Journal of the Association of Arab Universities for Basic and Applied Sciences (2016) 21, 96–102

University of Bahrain

Journal of the Association of Arab Universities for Basic and Applied Sciences

www.elsevier.com/locate/jaaubas
www.sciencedirect.com

ORIGINAL ARTICLE

Adjusted ridge estimator and comparison with Kibria's method in linear regression

A.V. Dorugade

Y.C. Mahavidyalaya, Halkarni, Tal-Chandgad, Kolhapur, Maharashtra- 416552, India

Received 12 November 2014; revised 6 April 2015; accepted 25 April 2015
Available online 21 May 2015

KEYWORDS

Ridge regression;
Ridge estimator;
Mean square error;
Simulation

Abstract This paper proposes an adjusted ridge regression estimator for β for the linear regression model. The merit of the proposed estimator is that it does not require estimating the ridge parameter k unlike other existing estimators. We compared our estimator with an ordinary least squares (LS) estimator and with some well known estimators proposed by Hoerl and Kennard (1970), ordinary ridge regression (RR) estimator and generalized ridge regression (GR) and some estimators proposed by Kibria (2003) among others. A simulation study has been conducted and compared for the performance of the estimators in the sense of smaller mean square error (MSE). It appears that the proposed estimator is promising and can be recommended to the practitioners.

1. Introduction

Regression analysis is one of the frequently used tools for forecasting in almost all disciplines; hence estimation of unknown parameters is a common interest for many users. These estimates can be found by various estimation methods. The easiest and the most common method of them is the ordinary least squares (LS) technique, which minimizes the squared distance between the estimated and observed values. Multicollinearity among the explanatory variables in the regression model is an important problem that exhibits serious undesirable effects on the analysis faced in applications. The LS estimator is sensitive to number 'errors', namely, there is an 'explosion' of the sampling variance of the estimators. Alternative estimators are designed to combat multicollinearity-yield-biased estimators.

One of the popular numerical techniques to deal with multicollinearity is the ridge regression due to Hoerl and Kennard (1970). Ridge regression approach has been studied by McDonald and Galarneau (1975), Swindel (1976), Lawless (1978), Singh and Chaubey (1987), Sarkar (1992), Saleh and Kibria (1993), Kibria (2003), Khalaf and Shukur (2005), Zhong and Yang (2007), Batah et al. (2008), Yan (2008), Yan and Zhao (2009), Muniz and Kibria (2009), Yang and Chang (2010), Khalaf (2012) and Dorugade (2014) and others. Ridge Regression estimator has been the benchmarked for almost all the estimators developed later in this context. Most of the researchers compare superiority of their suggested estimators with LS, RR, GR and other existing methods in terms of minimum MSE criterion in the presence of multicollinearity. In this article, our primary aim is to suggest an estimator by modifying the ordinary ridge regression (RR) estimator avoiding the computation of ridge parameter and secondly to evaluate the performance of our estimator with LS, RR and GR estimators in the presence of sever or extremely sever multicollinearity.

This article is organized as follows: in Section 2, we define model and parameter estimation methods with their bias and

E-mail address: adorugade@rediffmail.com

Peer review under responsibility of University of Bahrain.

http://dx.doi.org/10.1016/j.jaubas.2015.04.002
1815-3852

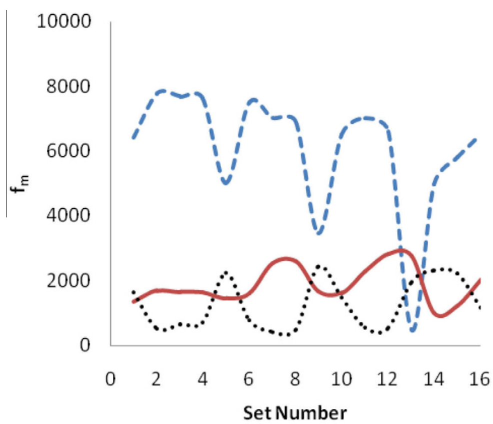

Figure 1 "f_m" for AR, RR and GR estimators (ρ = 0.95, p = 3 and β = (10, 4, 1, 8)′).

MSE. In Section 3, we have proposed biased estimator. We compare our new estimator in the MSE sense, with the RR estimator, in the same section. In Section 4, performances of the proposed estimators with respect to the scalar MSE criterion compared to LS, RR and GR estimators are evaluated on basis of the Monte Carlo Simulation results. Influence of choice of k to compute RR on the proposed estimator AR is also studied in the same section. Finally, article ends with some concluding remarks.

