Research Journal of Mathematical and Statistical Sciences ______________________________ISSN 2320-6047
Vol. **13(3),** 1-13, September **(2025)** *Res. J. Mathematical and Statistical Sci.*

New Ridge Parameters for Generalized Ridge Regression

A. V. Dorugade
Y.C. Mahavidyalaya, Halkarni, Tal- Chandgad, Kolhapur, MS - 416552, India
adorugade@rediffmail.com

Available online at: www.isca.in , www.isca.me
Received 19th April 2025, revised 30th May 2023, accepted 8th July 2025

Abstract

In Classical Linear Regression, in the presence of multicollinearity there are several methods to get rid of this problem and one of the most famous one is the ridge regression. In this paper, a new procedure to estimate the ridge parameters for Generalized Ridge Regression (GR) estimator is presented. We compared our estimators with other 30 estimators proposed elsewhere earlier according to mean squared error (MSE) criterion. All results are calculated by a Monte Carlo simulation. According to simulation study, our estimators perform equivalently to the estimator under the influence of multicollinearity. Also perform better than the others in the sense of MSE.

Keywords: Multicollinearity, Generalized ridge estimator, Mean square error.

Introduction

Regression analysis is one of the most common methods of analysis and modeling recognized for its practicality in forecasting and estimation. It is useful in several ways. The ordinary least-squares (LS) method is often favored for parameter estimation in regression analysis due to its appealing mathematical properties and computational ease. The estimator under certain assumptions has some very attractive statistical properties which have made it one of the most powerful estimators. Consider, widely used linear regression model

$$Y = X\beta + \varepsilon , \qquad (1)$$

where Y is a n×1 random vector of response variables, ***X*** is a known n×p matrix with full column rank, ε is the vector of errors E(ε) = 0 and Cov(ε) = $\sigma^2 I_n$. β is a p×1 vector of unknown regression parameters and σ^2 is the unknown variance parameter. For the sake of convenience, we assume that the matrix X and response variable Y are standardized in such a way that $X'X$ is a non-singular correlation matrix and $X'Y$ is the correlation between X and Y.

Let $\wedge$ and T be the matrices of eigen values and eigen vectors of $X'X$, respectively, satisfying $T'X'XT = \wedge$ = diagonal (λ_1, λ_2,..., λ_p), where λ_i being the ith eigen value of $X'X$ and $T'T = TT' = I_p$. We obtain the equivalent model[2]

$$Y = Z\alpha + \varepsilon , \qquad (2)$$

where Z = XT , it implies that $Z'Z = \wedge$, and α = $T'\beta$

Then LS estimator of α is given by

$$\hat{\alpha}_{LS} = (Z'Z)^{-1}Z'Y = \wedge^{-1} Z'Y . \qquad (3)$$

Therefore, LS estimator of β is given by

$$\hat{\beta}_{LS} = T\hat{\alpha}_{LS} .$$

Assumption that covariates are exogenous to the error, or in the case of fixed covariates, if the error has zero mean. This assumption, however, is violated if there is a close-linear relationship between the regressors, leading to the "multicollinearity". Indicators of multicollinearity include a very high correlation among two or more explanatory variables, a very small (near zero) eigen values of the correlation matrix of explanatory variables, a too large condition numbers. The separate effects of these variables may be confounded. Many techniques for dealing with multicollinearity have been developed, including adding more observations to the sample, principal component analysis, factor analysis, and ridge regression (RR). Ridge regression is a biased estimation technique and was developed as an alternative to LS, which is unbiased but prone to produce high variance in coefficient estimations. This would render misleading statistical inferences.

Ridge regression[1] defines a class of estimators indexed by a biasing parameter k. As 'k' increases from zero and continues up to infinity, the regression estimates tend toward zero. Though these estimators result in biased, for certain value of k, they yield minimum mean squared error (MMSE) compared to the LS estimator. This approach prevents the overinflating of regression coefficient variances, enhancing the reliability of statistical inferences regarding the coefficients.

RR estimator is given as:

$$\hat{\alpha}_{RR} = \left[I - k\left(\wedge + kI\right)^{-1}\right]\hat{\alpha}_{LS} \qquad (4)$$

Therefore, RR estimator of β is given by

$\hat{\beta}_{RR} = T\hat{\alpha}_{RR}$

and mean square error of $\hat{\alpha}_{RR}$ is

$$\text{MSE}(\hat{\alpha}_{RR}) = \hat{\sigma}^2 \sum_{i=1}^{p} \lambda_i / (\lambda_i + k)^2 + k^2 \sum_{i=1}^{p} \hat{\alpha}_i^{\,2} / (\lambda_i + k)^2 \quad (5)$$

We observe that, when $k = 0$ in (5), MSE of LS estimator of α is recovered. Hence

$$\text{MSE}(\hat{\alpha}_{LS}) = \hat{\sigma}^2 \sum_{i=1}^{p} 1/\lambda_i \quad (6)$$

In a ridge regression an additional parameter, the ridge parameter (k) plays an important role to control the bias of the regression towards the mean of the response variable. The value of ‘ k ’ is chosen small enough, for which the mean squared error of ridge estimator, is less than the mean squared error of LS estimator[3]. They shown that $k(HKB) = (p\hat{\sigma}^2 / \hat{\alpha}'\hat{\alpha})$ is used as an estimate of k, then significant improvement over LS is obtained in MSE properties of the estimate of ‘β’. Ridge parameter[3] k perform fairly well. The challenge is to determine a best choice for the ridge parameter k. Much of the discussions on ridge regression concern the problem of finding good empirical value of k in the statistical literature. However, there is not an explicit way of estimating k. Most recently, highlights the lack of a robust ridge parameter estimator[4] that can be applied in practice and perform well in case of real-life data with a wide range of p, n, ρ, and σ values. They considered 366 estimators for estimating ridge or shrinkage parameter k for regression models.

In addition, generalized ridge regression[1] (GR) procedure that allows separate ridge parameter for each regressor. Using GR, it is easier to find optimal values of ridge parameter i.e., values for which the MSE of the ridge estimator is minimum. In addition, if the optimal values for biasing constants differ significantly from each other then this estimator has the potential to save a greater amount of MSE than the LS estimator.

In the classic setting when $n > p$, it is well known that the ridge estimator uniformly shrinks the LS, thus reducing its squared-length relative to the LS. This variance reduction enables the ridge estimator to outperform the LS in correlated settings. However, the ridge estimator shrinks all coefficients uniformly to zero, which is effective only when all coefficients are small. On the other hand, GR generalizes ridge estimation by using a unique parameter k for each coefficient β, which allows GR to achieve non-uniform shrinkage, thus making it feasible to selectively shrink coefficients to zero, similar to the lasso. In the classic setting $n > p$, it has been shown that GR estimators can achieve oracle properties[5]. Although it is unlikely they envisioned using GR in problems where p could be larger than n, it is natural to wonder if it can be applied effectively in such contexts.

The GR estimator of α is defined by

$$\hat{\alpha}_{GR} = \left(I - KA^{-1}\right)\hat{\alpha}, \quad (7)$$

where K = diagonal $(k_1, k_2 \ldots k_p), k_i \geq 0$, i=1,2,...,p be the different ridge parameters for different regress or and A = $\wedge$ + K.

Hence, GR estimator for β is $\hat{\beta}_{GR}$ = $T\hat{\alpha}_{GR}$ and mean square error of $\hat{\alpha}_{RR}$ is MSE ($\hat{\alpha}_{GR}$) = Variance ($\hat{\alpha}_{GR}$) + [Bias ($\hat{\alpha}_{GR}$)]^2

$$= \hat{\sigma}^2 \sum_{i=1}^{p} \lambda_i / (\lambda_i + k_i)^2 + \sum_{i=1}^{p} k_i^2 \hat{\alpha}_i^{\,2} / (\lambda_i + k_i)^2 \quad (8)$$

Setting $k_1 = k_2 = ... = k_p = k$ and $k \geq 0$, the MSE ($\hat{\alpha}_{RR}$) is recovered.

