and elastic Net estimators versus regularization parameter when $\rho = 0.70, \rho = 0.80, \rho = 0.90$ and $\rho = 0.99$ and the optimal value of regularization parameters are summarized in Table 1-4.

We can observe in table 1-3, that Elastic Net was outperforms than Ridge and Lasso at $n = 200, 1000, 10000, 50000, 100000$ observations if degrees of multicollinearity are low, moderate, high.

In Table 4 we can show that Lasso method was outperforms than Ridge and Elastic Net when $\rho = 0.99$ (severe multicollinearity) at $n = 200, 1000, 10000$ observations, while at $n = 50000, 100000$ observations was showed that Elastic Net method was the best. Elastic Net method is the most accurate regression coefficients estimator.

Table 1 MSE values and optimal value of Regularization parameter when $\rho = 0.70$

$\rho = 0.7$											
Method	α	n=200		n=1000		n=10000		n=50000		n=100000	
		λ	MSE	λ	MSE	λ	MSE	λ	MSE	λ	MSE
Ridge	0	2.4943870	1.3542480	2.435132	1.77365	2.393143	2.071687	2.397843	1.90527	2.398618	1.925571
Elastic Net	0.1	0.8556920	0.8115550	0.5246333	1.09823	0.4280494	1.111337	0.3560721	1.026351	0.3561871	1.011122
	0.2	0.6812518	0.8255325	0.4584058	1.094004	0.3740143	1.111348	0.3414572	1.034078	0.3112236	1.013724
	0.3	0.4984487	0.8274622	0.3353999	1.086273	0.3003344	1.105096	0.2741909	1.030563	0.2499133	1.009879
	0.4	0.4502874	0.8562786	0.2760757	1.08463	0.2472124	1.10109	0.247698	1.034514	0.2257661	1.013563
	0.5	0.3953519	0.8766654	0.2919647	1.08985	0.2170523	1.101262	0.1981584	1.029075	0.1982224	1.014189
	0.6	0.3294599	0.8792247	0.2670257	1.091589	0.1808769	1.09714	0.1812321	1.031515	0.1651853	1.009974
	0.7	0.3099273	0.9037126	0.2085462	1.086522	0.1701533	1.100581	0.1704875	1.035552	0.155392	1.013765
	0.8	0.2251437	0.8662195	0.2002693	1.088843	0.1488841	1.098445	0.1359241	1.027507	0.135968	1.011472
	0.9	0.2410546	0.9089749	0.1953736	1.092268	0.1452446	1.103449	0.1208214	1.025793	0.1208604	1.008507
Lasso	1	0.2169491	0.9112325	0.1602154	1.088178	0.1307201	1.102091	0.1193412	1.031095	0.1087744	1.00979

Table 2 MSE values and optimal value of Regularization parameter when $\rho = 0.80$

$\rho = 0.8$											
Method	α	n=200		n=1000		n=10000		n=50000		n=100000	
		λ	MSE	λ	MSE	λ	MSE	λ	MSE	λ	MSE
Ridge	0	3.072037	2.377661	2.65657	1.925118	2.72915	2.103151	2.720972	2.013501	2.710478	2.047791
Elastic Net	0.1	0.9602314	1.001544	0.6281431	1.120764	0.4881492	1.111701	0.4040558	1.034427	0.4417403	1.02362
	0.2	0.76448	0.9836472	0.5000907	1.101727	0.4265273	1.110792	0.3874714	1.040867	0.3516877	1.016339
	0.3	0.5593439	0.9327243	0.401574	1.091794	0.375896	1.112307	0.3111404	1.03481	0.3099403	1.01822
	0.4	0.5052988	0.9567745	0.3305453	1.086529	0.281922	1.100099	0.2810773	1.03758	0.2551193	1.014054
	0.5	0.4436518	0.9626045	0.3495691	1.09122	0.2716609	1.10776	0.2467856	1.03770	0.2458337	1.022227
	0.6	0.3697098	0.9457097	0.3197097	1.091918	0.2062727	1.096673	0.2257057	1.039964	0.2048614	1.016942
	0.7	0.3477909	0.9656147	0.2496922	1.085563	0.2129625	1.107571	0.2123244	1.044483	0.1927159	1.020984
	0.8	0.2772824	0.9290207	0.2184807	1.08461	0.1863422	1.105263	0.1692793	1.034339	0.1686264	1.018251
	0.9	0.2968779	0.9840657	0.2131398	1.086901	0.1817869	1.111765	0.1504705	1.031538	0.1498901	1.01628
Lasso	1	0.2434537	0.9478264	0.1918258	1.086458	0.1636082	1.110342	0.1354234	1.032729	0.1349011	1.014802

Table 3 MSE values and optimal value of Regularization parameter when $\rho = 0.90$