2. Model specifications and the estimators

We consider the linear regression model with p predictors and n observations:

$$Y = X\beta + \varepsilon, \tag{1}$$

where $Y = (Y_1, Y_2, \ldots, Y_n)'$, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$and $X = (x_1, x_2, \ldots, x_p)$. ε_i's are independently and identically distributed as normal with mean 0 and variance σ^2. Assume that the Y_i's are centered and the covariates x_i's are standardized. Let Λ and T be the matrices of eigen values and eigen vectors of $X'X$, respectively, satisfying $T'X'XT = \Lambda = \text{diagonal}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, where λ_i being the ith

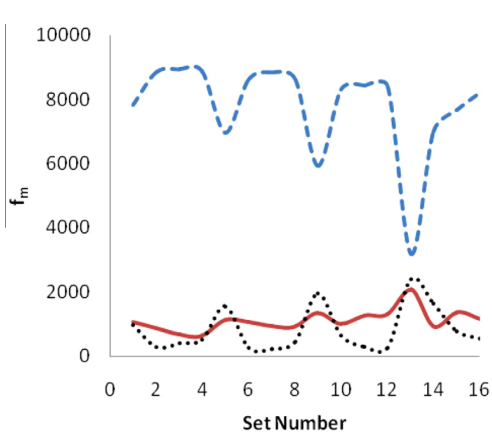

Figure 2 "f_m" for AR, RR and GR estimators (ρ = 0.99, p = 3 and β = (7, 4, 1, 8)′).

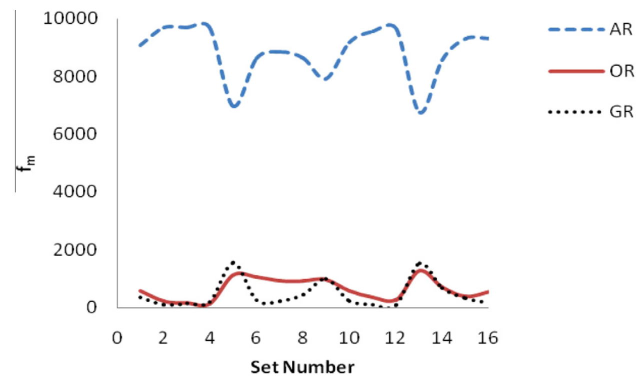

Figure 3 "f_m" for AR, RR and GR estimators (ρ = 0.999, p = 3 and β = (14, 5, 2, 6)′).

eigenvalue of $X'X$ and $T'T = TT' = I_p$. We obtain the equivalent model

$$Y = Z\sigma + \varepsilon, \tag{2}$$

where $Z = XT$, it implies that $Z'Z = \Lambda$, and $\alpha = T'\beta$ (see Montgomery et al., 2001).

Then LS estimator of α is given by

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

Therefore, LS estimator of β is given by

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

2.1. Generalized ridge regression estimator (GR)

In order to combat multicollinearity and improve the LS estimator, Hoerl and Kennard (1970) suggested an alternative estimator by adding a ridge parameter k to the diagonal elements of the least square estimator. They also suggested generalized ridge regression (GR) estimator by using separate ridge parameter for each regressor in the ridge regression. Also, 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 (Stephen and Christopher, 2001).

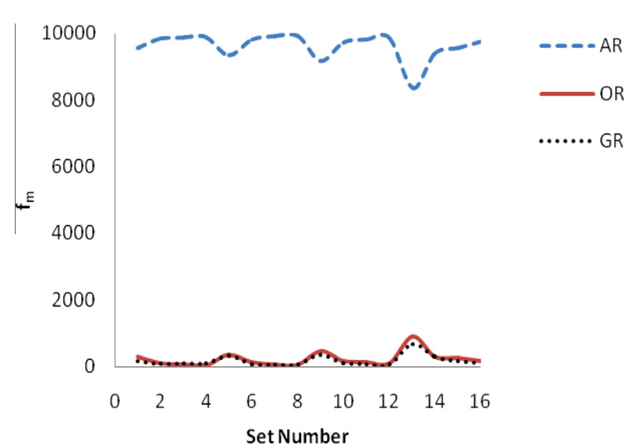

Figure 4 "f_m" for AR, RR and GR estimators (ρ = 0.9999, p = 3 and β = (10, 1, 1, 4)′).

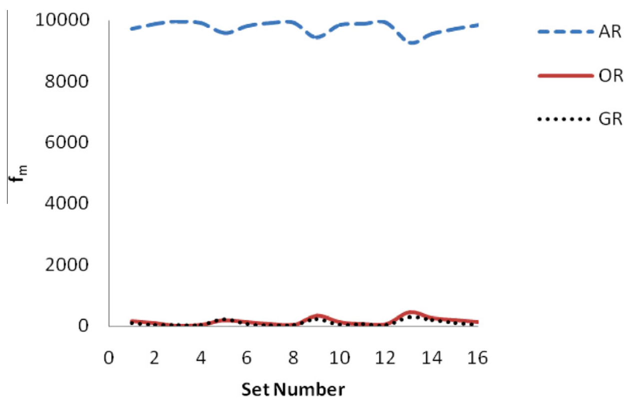

Figure 5 "f_m" for AR, RR and GR estimators (ρ = 0.99999, p = 3 and β = (8, 4, 11, 5)′).

The GR estimator of α is defined by

$$\hat{\alpha}_{GR} = (I - KA^{-1})\hat{\alpha}_{LS}, \tag{4}$$

where K = diagonal(k_1, k_2...k_p), $k_i \geqslant 0$, i = 1, 2, ...,p be the different ridge parameters for different regressor and $A = \Lambda + K$.