In case of GR, various methods are available in the literature to determine the separate ridge parameter for each regressor (Appendix).

Initially, the purpose of this paper is to judge the performance of the different methods of estimating ridge parameters in GR. Also, to develop generalized ridge regression by estimating the new ridge parameters. Moreover, we suggested methods for estimating ridge parameters in GR which produces ridge estimators mostly yields minimum MSE than most of the other estimators to be included in this study. In order to be able to evaluate the performance of the different methods of estimating ridge parameters we calculate the MSE for real data as well as simulated data using Monte Carlo simulations. With setting when $n > p$ as well as $n < p$, we chose to vary the sample size, the number of explanatory variables, error variance and the degree of correlation. In addition, we evaluate its performance when model defined in linear regression exhibits with not only multicollinearity but also heteroscedastics and/or correlated errors, non-normal errors and outliers, respectively.

Proposed Ridge Parameters

T the values of k_i which minimizes the MSE of GR[1] are given by $k_i = \sigma^2 / \alpha_i^{\,2}$, $i = 1,2,..., p$. Also, propose estimates of k_i by replacing the unknown quantities by their LS estimates given as $k_i(HK) = \hat{\sigma}^2 / \hat{\alpha}_i^{\,2}$, $i = 1,2,..., p$ where, $\hat{\alpha}_i$ is the i[th] element of $\hat{\alpha}_{LS}$, $i = 1,2,\ldots,p$ and $\hat{\sigma}^2$ is the LS estimator of σ^2 i.e. $\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}_{LS}'Z'Y}{n-p-1}$. $k_i(HK)$ is superior to all estimators within the class of biased estimators. However the k_i have no finite moments of any order and therefore may not yield a GR estimator, with good MSE properties. Iterative procedure[1] to estimate k_i, and explicit closed form[6] solution for the procedure. There are some exact finite sample properties for GR estimators. Approach[7] to estimate σ^2/α_i^2 as a single quantity rather than to estimate σ^2 and α_i separately. It is given by

$$k_i(F) = \lambda_i\hat{\sigma}^2 / \left[\lambda_i\hat{\alpha}_i^{\,2} + (n-p-1)\hat{\sigma}^2\right] \quad i = 1,2,..., p$$

It is well known that eigen values of $(X'X)$ are also used to detect multicollinearity among the various explanatory variables. Most of the researchers including[7] suggested ridge parameters k_i as the modification in $k_i(HK)$ based on these eigen value λ_i to keep the variation depends on the strength of the multicollinearity.

With above applications of LS estimate of 'α' and eigen values of $(X'X)$, we introduce a new approach to determine ridge parameters k_i in GR.

We denote our ridge parameters by $k_i(\lambda)$ given by

$$k_i(\lambda) = \frac{p\hat{\sigma}^2}{\lambda_{\max}\,\hat{\alpha}_i^{\,2}},\ i = 1,2,..., p$$

where, $\hat{\alpha}_i$ is the i[th] element of $\hat{\alpha}_{LS}$ and $\lambda_{\max} = \max(\lambda_i)$ $i = 1,2,\ldots,p$. λ_i being the i[th] eigen value of $(X'X)$.

Evaluation of the Performances of the Proposed and Review Estimators

Numerical Example: In this section, we demonstrate the performance of proposed estimator by considering numerical example; we use Hald cement data[2]. The Table-1 gives the estimated values of the regression coefficients using LS, RR and GR estimators. Also gives their MSE using equations (6), (5) & (8) respectively. Here we noted that using GR the estimated values of the regression coefficients and MSE ratios are represented only for $k_i(HK)$[1], $k_i(MU)$[8] and $k_i(\lambda)$ because these values for remaining choice of k_i having less importance for the comparative study.

From Table-1, we can see that RR as well as GR estimators are superior to LS estimator at $k_i(HK)$, $k_i(MU)$ and $k_i(\lambda)$. GR estimators at $k_i(HK)$, $k_i(MU)$ and $k_i(\lambda)$ are superior to LS estimator. Particularly, GR estimators at $k_i(HK)$, gives better performance than others.

Simulation Study: Consequently, parameter estimation methods are based on LS estimators do not assure the desirable results in the presence of multicollinearity. We evaluate the performance of the reviewed and proposed estimator compare to LS estimators using Monte Carlo Simulation study. We evaluate the performance of our suggested ridge parameters by considering following different situations in linear regression when data exhibits with multicollinearity.

Case II: Data generated using heteroscedastic errors.

Case III: Data generated using outlier observations.

Case IV: Data generated using outlier observations and heteroscedastic errors.

Case V: Data generated using non-normal errors.

Case I: Data generated using normal errors: We examined the ratios of average MSE (AMSE) of $\hat{\beta}_{GR}$ over the estimators for different degrees of multicollinearity. We consider the true model as $Y = X\beta + \varepsilon$. The explanatory variables are generated[9] by

$$x_{ij} = (1-\rho^2)^{1/2}u_{ij} + \rho u_{ip}, \quad i = 1,2,...,n \qquad j = 1,2,..., p.$$

Where, u_{ij} are independent standard normal pseudo-random numbers and ρ is specified so that the theoretical correlation between any two explanatory variables is given by ρ^2. In this study, we investigate the effects of different degrees of multicollinearity on the estimators.

Here, we investigate the performance of the suggested ridge parameters $k_i(\lambda)$ compare to ridge parameters with reference to GR listed in Appendix. Estimators $\hat{\beta}_{LS}$ and $\hat{\beta}_{GR}$ are computed and obtained the average MSE (AMSE) of estimators using the following expression. The experiment is repeated 3000 times.

$$AMSE(\hat{\beta}) = \frac{1}{3000}\sum_{i=1}^{p}\sum_{j=1}^{3000}(\hat{\beta}_{ij} - \beta_i)^2 .$$

where, $\hat{\beta}_{ij}$ denote the estimator of the i[th] parameter in the j[th] replication and β_i, $i = 1,2,\ldots,p$ are the true parameter values. We computed the AMSE ratios (AMSE ($\hat{\alpha}_{LS}$) / AMSE ($\hat{\alpha}_{GR}$)) of LS estimator over different estimators for various values of ρ, p, n, σ^2. We consider the method that leads to the maximum AMSE ratio to the best from the MSE point of view.

Case-I: Data generated using normal errors: Here 'ε' follows a normal distribution $N(0, \sigma^2 I_n)$.

Part A: Here we consider predictor variables p = 4 and ρ = 0.95. We have simulated the data with sample sizes n = 10, 80 and 500. The variance of the error terms is taken as σ^2 = 1, 8, 25 and 100. We computed the AMSE ratios (AMSE ($\hat{\alpha}_{LS}$) / AMSE ($\hat{\alpha}_{GR}$)) and reported in Table-2.

Table-1: Values of estimates and MSE.