$\rho = 0.9$											
Method	α	n=200		n=1000		n=10000		n=50000		n=100000	
		λ	MSE	λ	MSE	λ	MSE	λ	MSE	λ	MSE
Ridge	0	3.3935400	2.7443080	3.10191	2.095804	3.123437	2.490004	3.105966	2.234529	3.098342	2.260106
Elastic Net	0.1	1.0607240	1.1089750	0.8834351	1.183933	0.6131434	1.159529	0.4612261	1.037333	0.4600939	1.024278
	0.2	0.7694645	0.9855784	0.771914	1.164514	0.5357428	1.147965	0.4422951	1.041303	0.4020136	1.020438
	0.3	0.6781246	0.9557141	0.6198487	1.136541	0.4721469	1.141938	0.3897921	1.040322	0.3888352	1.028375
	0.4	0.5581806	0.9200789	0.6145527	1.144592	0.3886355	1.128757	0.3208472	1.034618	0.3200596	1.021084
	0.5	0.5378645	0.9313477	0.5921847	1.154328	0.3744903	1.137142	0.2817036	1.034401	0.3084104	1.029477
	0.6	0.4482204	0.9077476	0.4934872	1.135276	0.3120752	1.126321	0.2576411	1.037091	0.2570087	1.022753
	0.7	0.4216468	0.9163924	0.46423	1.140038	0.2935733	1.13074	0.2659968	1.050992	0.2417714	1.027734
	0.8	0.3361653	0.8828899	0.4062012	1.130843	0.2819218	1.138457	0.2120706	1.039766	0.21155	1.024674
	0.9	0.3599220	0.9240131	0.3610678	1.120571	0.2505971	1.13428	0.1885072	1.03851	0.1880445	1.022657
Lasso	1	0.2951528	0.8923225	0.324961	1.124729	0.2255374	1.131318	0.1861978	1.045797	0.1857407	1.030667

Table 4 MSE values and optimal value of Regularization parameter when $\rho = 0.99$

$\rho = 0.99$											
Method	α	n=200		n=1000		n=10000		n=50000		n=100000	
		λ	MSE	λ	MSE	λ	MSE	λ	MSE	λ	MSE
Ridge	0	3.654888	3.0121850	3.546463	3.483199	3.585424	4.344516	3.557993	3.859463	3.544116	3.739378
Elastic Net	0.1	1.040925	1.0764300	0.8385574	1.334766	0.7038333	1.387571	0.6984484	1.251602	0.6339181	1.216287
	0.2	0.9982006	1.0413740	0.8825415	1.313373	0.7407508	1.391776	0.7350835	1.251386	0.6671684	1.224578
	0.3	0.8015574	0.9893414	0.853611	1.305417	0.7164684	1.361745	0.7109868	1.25647	0.7082139	1.246001
	0.4	0.7241091	0.9845812	0.8463177	1.295802	0.7796048	1.379147	0.6422896	1.222855	0.7021628	1.25335
	0.5	0.6357671	0.9766932	0.8155141	1.300482	0.6844922	1.35112	0.6189121	1.223221	0.676606	1.227457
	0.6	0.5814613	0.9817787	0.7458548	1.274246	0.6870611	1.359158	0.6212348	1.237149	0.6188119	1.221965
	0.7	0.4983954	0.9641743	0.7016354	1.266754	0.6463274	1.353982	0.5844038	1.235244	0.5821245	1.227421
	0.8	0.4786148	0.9831136	0.6737885	1.272795	0.6206756	1.358227	0.5612097	1.244024	0.5590209	1.236977
	0.9	0.4669148	1.0107910	0.6573174	1.281283	0.5517117	1.322116	0.5474906	1.260935	0.5453553	1.257195
Lasso	1	0.3828918	0.9524901	0.5915857	1.242952	0.4965405	1.296512	0.4927415	1.244684	0.4908198	1.24159

4 Conclusion

According to the outcomes of simulation at $p = 10$ and $n = 200, 1000, 10000, 50000, 100000$ observations, containing different degrees of multicollinearity within all independent variables, it can be summarized in:

- Elastic Net method outperforms Ridge and Lasso methods to estimate the regression coefficients when degree of multicollinearity is low ($\rho=0.70$), moderate ($\rho=0.80$) and high ($\rho=0.90$) at all number of data ($n = 200, 1000, 10000, 50000, 100000$ observations).
- If data including cruel multicollinearity within all independent variables, the Lasso method outperforms Ridge and Elastic Net methods at ($n = 200, 1000, 10000$ observations).
- Elastic Net methods for the larger number of observations ($n = 50000, 100000$ observations) was outperforms Ridge and Lasso methods if data containing severe multicollinearity.
- Results suggest that performance of these methods are depending greatly on the values of α .
- Elastic net regression outperformed the other two methods in case of low, moderate and high level of multicollinearity when $0 < \alpha < 1$. Also, it can be incurred that severe multicollinearity requires higher value ($\alpha=1$). This is consistent with the theoretical framework.
- Ridge method was unsuitable regression coefficients estimator compared with Lasso and Elastic Net methods.
- General, in studying relationships and interconnecting economic and social factors, we recommend to the decision maker that use Elastic Net method for any sample size.
- We also recommend to the decision maker that use Lasso method for when using real data, and examine the relationships between the different variables (severe multicollinearity) at sample size less than 10000 observations.