Hence GR estimator for β is $\hat{\beta}_{GR} = T\hat{\alpha}_{GR}$. and mean square error of $\hat{\alpha}_{GR}$ is

$$\text{MSE}(\hat{\alpha}_{GR}) = \text{Variance}(\hat{\alpha}_{GR}) + [\text{Bias}(\hat{\alpha}_{GR})]^2$$

$$\text{MSE}(\hat{\alpha}_{GR}) = \hat{\sigma}^2\sum_{i=1}^{p}\lambda_i/(\lambda_i + k_i)^2 + \sum_{i=1}^{p}k_i^2\hat{\alpha}_i^2/(\lambda_i + k_i)^2 \tag{5}$$

Setting $k_1 = k_2 = \ldots = k_p = k$ and $k \geqslant 0$, GR estimator reduces to RR estimator of α denoted by $\hat{\alpha}_{RR}$.. Hence, mean square error of $\hat{\alpha}_{RR}$ is

$$\hat{\alpha}_{RR} = [I - k(\Lambda + kI)^{-1}]\ \hat{\alpha}_{LS} \tag{6}$$

Therefore, RR estimator of β is given by

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

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

$$\text{MSE}(\hat{\alpha}_{RR}) = \hat{\sigma}^2\sum_{i=1}^{p}\lambda_i/(\lambda_i + k)^2 + k^2\sum_{i=1}^{p}\hat{\alpha}_i^2/(\lambda_i + k)^2 \tag{7}$$

We observe that when k = 0 in (7), MSE of LS estimator of α is recovered. Hence

Table 1 "f_m" for AR, RR, GR and LS estimators (p = 4 and β = (2, 15, 3, 14, 8)′).

ρ	Estimator	n = 20				50				100				500			
		σ^2 = 1	9	25	100	1	9	25	100	1	9	25	100	1	9	25	100
0.6	AR	1420	4810	5420	4930	180	3600	4080	4470	10	2010	4050	4070	0	170	1350	3010
	RR	2420	2200	2970	4080	1950	2220	2630	3250	1850	2530	2140	2530	2020	2050	2180	2030
	GR	4870	2530	1460	910	5870	3390	2730	2050	5650	4230	3110	2900	4930	5570	4920	3990
	LS	1290	460	150	80	2000	790	560	230	2490	1230	700	500	3050	2210	1550	970
0.8	AR	2670	5980	6440	6020	700	4770	5570	5130	190	3700	4830	5330	0	700	2900	4180
	RR	2400	2060	2610	3440	2280	2250	2260	3390	1470	2390	2340	2800	980	1610	2300	2430
	GR	3950	1870	910	530	5370	2610	1920	1400	6730	3130	2450	1750	7190	6160	3930	2910
	LS	980	90	40	10	1650	370	250	80	1610	780	380	120	1830	1530	870	480
0.9	AR	4190	6940	7460	7430	2080	5880	6580	6540	690	4970	6220	6430	0	2020	4390	5310
	RR	2560	1750	1960	2200	2590	2010	2220	2770	1900	2070	1900	2250	900	2400	2340	1990
	GR	2670	1220	540	350	4460	1950	1130	680	5990	2610	1720	1230	7390	4490	2790	2330
	LS	580	90	40	20	870	160	70	10	1420	350	160	90	1710	1090	480	370
0.95	AR	5560	7810	8200	8080	3400	7080	7670	7580	1950	6190	7210	7500	0	3530	5410	6450
	RR	2200	1480	1420	1660	2520	1650	1670	1960	2210	2000	1580	1720	1120	2630	1890	1570
	GR	1940	660	370	260	3350	1170	650	460	4860	1590	1110	760	7640	3180	2270	1860
	LS	300	50	10	0	730	100	10	0	980	220	100	20	1240	660	430	120
0.99	AR	7560	9140	9350	9680	6600	8670	9160	9270	5100	8120	8620	9170	1450	5970	7820	8330
	RR	1580	620	580	280	1880	900	640	590	2620	1100	1020	680	2390	2220	1110	980
	GR	810	240	60	40	1330	410	200	140	1920	730	360	150	5190	1590	950	670
	LS	50	0	10	0	190	20	0	0	360	50	0	0	970	220	120	20
0.999	AR	9280	9780	9890	9900	8730	9640	9750	9850	8250	9410	9760	9850	6240	8690	9240	9530
	RR	500	160	80	100	880	270	170	140	1150	410	180	70	2340	890	540	370
	GR	220	60	30	0	360	90	80	10	560	180	60	80	1280	400	220	100
	LS	0	0	0	0	30	0	0	0	40	0	0	0	140	20	0	0
0.9999	AR	9770	9910	9960	9980	9500	9820	9900	9960	9390	9810	9900	9930	8770	9530	9730	9830
	RR	160	80	40	20	410	150	80	40	480	170	70	50	930	340	220	130
	GR	70	10	0	0	90	30	20	0	130	20	30	20	260	130	50	40
	LS	0	0	0	0	0	0	0	0	0	0	0	0	40	0	0	0
0.99999	AR	9900	9990	9980	10,000	9830	9990	9940	9980	9780	9960	9930	9980	9590	9810	9930	9920
	RR	80	10	20	0	120	10	60	20	190	30	60	20	310	140	40	60
	GR	20	0	0	0	50	0	0	0	30	10	10	0	100	50	30	20
	LS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2 "f_{m1}" and "f_{m2}" for AR, RR, GR and LS estimators (p = 7 and β = (10, 1, 8, 5, 12, 1, 4, 7)′).