Estimator	Ridge Parameter	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	MSE	MSE ratios over LS
LS	-	1.5511	0.5102	0.1019	-0.1441	1.37	1
RR	$k(HKB)$ = 0.0131	1.2995	0.2999	-0.1420	-0.3486	0.1488	9.2070
GR	$k_i(HK)$	0.0991	-0.2593	0.009	-0.5891	0.1470	9.3238
	$k_i(MU)$	0.01038	-0.2617	0.0166	-0.5891	0.1479	9.2616
	$k_i(\lambda)$	0.0580	-0.2379	0.0005	-0.5881	0.1502	9.1282

Table-2: Ratio of AMSE of LS over various GR estimators for different 'k_i' (p=4, ρ = 0.95).

k_i	$n=10$				$n=80$				$n=500$			
	$\sigma^2=1$	8	25	100	$\sigma^2=1$	8	25	100	$\sigma^2=1$	8	25	100
$k_i(1)$	4.129	5.866	5.621	7.073	3.286	4.708	5.536	5.845	3.152	4.022	4.939	5.342
$k_i(2)$	1.150	2.268	1.829	2.906	1.248	1.670	2.455	2.358	1.500	1.504	1.949	2.323
$k_i(3)$	2.001	2.021	2.016	2.005	2.000	2.001	2.004	2.001	2.000	2.000	2.001	2.001
$k_i(4)$	2.00	2.000	2.010	2.000	2.001	2.000	2.000	2.020	2.001	2.000	2.020	2.00
$k_i(5)$	3.019	4.818	4.506	5.873	2.111	3.585	4.507	4.725	1.964	2.777	3.830	4.322
$k_i(6)$	3.481	4.329	4.204	4.710	2.963	3.781	4.192	4.289	2.883	3.384	3.901	4.109
$k_i(7)$	2.549	2.697	2.676	2.731	2.040	2.052	2.053	2.053	2.006	2.008	2.008	2.008
$k_i(8)$	0.178	0.117	0.077	0.007	1.866	0.019	0.048	0.011	2.045	0.494	0.158	0.105
$k_i(9)$	0.432	2.004	1.612	2.550	0.036	0.973	1.959	1.879	0.005	0.285	1.163	1.824
$k_i(10)$	0.143	3.344	2.405	3.515	0.009	0.585	2.710	3.531	0.001	0.086	0.682	1.850
$k_i(11)$	1.310	2.810	2.318	3.549	0.688	2.107	3.164	3.079	0.561	1.319	2.432	3.016
$k_i(12)$	0.133	1.881	1.611	2.548	0.008	0.502	1.799	1.879	0.001	0.083	0.688	1.803
$k_i(13)$	4.127	5.831	5.594	7.025	3.132	4.027	4.367	4.450	2.485	2.623	2.675	2.693
$k_i(14)$	0.117	1.761	1.600	2.542	0.008	0.393	1.542	1.843	0.001	0.071	0.525	1.641
$k_i(15)$	1.312	2.815	2.322	3.551	0.699	2.227	3.327	3.192	0.586	1.611	2.956	3.513
$k_i(16)$	0.133	1.895	1.614	2.558	0.008	0.497	1.753	1.878	0.001	0.082	0.641	1.718
$k_i(17)$	3.925	5.464	5.253	6.559	3.179	4.446	5.148	5.450	3.052	3.867	4.645	4.978
$k_i(18)$	0.117	1.722	1.588	2.488	0.008	0.393	1.562	1.817	0.001	0.071	0.531	1.702

$k_i(19)$	0.117	1.628	1.546	2.395	0.008	0.385	1.462	1.701	0.001	0.071	0.516	1.618
$k_i(20)$	0.715	2.162	1.724	2.755	0.183	1.273	2.259	2.143	0.104	0.573	1.523	2.123
$k_i(21)$	0.126	1.839	1.629	2.568	0.008	0.456	1.748	1.908	0.001	0.078	0.624	1.831
$k_i(22)$	0.117	1.701	1.588	2.501	0.008	0.388	1.523	1.826	0.001	0.071	0.524	1.678
$k_i(23)$	1.564	3.070	2.565	3.858	0.979	2.489	3.551	3.475	0.851	1.707	2.840	3.400
$k_i(24)$	0.129	1.852	1.608	2.545	0.008	0.472	1.748	1.873	0.001	0.079	0.646	1.782
$k_i(25)$	2.200	2.660	2.630	2.720	2.008	2.047	2.052	2.052	2.001	2.007	2.008	2.008
$k_i(26)$	0.117	1.782	1.601	2.536	0.008	0.393	1.553	1.845	0.001	0.071	0.524	1.641
$k_i(27)$	1.983	3.036	2.499	3.817	1.768	3.112	3.709	3.624	1.562	2.524	2.792	2.851
$k_i(28)$	0.132	1.864	1.604	2.549	0.008	0.493	1.722	1.859	0.001	0.082	0.638	1.703
$k_i(29)$	3.261	3.955	3.869	4.328	2.831	3.518	3.839	3.948	2.756	3.211	3.612	3.768
$k_i(30)$	0.128	1.724	1.585	2.534	0.009	0.430	1.544	1.837	0.001	0.082	0.572	1.663
$k_i(\lambda)$	4.126	5.859	5.617	7.068	3.285	4.706	5.532	5.841	3.151	4.021	4.936	5.338

Above same procedure carried out for case II to V, and results are reported in Tables: from Table 3 to Table 6 respectively.

Part B: In the same vein, for different choice of β parameter vectors chosen arbitrarily for different combinations of p = 4,7,15; ρ = 0.99, 0.999; n = 20, 100, 400,1000 and σ^2 = 1, 8, 25. There are 72 set of (p, ρ ,n, σ^2) values. These are arranged as (4,0.99,20,1), (4,0.99,20,8), …, (15,0.999,1000,25) and it is numbered as 1, 2,…, 72 respectively. AMSE ratios only for $k_i(HK), k_i(MU)$ and $k_i(\lambda)$ are computed and represented in Figure-1.

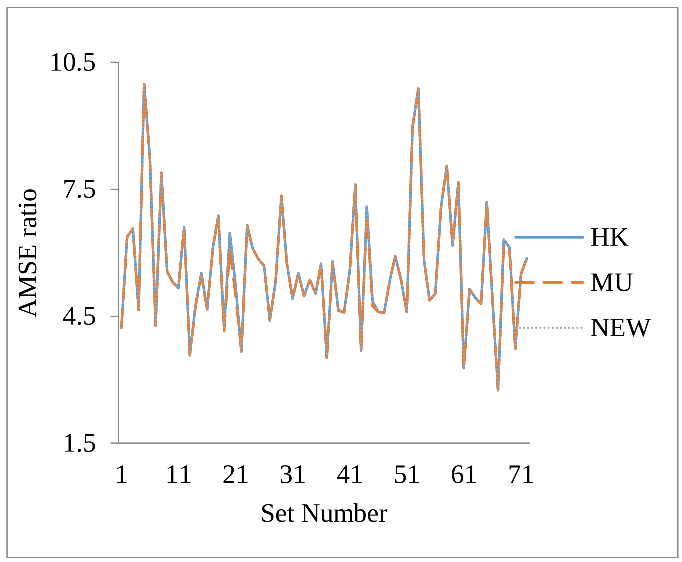

Figure-1: Ratio of AMSE of LS over various GR estimators.

Above same procedure carried out for case II to V, and results are represented in Figures: from Figure-2 to Figure-5 respectively.

