Fundamental Journal of Applied Sciences
Vol. 1, Issue 1, 2016, Pages 19-32
This paper is available online at http://www.frdint.com/
Published online February 18, 2016

A RIDGE REGRESSION ESTIMATION APPROACH WHEN MULTICOLLINEARITY IS PRESENT

GHADBAN KHALAF[*] and MOHAMED IGUERNANE

Department of Mathematics
Faculty of Science
King Khalid University
Saudi Arabia

Abstract

In regression problems, we usually try to estimate the parameters β in the general linear regression model $Y = X\beta + u$. We need a method to estimate the parameter vector β. The most common method is the Ordinary Least Squares (OLS) estimator. However, in the presence of multicollinearity, the OLS efficiency can be radically reduced because of the large variances of the estimates of the regression coefficients. An alternative to the OLS estimator has been recommended by Hoerl and Kennard [3], namely the ridge regression estimator. In this paper, a suggested method of finding the ridge parameter k is investigated and evaluated in terms of Mean Square Error (MSE) by simulation techniques. Results of a simulation study indicate that with respect to MSE criteria, the suggested estimators perform better than both the OLS estimators and the other estimators discussed here.

1. Introduction and Ridge Estimation of β

In multiple regression, it is known that the parameter estimates, based on

Keywords and phrases: linear regression model, multicollinearity, ridge estimators, simulation.

2010 Mathematics Subject Classification: 62J05, 62J07.

[*]Corresponding author

Received October 15, 2015

minimum residual sum of squares, have a high probability of being unsatisfactory if the prediction vectors, X, are multicollinear. In fact, the question of multicollinearity is not one of existence, but of degree. In the situation when the prediction vectors are far from being orthogonal, i.e., when strong multicollinearities exist in X, Hoerl and Kennard [3] suggested the ridge regression to deal with the problem of estimating the regression parameters.

Consider the standard multiple linear regression model:

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

where Y is an $(n \times 1)$ vector of observable random variable (the response or dependent variable), $X = (X_1, X_2, ..., X_p)$ is a known $(n \times p)$ matrix of the explanatory variables (the regressor or independent variables) of full rank p, $\beta = (\beta_1, \beta_2, ..., \beta_p)'$ is a $(p \times 1)$ vector of unknown regression coefficients, and finally, $u \sim N(0, \sigma^2 I)$ is an $(n \times 1)$ vector of uncorrelated errors. We have left a constant term $(\beta_0$ term), in order to simplify the discussion which follows. This is actually justifiable if we center all the data (i.e., offset it so that its mean is zero, both predictors variables and the response variable).

The most common method estimator of β is derived by the OLS estimator. We find the parameter values which minimize the sum of squared residuals (SSR)

$$SSR = \sum_i |Y - X\beta|^2. \tag{2}$$

The solution turns out to be a matrix equation

$$\hat{\beta} = (X'X)^{-1} X'Y, \tag{3}$$

where X' is the transpose of the matrix X, and the exponent "-1" indicates the matrix inverse of the given quantity. We expect the true parameters to give us nearly the most likely result, so the least squares solution, by minimizing the SSR, defined by (2), gives the maximum likelihood values of the parameter vector β. From the Gauss-Markov theorem, we know that the least squares estimate gives the best linear unbiased estimator of the parameters. And that is one of the reasons least squares is so popular. Its estimate are unbiased (the expected values of the parameters are the true values), and of all the unbiased estimators, it gives the least variance.

But, there are cases, however, for which the best linear unbiased estimator is not necessarily the "best" estimator. One pertinent case occurs when two or more of the predictor variables are very strongly correlated. Thus the matrix $X'X$ has a determinant which is close to zero, which makes it ill-conditioned so the matrix cannot be inverted with as much precision as we would like, there is uncomfortably large variance in the final parameter estimates. So it may be worth sacrificing some bias to achieve a lower variance.