ρ	Estimator		n = 20				100				500			
			σ^2 = 1	9	25	100	1	9	25	100	1	9	25	100
0.6	f_{m1}	AR	5630	6140	5780	6010	3890	6210	5820	5880	50	4080	4960	3650
		RR	2050	3050	3640	3390	2870	2230	3300	3750	2870	2250	2200	4020
		GR	1800	750	560	600	2340	1410	840	370	4410	2500	2250	2240
		LS	520	60	20	0	900	150	40	0	2670	1170	590	90
	f_{m2}	AR	1810	5600	6310	6070	80	3550	5470	6420	0	830	2710	3610
		RR	4620	3210	2840	2340	5190	4410	3340	2980	3130	3970	3500	3530
		GR	960	420	30	20	1220	720	620	140	2830	1510	1050	700
		LS	2610	770	820	1570	3510	1320	570	460	4040	3690	2740	2160
0.99	f_{m1}	AR	8840	9340	9550	9590	7190	8850	9040	9020	4570	8310	8750	8860
		RR	800	560	340	330	1800	760	790	920	3020	910	800	890
		GR	350	100	110	80	840	380	170	60	1940	720	410	250
		LS	10	0	0	0	170	10	0	0	470	60	40	0
	f_{m2}	AR	2710	7450	8570	8490	310	4540	7020	7890	0	1340	4180	6020
		RR	6330	2210	950	450	7720	4780	2680	1840	6200	7150	5150	3690
		GR	170	70	20	10	270	210	150	50	710	140	220	90
		LS	790	270	460	1050	1700	470	150	220	3090	1370	450	200
0.999	f_{m1}	AR	9570	9910	9920	9900	9060	9690	9810	9930	8060	9180	9730	9730
		RR	290	80	60	70	600	200	150	70	1370	570	220	210
		GR	140	10	20	30	320	110	40	0	480	240	50	60
		LS	0	0	0	0	20	0	0	0	90	10	0	0
	f_{m2}	AR	3180	7870	8310	8300	230	4590	6800	8430	0	1690	4250	6220
		RR	5710	1970	970	450	8340	4870	3040	1450	8470	7260	5280	3520
		GR	40	0	0	10	100	30	0	10	150	120	30	60
		LS	1070	160	720	1240	1330	510	160	110	1380	930	440	200
0.9999	f_{m1}	AR	9930	9980	10,000	9990	9750	9910	9950	9980	9280	9810	9950	9920
		RR	50	20	0	0	190	60	30	20	550	130	50	80
		GR	20	0	0	10	60	30	20	0	160	60	0	0
		LS	0	0	0	0	0	0	0	0	10	0	0	0
	f_{m2}	AR	3070	7610	8420	8530	300	4510	6980	8160	0	1740	4220	6230
		RR	5830	2260	890	370	8190	4930	2830	1600	8840	7060	5340	3540
		GR	0	20	0	0	20	0	10	0	10	10	0	10
		LS	1100	110	690	1100	1490	560	180	240	1150	1190	440	220

$$\text{MSE}\ (\hat{\alpha}_{LS}) = \hat{\sigma}^2\sum_{i=1}^{p}1/\lambda_i \tag{8}$$

There are different methods for estimating k that exists in the present literature. However, following we listed some of the well known methods for choosing ridge parameter value to compute RR estimator used in simulation study.

$$k_1 = p\hat{\sigma}^2/\sum_{i=1}^{p}\hat{\alpha}_i^2 \quad \text{(Hoerl et al., 1975)} \tag{9}$$

$$k_2 = \frac{\hat{\sigma}^2}{\left(\prod_{i=1}^{p}\hat{\alpha}_i^2\right)^{1/p}} \quad \text{(Kibria, 2003)} \tag{10}$$

$$k_3 = \text{Median}\left(\frac{\hat{\sigma}^2}{\hat{\alpha}_i^{2\prime}}\right) \quad i = 1, 2, ..., p \quad \text{(Kibria, 2003)} \tag{11}$$

$$k_4 = (\lambda_{\max}\hat{\sigma}^2)/((n-p-1)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_{\max}^2) \quad \text{(Khalaf and Shukur, 2005)} \tag{12}$$

$$k_5 = \left(\prod_{i=1}^{p}\sqrt{\hat{\alpha}_i^2/\hat{\sigma}^2}\right)^{\frac{1}{p}} \quad \text{(Muniz and Kibria, 2009)} \tag{13}$$

$$k_6 = \text{Median}\left(\prod_{i=1}^{p}\sqrt{\hat{\alpha}_i^2/\hat{\sigma}^2}\right) \quad \text{(Muniz and Kibria, 2009)} \tag{14}$$

$$k_7 = p\hat{\sigma}_1^2/\sum_{i=1}^{p}\hat{\alpha}_i^2 \quad \text{where,} \quad \hat{\sigma}_1^2 = \frac{Y'Y - \hat{\alpha}'_{LS}Z'Y}{n-p-1} \quad \text{(Khalaf, 2012)} \tag{15}$$