Case II: Data generated using heteroscedastic errors: Here we consider the problem of heteroscedasticity. Means we assume that the elements of the random vector 'ε' were not independent and identically distributed random variables. To introduce Heteroscedastic and/or correlated errors with the assumptions E(ε) = 0 and Cov(ε) = σ^2 V, where V is a known n × n symmetric positive definite (pd) matrix there exists a nonsingular symmetric matrix P such that $V^{-1} = P'P$ and $\sigma^2 > 0$ is the unknown variance parameter. Then the linear model given in Equation (2) can be written as

$$\tilde{Y} = \tilde{Z}\alpha + \tilde{\varepsilon} \;,$$

Where, $\tilde{Y} = PY$, $\tilde{\varepsilon} = P\varepsilon$ and $\tilde{Z} = PZ$

Matrix V will be estimated by method[7]. In the present study we choose $\rho = 0.95$ and consider matrix V is estimated as below

$$V = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix}$$

Table-3: Ratio of AMSE of LS over various GR estimators for different 'k_i' (p=4, ρ = 0.95).

k_i	$n=10$				$n=80$				$n=500$			
	$\sigma^2=1$	8	25	100	$\sigma^2=1$	8	25	100	$\sigma^2=1$	8	25	100
$k_i(1)$	4.119	5.225	5.815	6.809	3.104	4.239	4.781	4.862	2.915	3.756	4.455	5.015
$k_i(2)$	1.130	1.751	1.771	2.352	0.994	1.449	1.661	1.652	1.177	1.351	1.538	1.968
$k_i(3)$	2.001	2.025	2.017	2.005	2.000	2.001	2.003	2.001	2.000	2.000	2.001	2.001
$k_i(4)$	2.010	2.010	2.000	2.000	2.01	2.000	2.011	2.020	2.012	2.000	2.020	2.000
$k_i(5)$	3.064	4.168	4.647	5.572	1.980	3.186	3.724	3.792	1.828	2.614	3.360	3.971
$k_i(6)$	3.491	4.044	4.274	4.625	2.842	3.551	3.836	3.868	2.731	3.254	3.656	3.956
$k_i(7)$	2.562	2.660	2.681	2.719	2.040	2.051	2.052	2.052	2.006	2.008	2.008	2.008
$k_i(8)$	1.118	0.204	0.005	0.015	2.029	0.070	0.059	0.003	0.011	0.003	0.182	0.049
$k_i(9)$	0.442	1.513	1.478	2.025	0.036	0.796	1.195	1.223	0.005	0.272	0.866	1.377
$k_i(10)$	0.153	2.607	2.360	2.862	0.009	0.582	2.416	2.876	0.001	0.087	0.673	1.824
$k_i(11)$	1.297	2.132	2.237	2.924	0.623	1.794	2.180	2.181	0.533	1.240	1.926	2.540
$k_i(12)$	0.140	1.441	1.476	2.025	0.008	0.450	1.126	1.226	0.001	0.082	0.570	1.359
$k_i(13)$	4.118	5.203	5.792	6.784	2.974	3.741	4.005	4.015	2.430	2.586	2.627	2.669
$k_i(14)$	0.126	1.374	1.465	2.015	0.008	0.361	1.019	1.204	0.001	0.071	0.453	1.264
$k_i(15)$	1.299	2.135	2.239	2.927	0.633	1.884	2.276	2.264	0.560	1.505	2.322	2.979
$k_i(16)$	0.140	1.446	1.477	2.028	0.008	0.446	1.112	1.223	0.001	0.081	0.538	1.313
$k_i(17)$	3.902	4.896	5.426	6.325	3.012	4.013	4.489	4.570	2.825	3.604	4.215	4.692
$k_i(18)$	0.126	1.354	1.461	2.002	0.008	0.361	1.021	1.204	0.001	0.071	0.457	1.292
$k_i(19)$	0.126	1.328	1.422	1.926	0.008	0.356	0.991	1.174	0.001	0.071	0.446	1.250
$k_i(20)$	0.730	1.636	1.609	2.196	0.173	1.063	1.414	1.426	0.103	0.547	1.150	1.667
$k_i(21)$	0.134	1.424	1.493	2.053	0.008	0.414	1.111	1.256	0.001	0.077	0.527	1.379
$k_i(22)$	0.126	1.345	1.459	2.004	0.008	0.357	1.004	1.204	0.001	0.071	0.452	1.278
$k_i(23)$	1.524	2.321	2.488	3.216	0.877	2.127	2.520	2.512	0.793	1.593	2.283	2.921
$k_i(24)$	0.137	1.426	1.473	2.022	0.008	0.426	1.105	1.222	0.001	0.079	0.541	1.346

$k_i(25)$	2.251	2.625	2.646	2.709	2.008	2.047	2.051	2.052	2.001	2.007	2.008	2.008
$k_i(26)$	0.126	1.389	1.467	2.013	0.008	0.361	1.029	1.206	0.001	0.071	0.453	1.265
$k_i(27)$	1.887	2.319	2.330	3.110	1.702	2.900	3.155	3.148	1.554	2.485	2.717	2.803
$k_i(28)$	0.140	1.426	1.470	2.021	0.008	0.442	1.093	1.211	0.001	0.081	0.534	1.302
$k_i(29)$	3.257	3.724	3.937	4.253	2.727	3.312	3.548	3.581	2.616	3.080	3.409	3.642
$k_i(30)$	0.136	1.356	1.454	2.010	0.009	0.391	1.021	1.201	0.001	0.081	0.488	1.278
$k_i(\lambda)$	4.114	5.218	5.810	6.804	3.103	4.236	4.778	4.859	2.914	3.754	4.453	5.012

Case III: Data generated using outlier observations: By introducing two outliers for different model specifications.

Case IV: Data generated using outlier observations and heteroscedastic errors: Here we evaluate the performance of estimators for the simulated data exits with one or multiple outliers and heteroscedastic and/or correlated errors, in linear regression model in the presence of multicollinearity. We introduce two outliers in the simulated data with heteroscedastic and/or correlated errors, where heteroscedastic and/or correlated errors are introduced using the same method as given in Case: II.

Case V: Data generated using non-normal errors: Like heteroscedastic and/or correlated errors, departure from the normality assumption is also one of` the common problems in regression.

Here we attempt the same problem and to evaluate performance of the proposed estimator, we assumed that 'ε' follows a non-normal distribution. To examine the robustness of all estimators under consideration, random numbers are generated for the error terms (ε) from each of the Chi-square, F and t distributions respectively.

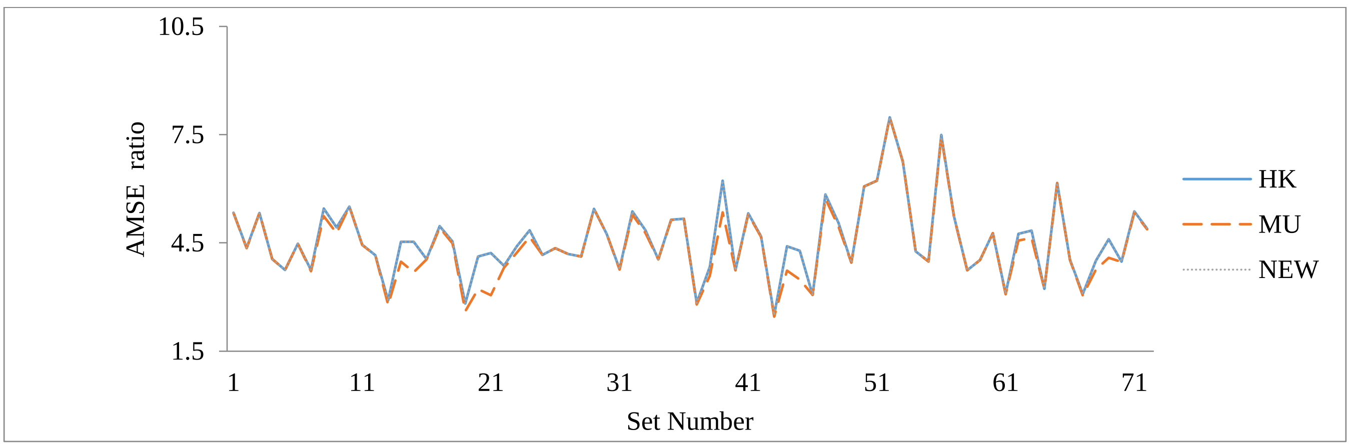

Figure-2: Radio of AMSE of LS over various GR estimators.