References:

- Biswas, S. (2019). How Regularization Helps in Data Overfitting. <https://medium.com/towards-artificial-intelligence/how-regularization-can-help-in-overfitting-the-data-ad9ff80f9ccc>.
- Chen, S. S., Donoho, D. & Saunders, M. (1998). Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1), 33–61. <https://doi.org/10.1137/S1064827596304010>
- Doreswamy, A., & Vastrad, C. M. (2013). Performance Analysis of Regularized Linear Regression Models for Oxazolines and Oxazoles Derivatives Descriptor Dataset. *International Journal of Computational Science and Information Technology (IJCSITY)*, 1(4), 111-123. DIO: 10.5121/ijcsity.2013.1408.
- Draper, N.R., & Smith, H. (1998). *Applied Regression Analysis*. 3rd ed., New York: Wiley. <https://doi.org/10.1002/9781118625590>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent, *Journal of Statistical Software*, 33(1),1-22. DOI: [10.1163/ej.9789004178922.i-328.7](https://doi.org/10.1163/ej.9789004178922.i-328.7)
- Gujarati, D. (1995). *Basic Econometrics*. 4th ed., New York: McGraw–Hill.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., Springer-Verlag New York Inc. <https://www.springer.com/gp/book/9780387848570>
- Herawati, N., Nisa, K., Setiawan, E., & Nusyirwan, T. (2018). Regularized multiple regression methods to deal with severe multicollinearity. *International Journal of Statistics and Applications*, 8(4), 167-172. DOI: [10.5923/j.statistics.20180804.02](https://doi.org/10.5923/j.statistics.20180804.02)
- Hoerl, A.E., & Kennard, R.W. (2000). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 42, 80-86. <https://doi.org/10.2307/1271436>
- Hoerl, A.E. (1962). Application of ridge analysis to regression problems. *Chem. Eng. Prog.*, 58, 54-59.
- Hoerl, A., Kannard, R., & Baldwin, K.F. (1975). Ridge regression: Some simulations. *Communication of Statistics*, 4, 105-123. <https://doi.org/10.1080/03610927508827232>

- James, G., Witten D., Hastie T., & Tibshirani R. (2013). *An Introduction to Statistical Learning: With Applications in R*. New York: Springer Publishing Company, Inc. <https://link.springer.com/book/10.1007/978-1-4614-7138-7>
- Judge, G.G. (1988). *Introduction to Theory and Practice of Econometrics*. New York: John Willy and Sons. <https://doi.org/10.1002/jae.3950050311>
- Kutner, M. H., Nachtsheim, C., Neter, & William, N. (2005). *Applied Linear Statistical Models*. 5th Edition. New York: McGraw-Hill. <https://www.amazon.com/Applied-Linear-Statistical-Models-Michael/dp/007310874X>
- McDonald G.C., & Galarneau, D.I. (1975). A Monte Carlo evaluation of some ridge type estimators. *J. Amer. Statist. Assoc.*, 20, 407-416. <https://www.tandfonline.com/doi/abs/10.1080/01621459.1975.10479882>
- Montgomery, D.C. & Peck, E.A. (1992). *Introduction to Linear Regression Analysis*. New York: John Willy and Sons. <https://doi.org/10.1111/biom.12129>
- Myers, R. H. (1986). *Classical and modern regression with applications*, 2nd Ed, USA: PWSKENT Publishing Company. <https://lib.ugent.be/catalog/rug01:000851135>
- Ogutu, J. O., Schulz-Streeck, T., & Piepho, H. P. (2012). Genomic selection using regularized linear regression models: Ridge regression, Lasso, Elastic Net and their extensions. *BMC proceedings*, 6 (2), 10. <https://doi.org/10.1186/1753-6561-6-S2-S10>
- Slinker, B.K., & Glantz, S.A. (1985). Multiple regression for physiological data analysis: the problem of multicollinearity. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 249(1), R1–R12. <https://doi.org/10.1152/ajpregu.1985.249.1.R1>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal Royal. Statist. Soc B.*, 58(1), 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of The Royal Statistical Society: Series B (Statistical Methodology)*, 67 (2), 301-320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>