Also, in case of generalized ridge regression, the following well known method for determination of ridge parameter for each regressor, given by Hoerl and Kennard (1970), is used to compute $\hat{\alpha}_{GR}$.

$$k_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}, \quad i = 1, 2, ..., p \tag{16}$$

where, $\hat{\alpha}_i$ is the ith element of $\hat{\alpha}_{LS}$, $i = 1, 2, ..., p$ and $\hat{\sigma}^2$ is the LS estimator of σ^2 i.e. $\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}'_{LS}Z'Y}{n-p-1}$.

3. Proposed estimator

The Ridge Regression (RR) estimator proposed by Hoerl and Kennard (1970) is such an estimator widely used by statisticians in the presence of multicollinearity. However, RR

Table 3 "f_m" for AR and RR estimators for a different choice of k (p = 4 and β = (4, 2, 10, 1, 3)′).

ρ	Estimator	k	n = 20				100				500			
			σ^2 = 1	9	25	100	1	9	25	100	1	9	25	100
0.6	AR		40	2070	2810	1590	0	100	940	1950	0	0	10	410
	RR	k_1	580	940	630	460	240	970	930	1080	60	330	660	860
		k_2	2570	2560	2280	1780	1640	2310	2710	2640	700	1660	2230	2780
		k_3	3390	2470	1650	1140	3770	3510	2930	2220	4460	3770	3370	3110
		k_4	230	580	430	430	100	350	440	550	70	120	270	340
		k_5	0	90	930	2610	0	0	10	130	0	0	0	0
		k_6	0	60	350	1160	0	0	0	90	0	0	0	0
		k_7	0	0	0	0	0	0	0	0	0	0	0	0
0.99	AR		2450	5300	4930	4290	240	3030	4570	5090	0	750	2100	3060
	RR	k_1	700	210	110	90	960	620	300	210	520	1120	810	480
		k_2	2990	1680	1260	840	3430	2790	2510	1710	2690	3180	3230	2900
		k_3	2510	1500	1570	1570	3310	2200	1600	1420	4000	3070	2480	2170
		k_4	390	450	230	140	420	600	400	380	200	530	440	470
		k_5	0	350	1020	1780	0	0	110	500	0	0	0	0
		k_6	0	190	700	1130	0	0	0	350	0	0	0	0
		k_7	0	0	0	0	0	0	0	0	0	0	0	0
0.999	AR		5660	7140	7150	7270	2830	6070	6920	7190	690	4020	5460	6590
	RR	k_1	280	80	90	60	440	160	90	70	1060	500	240	100
		k_2	2000	750	660	490	3210	1360	1090	760	3270	2490	2020	1250
		k_3	1520	1770	1720	1490	2200	1620	1530	1760	3150	1950	1490	1570
		k_4	400	150	130	20	530	600	180	120	530	480	520	350
		k_5	0	0	130	460	0	0	0	0	0	0	0	0
		k_6	0	10	60	140	0	0	0	0	0	0	0	0
		k_7	0	0	0	0	0	0	0	0	0	0	0	0
0.9999	AR		6790	7770	8410	8460	6110	7100	7540	7700	4440	6520	7050	7210
	RR	k_1	100	10	20	30	130	130	60	50	470	70	110	100
		k_2	860	530	310	300	1890	780	630	500	2610	1240	750	840
		k_3	1970	1620	1190	1170	1420	1730	1620	1680	1570	1650	1810	1600
		k_4	120	20	30	20	300	130	50	60	470	360	190	150
		k_5	0	0	10	20	0	0	0	0	0	0	0	0
		k_6	0	0	0	0	0	0	0	0	0	0	0	0
		k_7	0	0	0	0	0	0	0	0	0	0	0	0

estimator has some disadvantages; mainly it is a nonlinear function of the ridge parameter (or biasing constant) k.

This leads to complicated equations, when k is selected. There is no explicit formula for this ridge parameter. Many authors proposed different approximations for it. The conventional wisdom is that no single method would be uniformly better than all the others. Also, as pointed out by Liu (2003) when there exits sever multicollinearity the ridge parameter k selected for ridge regression may not fully remedy the problem of multicollinearity. To avoid calculating the value of k in this article, we suggest the modification in the Ridge Regression (RR) estimator proposed by Hoerl and Kennard (1970) by avoiding the determination of optimal ridge parameter k. Now the idea is that the correlation coefficient between the regressors is helpful in detecting the near linear dependency between the same pairs of regressors only which plays an important role in detecting problem of multicollinearity. Rodgers and Nicewander (1988) present a longer review of ways to interpret the correlation coefficient. Also, as interpreted by Nefzger and Drasgow (1957), for the bivariate data (X, Y) when we standardize the two raw variables, the standard deviations become unity and the slope of the regression line of Y on X becomes the correlation coefficient. Obliviously, $Z'Y$ is the vector of correlation coefficient between Z and Y. By using the same vector with modification in RR estimator, we proposed a new estimator of α which is termed as Adjusted Ridge (AR) Estimator and is given by:

$$\hat{\alpha}_{AR} = [\Lambda + C]^{-1} Z'Y \quad \text{where,} \quad C = \text{diagonal}\left[(|Z'Y|)^{1/2}\right]$$

or

$$\hat{\alpha}_{AR} = [I - CA^{-1}]\, \hat{\alpha}_{LS} \quad \text{where,} \quad A = (\Lambda + C)$$

Hence, Adjusted Ridge Estimator of β is:

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

3.1. Bias, variance and MSE of $\hat{\alpha}_{AR}$

Bias of $\hat{\alpha}_{AR}$:

$$\begin{aligned} \text{Bias}(\hat{\alpha}_{AR}) &= E[\hat{\alpha}_{AR}] - \alpha \\ &= -CA^{-1}\alpha \end{aligned}$$

Variance of $\hat{\alpha}_{AR}$:

$$\begin{aligned}\mathrm{Var}(\hat{\alpha}_{\mathrm{AR}}) &= E\big[(\hat{\alpha}_{\mathrm{AR}} - E(\hat{\alpha}_{\mathrm{AR}}))\ (\hat{\alpha}_{\mathrm{AR}} - E(\hat{\alpha}_{\mathrm{AR}}))'\big] \\ &= (I - CA^{-1})\sigma^2\Lambda^{-1}(I - CA^{-1})'\end{aligned}$$

where $\mathrm{MSE}(\hat{\alpha}_{\mathrm{LS}}) = V(\hat{\alpha}_{\mathrm{LS}}) = \sigma^2(Z'Z)^{-1} = \sigma^2\Lambda^{-1}$
MSE of $\hat{\alpha}_{\mathrm{AR}}$:

$$\begin{aligned}\mathrm{MSE}(\hat{\alpha}_{\mathrm{AR}}) &= V(\hat{\alpha}_{\mathrm{AR}}) + [\mathrm{Bias}(\hat{\alpha}_{\mathrm{AR}})]^2 \\ &= (I - CA^{-1})\sigma^2\Lambda^{-1}(I - CA^{-1})' + CA^{-1}\alpha\alpha' CA^{-1}\end{aligned}$$

$$\mathrm{MSE}(\hat{\alpha}_{\mathrm{AR}}) = \sum_{i=1}^{p} \frac{\sigma^2\lambda_i + (\alpha_i c_i)^2}{(\lambda_i + c_i)^2}$$

where, c_i is the ith diagonal element of C, $i = 1, 2, \ldots, p$
Or

$$\mathrm{MSE}(\hat{\alpha}_{\mathrm{AR}}) = \sum_{i=1}^{p} \frac{\hat{\sigma}^2\lambda_i + (\hat{\alpha}_i c_i)^2}{(\lambda_i + c_i)^2} \tag{17}$$

where, $\hat{\alpha}_i$ is the ith element of $\hat{\alpha}_{\mathrm{LS}}$, $i = 1, 2, \ldots, p$ and $\hat{\sigma}^2$ are the LS estimator of σ^2 i.e. $\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}'_{\mathrm{LS}} Z'Y}{n-p-1}$.

3.2. Comparison between the $\hat{\alpha}_{\mathrm{AR}}$ and $\hat{\alpha}_{\mathrm{RR}}$

It is well known that, the value of ridge parameter 'k' is chosen small enough, for which the mean squared error of RR estimator, is less than the mean squared error of LS estimator. Also most of the researchers studied comparison between RR and GR estimators. Hence, in the following, we compare our proposed estimator to the RR estimator only. Using (7) and (17) we investigate the following difference

$$\begin{aligned}\mathrm{MSE}(\hat{\alpha}_{\mathrm{RR}}) - \mathrm{MSE}(\hat{\alpha}_{\mathrm{AR}}) &= \sum_{i=1}^{p}\left[\frac{\left(\hat{\sigma}^2\lambda_i + k^2\hat{\alpha}_i^2\right)}{(\lambda_i + k)^2}\right] - \sum_{i=1}^{p}\frac{\hat{\sigma}^2\lambda_i + (\hat{\alpha}_i c_i)^2}{(\lambda_i + c_i)^2} \\ &= \sum_{i=1}^{p}\frac{\left[(\lambda_i + c_i)^2(\hat{\sigma}^2\lambda_i + k^2\hat{\alpha}_i^2) - (\hat{\sigma}^2\lambda_i + (\hat{\alpha}_i c_i)^2)(\lambda_i + k)^2\right]}{(\lambda_i + c_i)^2(\lambda_i + k)^2} \\ &= \sum_{i=1}^{p}\frac{\left\{\left[(\lambda_i + c_i)^2 - (\lambda_i + k)^2\right]\hat{\sigma}^2\lambda_i + \left[(\lambda_i + c_i)^2k^2 - (c_i)^2(\lambda_i + k)^2\right]\hat{\alpha}_i^2\right\}}{(\lambda_i + c_i)^2(\lambda_i + k)^2}\end{aligned}$$