Table-4: Ratio of AMSE of LS over various GR estimators for different 'k_i' (p=4, ρ = 0.95).

k_i	$n=10$				$n=80$				$n=500$			
	$\sigma^2=1$	8	25	100	$\sigma^2=1$	8	25	100	$\sigma^2=1$	8	25	100
$k_i(1)$	6.129	5.534	5.933	5.999	5.865	5.285	5.961	5.465	5.388	6.172	5.299	5.961
$k_i(2)$	2.010	1.993	1.968	2.276	2.463	2.020	2.584	2.128	2.361	2.379	2.171	2.791
$k_i(3)$	2.043	2.029	2.030	2.003	2.029	2.020	2.019	2.002	2.017	2.024	2.008	2.001
$k_i(4)$	2.000	2.000	2.020	2.010	2.000	2.010	2.000	2.000	2.014	2.000	2.020	2.007

$k_i(5)$	4.935	4.513	4.766	4.926	4.798	4.254	4.914	4.415	4.393	4.971	4.297	4.918
$k_i(6)$	4.384	4.197	4.310	4.368	4.313	4.079	4.359	4.152	4.137	4.396	4.096	4.355
$k_i(7)$	2.694	2.683	2.687	2.700	2.053	2.052	2.053	2.053	2.008	2.008	2.008	2.008
$k_i(8)$	1.820	1.615	1.773	1.996	2.015	1.631	2.115	1.733	1.830	1.831	1.612	2.213
$k_i(9)$	1.827	1.624	1.780	2.003	2.033	1.647	2.135	1.747	1.852	1.843	1.625	2.238
$k_i(10)$	1.821	1.616	1.774	1.995	2.016	1.634	2.116	1.735	1.834	1.829	1.614	2.216
$k_i(11)$	2.488	2.456	2.404	2.775	3.201	2.661	3.369	2.800	3.044	3.137	2.805	3.574
$k_i(12)$	1.825	1.619	1.772	2.005	2.042	1.651	2.129	1.751	1.859	1.851	1.618	2.234
$k_i(13)$	6.093	5.501	5.892	5.965	4.459	4.176	4.455	4.234	2.691	2.728	2.675	2.697
$k_i(14)$	1.824	1.620	1.777	2.000	2.020	1.637	2.121	1.738	1.832	1.827	1.612	2.214
$k_i(15)$	2.489	2.457	2.404	2.777	3.312	2.754	3.498	2.894	3.488	3.463	3.175	4.004
$k_i(16)$	1.831	1.628	1.785	2.010	2.042	1.654	2.146	1.755	1.849	1.841	1.624	2.235
$k_i(17)$	5.731	5.139	5.547	5.566	5.461	4.929	5.535	5.097	5.015	5.764	4.933	5.535
$k_i(18)$	1.801	1.596	1.738	1.986	1.990	1.612	2.053	1.704	1.813	1.805	1.555	2.132
$k_i(19)$	1.762	1.572	1.671	1.941	1.938	1.575	1.970	1.661	1.797	1.697	1.550	2.050
$k_i(20)$	1.928	1.772	1.879	2.151	2.286	1.854	2.404	1.960	2.151	2.120	1.884	2.584
$k_i(21)$	1.840	1.638	1.782	2.034	2.098	1.692	2.180	1.792	1.927	1.896	1.650	2.283
$k_i(22)$	1.806	1.601	1.745	1.989	2.001	1.619	2.065	1.711	1.818	1.813	1.562	2.151
$k_i(23)$	2.715	2.713	2.612	3.017	3.582	3.009	3.769	3.155	3.421	3.573	3.185	3.987
$k_i(24)$	1.824	1.618	1.772	2.003	2.036	1.647	2.125	1.747	1.853	1.845	1.616	2.229
$k_i(25)$	2.690	2.651	2.672	2.687	2.052	2.052	2.052	2.052	2.008	2.008	2.008	2.008
$k_i(26)$	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
$k_i(27)$	2.685	2.598	2.591	3.075	3.689	3.412	3.699	3.477	2.853	2.847	2.815	2.868
$k_i(28)$	1.827	1.623	1.780	2.003	2.024	1.640	2.124	1.741	1.836	1.830	1.614	2.219
$k_i(29)$	4.042	3.838	3.978	4.000	3.955	3.747	3.987	3.813	3.785	4.057	3.752	3.987
$k_i(30)$	1.820	1.616	1.773	1.995	2.015	1.633	2.114	1.734	1.831	1.826	1.611	2.212
$k_i(\lambda)$	6.125	5.530	5.929	5.995	5.861	5.282	5.957	5.462	5.385	6.169	5.296	5.957

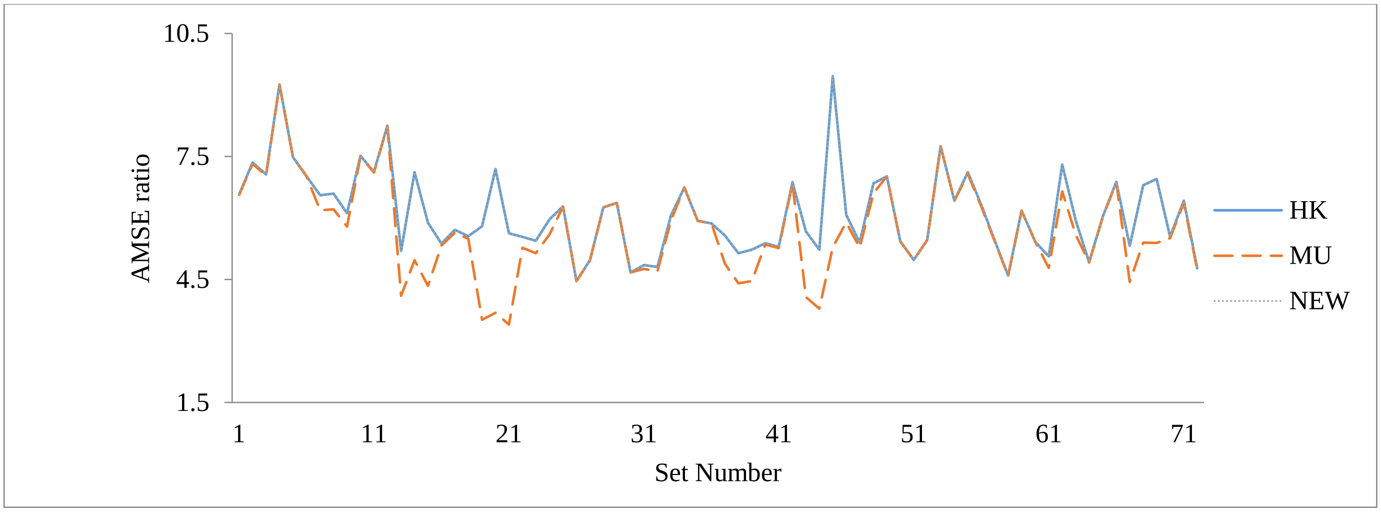

Figure-3: Radio of AMSE of LS over various GR estimators.