One approach is to use an estimator which is no longer unbiased, but can greatly reduce the variance, resulting in a better MSE. This estimator is called "ridge regression estimator". Ridge regression is like least squares but shrinks the estimated coefficients towards zero. Given a response vector *Y* and a predictor matrix *X*, the ridge regression coefficients are defined as

$$\hat{\beta}(k) = (X'X + kI_p)^{-1} X'Y, \tag{4}$$

where $k, k > 0,$ is the ridge parameter and *I* is the identity matrix. The amount of shrinkage is controlled by the ridge parameter *k*. Small positive values of *k* improve the conditioning of the problem and reduce the variance of the estimates. While biased, the reduced variance of ridge estimates often results in a smaller MSE when compared to least-squares estimates. Obviously the question is how to determine the parameter *k*. Choosing an appropriate value of *k* is important and also difficult.

For selecting the best ridge parameter estimator, in a given application, several criteria have been proposed in the literature (see for example, Hoerl and Kennard [3], Hoerl et al. [4], McDonald and Galarneau [9], Nomura [10], Hag and Kibria [2], Khalaf and Shukur [8], Muniz and Kibria [11], Khalaf [5], Khalaf [6] and Khalaf and Iguernane [7].

2. Estimators included in the Study

In this section, we discuss some formulas for determining the value of *k* to be used in (4). The classical choice is the ridge trace method, proposed by Hoerl and Kennard [3]. They suggested that the best method for achieving an improved estimate $\hat{\beta}(k)$ (with respect to MSE) is to employ a ridge trace. The ridge trace is a graph of the estimates of the regression coefficients plotted against the corresponding *k*-

values $(0 \leq k \leq 1)$ with the aid of which one selects a single value of k and a unique improved estimator for β.

In using the ridge-trace, a value of k is chosen at which the regression coefficients have reasonable magnitude, sign and stability, while the level of the MSE is not grossly inflated. In fact, letting $\beta_{\max}$ denote the maximum of β, Hoerl and Kennard [3] showed that choosing

$$\hat{k} = \frac{\hat{\sigma}^2}{\hat{\beta}^2_{\max}}, \tag{5}$$

implies that $MSE(\hat{\beta}(k)) < MSE(\hat{\beta})$, where $\hat{\sigma}^2$ is the usual estimate of σ^2, defined by

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n - p - 1}. \tag{6}$$

This estimator will be denoted by HK. Hoerl et al. [4] suggested that, the value of k is chosen small enough, for which the MSE of ridge estimator is less than the MSE of OLS estimator. They showed, through simulation, that the use of the ridge with biasing parameter given by

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \tag{7}$$

has a probability greater than 0.50 of producing estimator with a smaller MSE than the OLS estimator, where $\hat{\sigma}^2$ is the usual estimator of σ^2, defined by (6). The ridge estimator using eq. (7) will be denoted by HKB.

Alkhamisi and Shukur [1] used the estimator

$$\hat{k}_{AS} = \max\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2} + \frac{1}{\lambda_i}\right), \tag{8}$$

where λ_i, $i = 1, 2, ..., p$, is the ith eigenvalue of the matrix $X'X$. They concluded that the ridge estimator using $\hat{k}_{AS}$, given by (8), performed very well indeed and that it was substantially better than any of the other estimators included in their study. The ridge estimator using $\hat{k}_{AS}$ will be denoted by AS.

In the light of the above remarks, which indicate the satisfactory performance of $\hat{k}_{AS}$ on the one hand, and the potential for improvement on the other hand, we propose the following modification of the ridge estimator using $\hat{k}_{AS}$ to suggest the following three estimators

$$\hat{k}_1 = \sqrt{\max\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2} + \frac{1}{\lambda_i}\right)}, \tag{9}$$

$$\hat{k}_2 = \text{median}\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2} + \frac{1}{\lambda_i}\right), \tag{10}$$

$$\hat{k}_3 = \sqrt{\text{median}\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2} + \frac{1}{\lambda_i}\right)}. \tag{11}$$

The ridge estimators using $\hat{k}_1, \hat{k}_2$ and $\hat{k}_3$ will be denoted by KI_1, KI_2 and KI_3, respectively.