Since the quantity $(\lambda_i + c_i)^2 - (\lambda_i + k)^2$ is always positive, from above equation, it can be shown that $\mathrm{MSE}(\hat{\alpha}_{\mathrm{RR}}) \geqslant \mathrm{MSE}(\hat{\alpha}_{\mathrm{AR}})$ if and only if $(\lambda_i + c_i)^2k^2 \geqslant (c_i)^2(\lambda_i + k)^2$.

4. Simulation study

We are now ready to illustrate the behavior of the proposed estimator via a Monte Carlo simulation. We performed our simulations with MATLAB, using different sample sizes and error variances examined the MSE of the estimators LS, RR, GR and AR for different degrees of multicollinearity. For the simulations, we supposed the regression model defined in Eq. (1). Following McDonald and Galerneau (1975) the explanatory variables are generated by

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

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, to investigate the effects of different degrees of multicollinearity on the estimators, we consider, ρ = 0.6, 0.8, 0.90, 0.95, 0.99, 0.999, 0.9999 and 0.99999. Ten thousand simulations are run for all combinations of σ^2 = 1, 9, 25 100 and n = 20, 50, 100 and 500. Here we used well known ridge parameter k_1 given by Hoerl et al. (1975). MSE of estimators computed using the following expression,

$$\mathrm{MSE}(\hat{\beta}) = \sum_{i=1}^{p}(\hat{\beta}_i - \beta_i)^2$$

.where, $\hat{\beta}_i$ denote the estimator of the ith parameter and β_i, $i = 1, 2, \ldots, p$ are the true parameter values. However, β parameter vectors are chosen arbitrarily for number of regressors p = 4. For each simulation, the dependent variables are computed by the specified protocol. Values of "$\mathrm{f_m}$" reported in Table 1, indicate the frequency with which each estimator had the lowest $\mathrm{MSE}(\hat{\beta})$. We consider the method that leads to the maximum "$\mathrm{f_m}$" to the best from the MSE point of view.

The same procedure above for another choice of p = 3 and arbitrarily chosen parameter vectors β are done and values of "$\mathrm{f_m}$" are computed and represented in Figures. Here we noted that values of "$\mathrm{f_m}$" only for RR, GR and AR are represented because these values for LS have less importance for the comparative study. To compute and represent the results with respect to the presence of moderate or extremely sever multicollinearity, we consider ρ = 0.95, 0.99, 0.999, 0.9999 and 0.99999. Here input values are n and σ^2. These input values are ordered according to the increase of values. For fixed value of 'n' changes the values of σ^2.

There are 16 sets of (n, σ^2) values. These are arranged as (20, 1), (20, 9),..., (500, 100) and it is numbered as 1, 2,..., 16 respectively. Obtained results are represented in Figs. 1–5.

In addition to demonstrate the other performances of the proposed method, we have computed the relative error sum of squares of parameters ($\mathrm{RESS}(\beta)$) as well as prediction mean square error ($\mathrm{MSE}(y)$) to show the predicting ability of the model developed by the proposed and the other estimators using following expressions.

$$\mathrm{MSE}(y) = \sum_{i=1}^{p}(y - \hat{y})^2$$

$$\mathrm{RESS}(\beta) = \sum_{i=1}^{p}\left[\frac{(\beta - \hat{\beta})}{\beta}\right]^2$$

Results on "$\mathrm{f_{m1}}$" and "$\mathrm{f_{m2}}$" which are reported in Table 2, indicate the frequencies with which each estimator had the lowest $\mathrm{RESS}(\beta)$ and $\mathrm{MSE}(y)$, respectively. We consider the method that leads to the maximum "$\mathrm{f_{m1}}$" and "$\mathrm{f_{m2}}$" to the best from the MSE point of view.

Tables 1 and 2 indicate that when multicollinearity is nonexistent with lower error variance σ^2 only (at ρ = 0.6 and σ^2 = 1) the improvement of AR is not very substantial, since in this case RR and GR are themselves fine estimators. However, when multicollinearity is moderate or sever or extremely sever, the improvement is extremely effective and dramatic, because in this case not only LS but also RR and GR perform very poorly as shown by the simulation. Especially under the situation, when multicollinearity is extremely severed; the level of multicollinearity influences the improvement of the AR over other estimators. Similarly, increasing the error variance seems to improve the accuracy of AR. However, in Tables 1 and 2, it is also seen that increasing the sample size at lower error variance for nonexistent or moderate multicollinearity GR is superior to LS, RR and AR estimators. But, for sever or extremely sever multicollinearity AR is superior to others for large sample size, even error variances are small. For sever or extremely sever multicollinearity AR is consistently superior to LS, RR and GR estimators for different combinations of size of the sample (n), variance of the error term (σ^2) and number of predictors (p). Two novel features of the proposed estimator are that its computation does not depend on any unknown ridge parameter k and it can be used without any modification in the proposed estimator. It is a better alternative to overcome the problem of multicollinearity, particularly with sever or extremely sever multicollinearity and increasing error variances in linear regression. We observe that represented results in Fig. 1 to Fig. 4 are also supported to the conclusions, drawn from Tables 1 and 2.