Table-5: Ratio of AMSE of LS over various GR estimators for different ‘ k_i ’ (p=4, ρ = 0.95).

k_i	$n=10$				$n=80$				$n=500$			
	$\sigma^2=1$	8	25	100	$\sigma^2=1$	8	25	100	$\sigma^2=1$	8	25	100
$k_i(1)$	5.147	6.498	7.353	5.151	6.076	5.363	5.703	5.405	5.709	5.793	5.856	5.928
$k_i(2)$	1.739	2.470	2.857	1.852	2.502	2.138	2.404	2.409	2.387	2.432	2.604	2.546
$k_i(3)$	2.025	2.055	2.078	2.002	2.043	2.022	2.010	2.002	2.025	2.023	2.009	2.002
$k_i(4)$	2.023	2.040	2.000	2.010	2.000	2.002	2.010	2.000	2.000	2.020	2.004	2.012
$k_i(5)$	4.146	5.304	6.043	4.187	4.956	4.334	4.679	4.448	4.669	4.726	4.804	4.845
$k_i(6)$	4.030	4.513	4.768	4.047	4.381	4.110	4.264	4.153	4.259	4.284	4.313	4.331
$k_i(7)$	2.662	2.711	2.732	2.667	2.053	2.053	2.053	2.053	2.008	2.008	2.008	2.008
$k_i(8)$	1.554	2.177	2.544	1.649	2.010	1.762	1.968	1.988	1.929	1.954	2.078	2.067
$k_i(9)$	1.560	2.192	2.560	1.659	2.027	1.776	1.988	2.009	1.946	1.974	2.101	2.088
$k_i(10)$	1.554	2.178	2.545	1.650	2.011	1.763	1.971	1.989	1.931	1.957	2.080	2.069
$k_i(11)$	2.137	3.011	3.482	2.267	3.246	2.781	3.137	3.100	3.126	3.174	3.376	3.322
$k_i(12)$	1.560	2.191	2.559	1.661	2.029	1.777	1.990	2.006	1.942	1.956	2.103	2.091
$k_i(13)$	5.127	6.468	7.320	5.132	4.630	4.311	4.423	4.252	2.688	2.693	2.734	2.694
$k_i(14)$	1.557	2.185	2.552	1.654	2.015	1.766	1.976	1.994	1.929	1.955	2.078	2.067
$k_i(15)$	2.138	3.013	3.484	2.269	3.356	2.885	3.261	3.226	3.470	3.504	3.826	3.680
$k_i(16)$	1.564	2.198	2.566	1.663	2.036	1.784	1.998	2.018	1.944	1.971	2.097	2.085
$k_i(17)$	4.795	6.046	6.846	4.788	5.654	4.999	5.305	5.015	5.309	5.388	5.445	5.520
$k_i(18)$	1.548	2.165	2.524	1.648	1.978	1.730	1.935	1.948	1.868	1.878	2.031	2.015
$k_i(19)$	1.542	2.054	2.419	1.648	1.888	1.673	1.937	1.907	1.828	1.801	1.989	1.914
$k_i(20)$	1.653	2.350	2.734	1.767	2.299	1.992	2.233	2.261	2.198	2.241	2.411	2.370

$k_i(21)$	1.581	2.221	2.589	1.689	2.079	1.819	2.050	2.063	1.983	1.989	2.169	2.137
$k_i(22)$	1.548	2.168	2.531	1.649	1.988	1.735	1.941	1.953	1.877	1.892	2.042	2.031
$k_i(23)$	2.327	3.270	3.773	2.468	3.644	3.117	3.514	3.448	3.516	3.567	3.782	3.723
$k_i(24)$	1.558	2.188	2.556	1.659	2.024	1.773	1.985	2.001	1.939	1.954	2.097	2.086
$k_i(25)$	2.684	2.700	2.728	2.655	2.052	2.052	2.053	2.052	2.008	2.008	2.008	2.008
$k_i(26)$	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
$k_i(27)$	2.310	3.339	3.837	2.518	3.726	3.548	3.671	3.688	2.842	2.849	2.882	2.854
$k_i(28)$	1.560	2.190	2.557	1.658	2.018	1.769	1.979	1.998	1.932	1.959	2.083	2.072
$k_i(29)$	3.694	4.160	4.405	3.696	4.025	3.776	3.898	3.791	3.901	3.930	3.951	3.976
$k_i(30)$	1.554	2.177	2.544	1.649	2.010	1.762	1.970	1.987	1.928	1.953	2.076	2.065
$k_i(\lambda)$	5.142	6.493	7.346	5.147	6.073	5.359	5.699	5.401	5.706	5.789	5.852	5.924

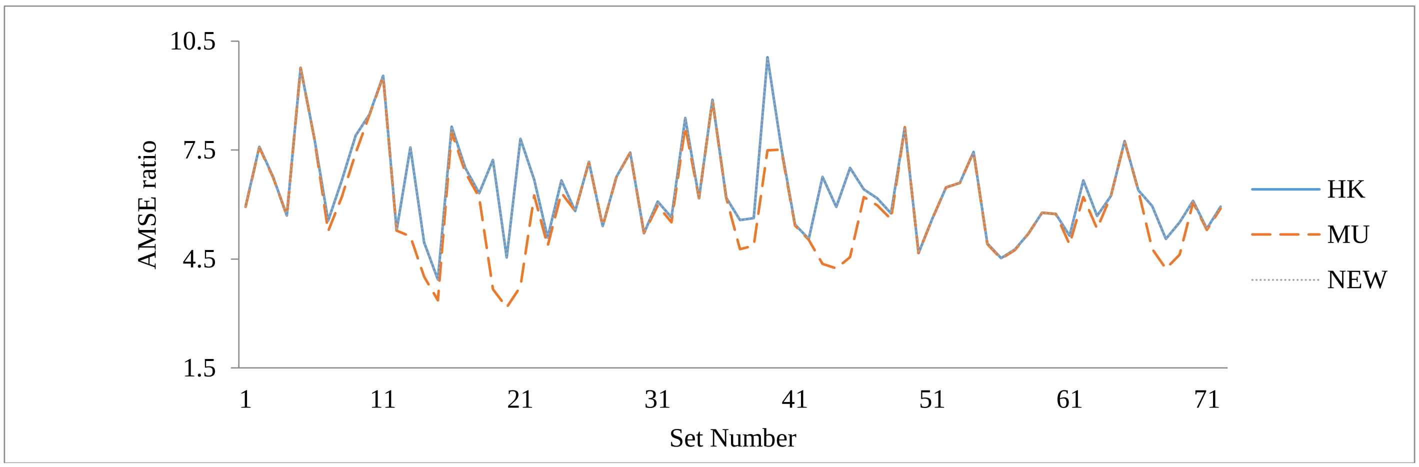

Figure-4: Radio of AMSE of LS over various GR estimators.

Table-6: Ratio of AMSE of LS over various GR estimators for different ‘ k_i ’ (p=4, ρ = 0.95).

k_i	$n=10$			$n=80$			$n=500$		
	$\varepsilon\sim$ Chi(6)	F(2,3)	T(8)	Chi(6)	F(2,3)	T(8)	Chi(6)	F(2,3)	T(8)
$k_i(1)$	5.444	7.645	5.031	4.063	4.320	3.411	3.714	4.723	3.179
$k_i(2)$	1.799	2.048	1.713	1.501	1.701	1.258	1.629	1.928	1.412
$k_i(3)$	2.013	2.005	2.002	2.000	2.002	2.000	2.000	2.001	2.000
$k_i(4)$	2.010	2.010	2.020	2.002	2.000	2.003	2.001	2.000	2.000
$k_i(5)$	4.305	5.873	3.783	2.896	3.338	2.179	2.413	3.490	1.979
$k_i(6)$	4.113	4.735	3.918	3.433	3.622	3.021	3.206	3.766	2.895
$k_i(7)$	2.664	2.710	2.620	2.049	2.052	2.041	2.008	2.008	2.006

$k_i(8)$	0.714	0.087	0.857	0.873	0.118	0.102	1.968	0.354	2.010
$k_i(9)$	1.431	1.731	0.592	0.357	0.921	0.045	0.061	0.516	0.007
$k_i(10)$	1.650	2.392	0.208	0.109	0.552	0.011	0.015	0.162	0.002
$k_i(11)$	2.237	2.668	1.882	1.462	2.075	0.735	0.872	1.988	0.562
$k_i(12)$	1.052	1.573	0.194	0.104	0.457	0.011	0.015	0.159	0.002
$k_i(13)$	5.425	7.615	5.025	3.694	3.771	3.221	2.584	2.681	2.498
$k_i(14)$	0.900	1.497	0.169	0.089	0.383	0.010	0.014	0.137	0.002
$k_i(15)$	2.241	2.671	1.886	1.530	2.188	0.748	0.968	2.454	0.589
$k_i(16)$	1.053	1.574	0.194	0.104	0.454	0.011	0.015	0.157	0.002
$k_i(17)$	5.093	7.207	4.755	3.880	4.059	3.312	3.593	4.480	3.080
$k_i(18)$	0.896	1.492	0.169	0.089	0.384	0.010	0.014	0.137	0.002
$k_i(19)$	0.843	1.198	0.167	0.089	0.379	0.010	0.014	0.136	0.002
$k_i(20)$	1.616	1.875	1.030	0.691	1.295	0.207	0.250	0.990	0.108
$k_i(21)$	0.991	1.556	0.183	0.098	0.429	0.011	0.014	0.149	0.002
$k_i(22)$	0.886	1.484	0.168	0.089	0.381	0.010	0.014	0.136	0.002
$k_i(23)$	2.488	3.002	2.237	1.842	2.416	1.036	1.231	2.453	0.847
$k_i(24)$	1.013	1.555	0.187	0.100	0.437	0.011	0.015	0.153	0.002
$k_i(25)$	2.545	2.635	2.275	2.034	2.048	2.010	2.005	2.008	2.002
$k_i(26)$	0.905	1.500	0.169	0.089	0.384	0.010	0.014	0.136	0.002
$k_i(27)$	2.500	2.774	2.798	2.645	3.175	1.814	2.032	2.711	1.578
$k_i(28)$	1.038	1.565	0.194	0.104	0.450	0.011	0.015	0.157	0.002
$k_i(29)$	3.806	4.482	3.649	3.230	3.349	2.900	3.059	3.524	2.770
$k_i(30)$	0.918	1.503	0.184	0.102	0.420	0.012	0.016	0.156	0.002
$k_i(\lambda)$	5.439	7.640	5.027	4.061	4.317	3.410	3.712	4.721	3.178

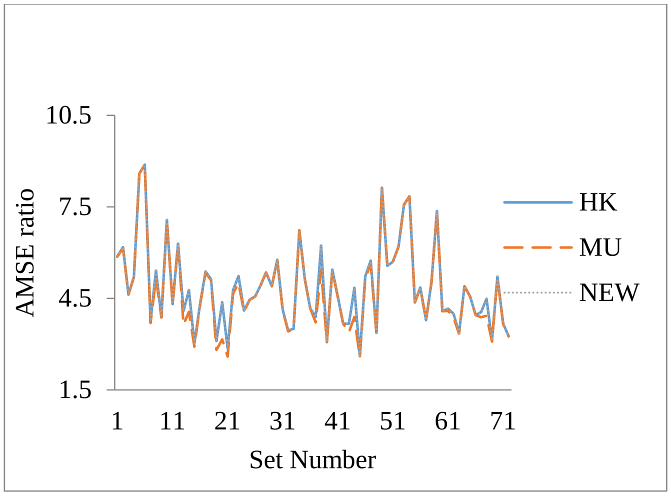

Figure-5: Radio of AMSE of LS over various GR estimators.

From Table-2 to Table-6 and Figure-1 to Figure-5, we observe that $k_i(\lambda)$ performs equivalently to $k_i(HK)$ ridge parameters[1]. Whereas, it gives better performance than $k_i(MU)$ and other ridge parameters reviewed in this article for all combinations of number of predictors (*p*), correlation between predictors (ρ), sample size (n) and variance of the error term (σ^2) used in this simulation study in all the cases from I to V under the influence of multicollinearity.

Conclusion

A method for specifying 'k_i' is proposed herein and were evaluated in terms of MSE via simulation techniques. Comparisons were made with other 30 estimators proposed elsewhere earlier according to MSE criterion. The simulation study showed that the improvement of the suggested estimators is substantial from the MSE point of view. We believe that the findings of this paper will be definitely useful for the practitioners to select the optimal ridge parameters in the context of ridge parameters in Generalized Ridge Regression.

Appendix: List of the ridge parameters considered for GR estimator:

Hoerl and Kennard (1970)[1]

$$k_i(HK)\ or\ k_i(1) = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^{\,2}} \quad , \qquad i = 1,2,\dots,p$$

Lawless and Wang (1976)[10]

$$k_i(2) = \frac{\hat{\sigma}^2}{\lambda_i\,\hat{\alpha}_i^{\,2}} \quad , \qquad i = 1,2,\dots,p$$

Myoken and Nagoya, (1977)[11]

$$k_i(3) = \min\left[\frac{\hat{\sigma}^2}{\hat{\alpha}_i^{\,2}}, \frac{2\hat{\sigma}^2}{\left|\hat{\beta}'\hat{\beta} - \hat{\sigma}^2\left((1/\lambda_i) + \sum_{i=1}^{p}(1/\lambda_i)\right)\right|}\right] \quad i = 1,2,\dots,p$$

Hemmerle and Brantle, (1978)[12]

$$k_i(4) = \lambda_i\hat{\sigma}^2 \big/ \left(\lambda_i\hat{\alpha}_i^{\,2} - \hat{\sigma}^2\right) \quad , \quad i = 1,2,\dots,p$$

Masuo Nomura (1988)[13]

$$k_i(5) = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^{\,2}}\left\{1 + \left[1 + \lambda_i\left(\hat{\alpha}_i^{\,2}/\hat{\sigma}^2\right)\right]^{\frac{1}{2}}\right\} \quad , \quad i = 1,2,\dots,p$$

Troskie and Chalton (1996)[14]

$$k_i(6) = \lambda_i\hat{\sigma}^2 \big/ \left(\lambda_i\hat{\alpha}_i^{\,2} + \hat{\sigma}^2\right), \quad i = 1,2,\dots,p$$

Fringuetti (1999)[7]

$$k_i(F)\ or\ k_i(7) = \lambda_i\hat{\sigma}^2 \big/ \left[\lambda_i\hat{\alpha}_i^{\,2} + (n-p-1)\hat{\sigma}^2\right], \quad i = 1,2,\dots,p$$

Alkhamisi et al (2006)[15]

$$k_i(8) = \frac{\hat{\sigma}^2}{\hat{\alpha}_i} \quad , \qquad i = 1,2,\dots,p$$

Alkhamisi and Shukur (2007)[16]

$$k_i(9) = \left[\frac{\hat{\sigma}^2}{\hat{\alpha}_i^{\,2}} + \frac{1}{\lambda_i}\right] \quad , \qquad i = 1,2,\dots,p$$

Batah et al. (2008)[17]

$$k_i(10) = \hat{\sigma}^2\left\{\left[\left(\hat{\alpha}_i^{\,4}\lambda_i^2/4\hat{\sigma}^2\right) + \left(6\hat{\alpha}_i^{\,4}\lambda_i/\hat{\sigma}^2\right)\right]^{1/2} - \left(\hat{\alpha}_i^{\,2}\lambda_i/2\hat{\sigma}^2\right)\right\} \big/ \hat{\alpha}_i^{\,2} \quad i = 1,2,\dots,p$$

Muniz and Kibria, (2009)[18]

$$k_i(11) = \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^{\,2}}} \quad , \qquad i = 1,2,\dots,p$$

$$k_i(12) = 1 \Big/ \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^{\,2}}} \quad , \qquad i = 1,2,\dots,p$$