In case of multicollinearity, we have used RR estimator, in which ridge parameter k plays an important role. A natural question arising at this stage is that what should be the suitable choice of k and how does it influence the performance of the proposed estimator. Here, we attempt to answer this question. RR estimator computed using different ridge parameters given in 9–15 and AR estimators computed for number of regressors $p = 4$ and values of "f_m" are computed and reported in Table 3. We consider the method that leads to the maximum "f_m" to the best from the MSE point of view. Table 3, clearly indicates that choice of k to compute RR does not influence the performance of the proposed estimator AR.

5. Conclusion

This article introduces a new method for regression parameter estimation which aims at totally avoiding computational part for optimal ridge parameter k in ridge regression. Our suggested estimator is termed as AR since it is obtained by adjusting RR estimator, given by Hoerl and Kennard (1970). New estimator is a better alternative to RR and GR estimators in the presence of sever or extremely sever multicollinearity with increasing error variance in linear regression. We believe that AR is a fine estimator, not only in theory but also in practice.

Acknowledgements

The author wishes to thank the anonymous referee for his/her comments and the editor who improved the original version. The present studies were supported in part Research Award Scheme by UGC, India Project No. F. 30-1/2014/RA-2014-16-GE-MAH-5958 (SA-II).

References

Batah, F.S., Ramnathan, T., Gore, S.D., 2008. The efficiency of modified jackknife and ridge type regression estimators: a comparison. Surv. Math. App. 24 (2), 157–174.

Dorugade, A.V., 2014. New ridge parameters for ridge regression. J. Assoc. Arab Univ. Basic Appl. Sci. 15, 94–99.

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

Hoerl, A.E., Kennard, R.W., Baldwin, K.F., 1975. Ridge regression: some simulation. Commun. Stat. 4, 105–123.

Khalaf, G., 2012. A proposed ridge parameter to improve the least squares estimator. J. Mod. Appl. Stat. Methods 11, 443–449.

Khalaf, G., Shukur, G., 2005. Choosing ridge parameter for regression problem. Commun. Stat. Theor. Methods 34, 1177–1182.

Kibria, B.M.G., 2003. Performance of some new ridge regression estimators. Commun. Stat. Simul. Comput. 32 (2), 419–435.

Lawless, J.F., 1978. Ridge and related estimation procedure. Commun. Stat. 7, 139–164.

Liu, K.J., 2003. Using Liu-type estimator to combat collinearity. Commun. Stat. Theor. Methods 32, 1009–1020.

McDonald, G.C., Galarneau, D.I., 1975. A monte carlo evaluation of some ridge-type estimators. J. Am. Stat. Assoc. 70, 407–416.

Montgomery, D.C., Peck, E.A., Vining, G.G., 2001. Introduction to Linear Regression Analysis. John Wiley and Sons, New York.

Muniz, G., Kibria, B.M.G., 2009. On some ridge regression estimators: an empirical comparison. Commun. Stat. Simul. Comput. 38, 621–630.

Nefzger, M.D., Drasgow, J., 1957. The needless assumption of normality in Pearson's r. Am. Psychol. 12, 623–625.

Rodgers, J.L., Nicewander, W.A., 1988. Thirteen ways to look at the correlation coefficient. Am. Stat. 42 (1), 59–66.

Saleh, A.K.Md.E., Kibria, B.M.G., 1993. Performances of some new preliminary test ridge regression estimators and their properties. Commun. Stat. 22, 2747–2764.

Sarkar, N., 1992. A new estimator combining the ridge regression and the restricted least squares method of estimation. Commun. Stat. 21, 1987–2000.

Singh, B., Chaubey, Y.P., 1987. On some improved ridge estimators. Stat. Pap. 28, 53–67.

Stephen, G.W., Christopher, J.P., 2001. Generalized ridge regression and a generalization of the Cp statistic. J. Appl. Stat. 28 (7), 911–922.

Swindel, B.F., 1976. Good ridge estimators based on prior information. Commun. Stat. Theor. Methods 11, 1065–1075.

Yan, X., 2008. Modified nonlinear generalized ridge regression and its application to develop naphtha cut point soft sensor. Comput. Chem. Eng. 32 (3), 608–621.

Yan, X., Zhao, W., 2009. Concentration soft sensor based on modified bak propagation algorithm embedded with ridge regression. Intell. Autom. Soft Comput. 15 (1), 41–51.

Yang, H., Chang, X., 2010. A new two-parameter estimator in linear regression. Comm. Stat. Theor. Methods 39, 923–934.

Zhong, Z., Yang, H., 2007. Ridge estimation to the restricted linear model. Commun. Stat. Theor. Methods 36, 2099–2115.