Muniz et al.,(2012)[8] $k_i(MU)$ *or*

$$k_i(13) = \left[\frac{\lambda_{\max}\hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_i^{\,2}}\right], i = 1,2,\dots,p$$

$$k_i(14) = 1 \Big/ \left[\frac{\lambda_{\max}\hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_i^{\,2}}\right] \quad , \qquad i = 1,2,\dots,p$$

$$k_i(15) = \sqrt{\frac{\lambda_{\max}\hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_i^{\,2}}} \quad , \qquad i = 1,2,\dots,p$$

$$k_i(16) = 1 \Big/ \sqrt{\frac{\lambda_{\max}\hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_i^{\,2}}} \quad , \qquad i = 1,2,\dots,p$$

Dorugade (2014)[19]

$$k_i(17) = \frac{2\hat{\sigma}^2}{\lambda_{\max}\hat{\alpha}_i^{\,2}} \quad , \qquad i = 1,2,\dots,p$$

Lukman and Ayinde (2017)[20]

$$k_i(18) = 1 \Big/ \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \quad , \qquad i = 1,2,...,p$$

$$k_i(19) = 1 \Big/ \frac{\hat{\sigma}^2}{\lambda_i \, \hat{\alpha}_i^2} \quad , \qquad i = 1,2,...,p$$

$$k_i(20) = \sqrt{\frac{\hat{\sigma}^2}{\lambda_i \, \hat{\alpha}_i^2}} \quad , \qquad i = 1,2,...,p$$

$$k_i(21) = 1 \Big/ \sqrt{\frac{\hat{\sigma}^2}{\lambda_i \, \hat{\alpha}_i^2}} \quad , \qquad i = 1,2,...,p$$

Lukman et al (2017)[20]

$$k_i(22) = 1 \Big/ \frac{2\hat{\sigma}^2}{\lambda_{\max} \hat{\alpha}_i^2} \quad , \qquad i = 1,2,...,p$$

$$k_i(23) = \sqrt{\frac{2\hat{\sigma}^2}{\lambda_{\max} \hat{\alpha}_i^2}} \quad , \qquad i = 1,2,...,p$$

$$k_i(24) = 1 \Big/ \sqrt{\frac{2\hat{\sigma}^2}{\lambda_{\max} \hat{\alpha}_i^2}} \quad , \qquad i = 1,2,...,p$$

Lukman et al (2018)[21]

$$k_i(25) = \left[\frac{\lambda_i \hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_i \hat{\alpha}_{\max}^2} \right] \quad , \qquad i = 1,2,...,p$$

$$k_i(26) = 1 \Big/ \left[\frac{\lambda_i \hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_i \hat{\alpha}_{\max}^2} \right] \quad , \qquad i = 1,2,...,p$$

$$k_i(27) = \sqrt{\frac{\lambda_i \hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_i \hat{\alpha}_{\max}^2}} \quad , \qquad i = 1,2,...,p$$

$$k_i(28) = 1 \Big/ \sqrt{\frac{\lambda_i \hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_i \hat{\alpha}_{\max}^2}} \quad , \qquad i = 1,2,...,p$$

Kibria and Lukman (2020)[22]

$$k_i(29) = \left[\frac{\hat{\sigma}^2}{2\hat{\alpha}_i^2 + (\hat{\sigma}^2 / \lambda_i)} \right] \quad , \qquad i = 1,2,...,p$$

Qasim et al. (2021)[23]

$$k_i(30) = \left[\frac{(n-p-1)\hat{\sigma}^2 + \lambda_{\max} \hat{\alpha}_i^2}{\lambda_{\min}} \right] \quad , \qquad i = 1,2,...,p$$

Where, $\hat{\alpha}_i$ is the ith element of $\hat{\alpha}$, $i = 1,2,\ldots,p$ and $\hat{\sigma}^2$ is the OLS estimator of σ^2 i.e.

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}' Z' Y}{n-p-1}.$$

References

1. Hoerl A.E., Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55–67.
2. Montgomery, D.C., Peck, A.E. and Vining, G.G. (2001). Introduction to Linear Regression Analysis. New York: Wiley.
3. Hoerl A.E., Kennard, R. W., Baldwin, K. F. (1975). Ridge regression: Some simulations. *Communications in Statistics - Theory and Methods*, 4, 105–123.
4. Mermi S., G¨oktas¸, A., Akkus¸, ¨O. & Necla G. (2024). A new robust ridge parameter estimator having no outlier and ensuring normality for linear regression model. *Journal of Radiation Research and Applied Sciences*, 17(1), 100788.
5. Ishwaran H. and Rao J. S. (2011). Consistency of spike and slab regression. *Statist. Probab. Lett*., 81(12), 1920–1928.
6. Hemmerle, W. J. (1975). An Explicit Solution for Generalized Ridge Regression. *Technometrics*, 17, 309-314.
7. Firinguetti, L. (1999). A generalized ridge regression estimator and its finite sample properties. *Communications in Statistics-Theory and Methods*, 28(5), 1217-1229.
8. Muniz, G., Kibria, B.M.G., Kristofer, M. and Shukur, G. (2012). On Developing Ridge Regression Parameters: *A Graphical investigation. SORT-Statistics and Operations Research Transactions*, 36(2), 115-138.
9. McDonald, G.C. and Galarneau, D.I. (1975). A monte carlo evaluation of some ridge-type estimators. *J. Amer. Statist. Assoc*., 70, 407–416.
10. Lawless, J.F. and Wang, P. (1976). A Simulation Study of Ridge and Other Regression Estimators. *Communications in Statistics- Theory and Methods*, 14, 1589-1604
11. Myoken H. and Nagoya Y. U. (1977). Adjustments for Regression Parameters. *Metrika*, 24, 113.
12. Hemmerle, W. J. and Brantle, T. F. (1978). An explicit and constrained generalized ridge estimation. *Technometrics*, 20, 109–119.
13. Masuo, N. (1988). On the almost unbiased ridge regression estimation. *Communications in Statistics - Simulation and Computation*, 17(3), 729-743.
14. Troskie, C. G. and Chalton, D. O. (1996). A Bayesian estimate for the constants in ridge regression. *South African Statistical Journal*, 30, 119–137.
15. Alkhamisi, M., Khalaf, G. and Shukur, G. (2006). Some modifications for choosing ridge parameters. *Communications in Statistics - Theory and Methods*, 35, 2005–2020.

16. Alkhamisi, M.A. and Shukur, G.A. (2007). Monte carlo study of recent ridge parameters. *Communications in Statistics - Simulation and Computation*, 36, 535–547.

17. Batah, F.S., Ramnathan, T. and Gore, S.D. (2008). The efficiency of modified jackknife and ridge type regression estimators: a comparison. *Surveys in Mathematics and its Applications*, 24(2), 157-174.

18. Muniz, G. and Kibria, B. M. (2009). On some ridge regression estimators: An empirical comparison. *Communications in Statistics - Simulation and Computation*, 38, 621–630.

19. Dorugade, A.V. (2014). New ridge parameters for ridge regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, 15, 94-99.

20. Lukman, A, F, and Ayinde, K. (2017). Review and classifications of the ridge parameter estimation techniques. *Hacettepe Journal of Mathematics and Statistics*, 46(5), 953-967.

21. Lukman, A. F., Oluyemi, O. A., Akanbi, O. B., & Clement, O. A. (2018). Classification-based ridge estimation techniques of Alkhamisi methods. *Journal of Probability and Statistical Science,* 16, 165–181.

22. Kibria, B. M. G., & Lukman, A. F. (2020). A new ridge-type estimator for the linear regression model: Simulations and applications. *Scientifica*. https://doi.org/10.1155/2020/9758378

23. Qasim, M., Mansson, K., & Kibria, B. M. G. (2021). On some beta ridge regression estimators: Method, simulation and application. *Journal of Statistical Computation and Simulation*, 91, 1699–1712.