3. Simulation Study

In this section, we describe the simulation techniques which were used to examine the performance, relative to the OLS estimator and other ridge estimators, of the new ridge estimators $\mathrm{KI}_1, \mathrm{KI}_2$ and KI_3 using $\hat{k}_1, \hat{k}_2$ and $\hat{k}_3$, defined by, respectively, (9), (10) and (11). Since $\mathrm{KI}_1, \mathrm{KI}_2$ and KI_3 are modifications of AS, given by (8). This estimator was included for purposes of comparison in addition to the estimators HK and HKB, defined by (5) and (7), respectively.

Following McDonald and Galarneau [9], the explanatory variables are generated by

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

where z_{ij} are independent standard normal pseudo-random numbers, and ρ is specified so that the correlation between any two explanatory variables is given by ρ^2. Three different sets of correlation are considered, corresponding to $\rho = 0.85$, 0.95 and 0.99. The explanatory variables are then standardized so that $X'X$ is in correlation form.

Observations on the dependent variable are determined by

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{ip} + e_i, \quad i = 1, 2, \ldots, n,$$

where β_0 is taken to be identically zero. Five values of σ^2 are investigated which are $0.01, 0.05, 0.10, 0.5$ and 1.00. Then the dependent variable is standardized so that $X'y$ is the vector of correlation of dependent variable with each explanatory variable. In this experiment, we choose $p = 10$ and 15 for $n = 50$ and 100. Then the experiment is replicated 8000 times by generating new error terms.

3.1. Judging the performance of the estimators

To investigate the performance of the different proposed ridge regression estimators and the OLS method, we calculate the MSE using the following equation:

$$MSE = \frac{\sum_{i=1}^{R} (\hat{\beta} - \beta)'_i (\hat{\beta} - \beta)_i}{R},$$

where $\hat{\beta}$ is the estimator of β obtained from the OLS or the other different ridge parameters, and *R* equals 8000 which corresponds to the number of replications used in the simulation.

4. Results and Discussion

Ridge estimators are constructed with the aim of having smaller MSE than the MSE for the least squares. Improvement, if any, can therefore be studied by looking at the MSE of ridge estimator and that of least squares. These MSEs are reported in Tables (1) and (2).

The MSEs are always less than the MSE of the OLS for all estimators (they exceeded the MSE of the OLS for AS at certain values of σ^2 and σ). This is to say that ridge estimators dominate least squares. Further, they do not exceed the MSE of the OLS for all estimators when $\sigma^2 = 0.01, 0.05$ and 0.1 for the different values of ρ.

Table 1. The Estimated MSE when $p = 10$

$\sigma^2 = 0.01$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	11685	5146	2756	9.9998	9.9872	9.9989	9.9702
	100	5144	2314	1254	9.9991	9.9613	9.9955	9.9144

$\sigma^2 = 0.01$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	37631	15942	8545	9.9999	9.9924	9.9996	9.9795
	100	16371	7186	3810	9.9997	9.9767	9.9982	9.9403

$\sigma^2 = 0.01$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	204160	82595	43822	10	9.9969	9.9999	9.9909
	100	89874	36713	20101	10	9.9905	9.9996	9.9724

$\sigma^2 = 0.05$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	457	202	109	9.88	8.65	9.39	7.22
	100	204	94	51	9.53	6.84	7.88	4.53

$\sigma^2 = 0.05$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	1477	620	332	9.96	9.12	9.73	7.89
	100	655	284	153	9.83	7.72	8.98	5.43

$\sigma^2 = 0.05$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	8231	3372	1786	9.99	9.61	9.94	8.89
	100	3587	1498	796	9.97	8.91	9.76	7.27

$\sigma^2 = 0.1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	116	53	29	8.70	4.76	4.99	2.25
	100	51	26	14	6.92	3.03	1.86	1.14

$\sigma^2 = 0.1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	373	157	85	9.46	5.54	6.94	2.71
	100	162	72	39	8.08	2.85	3.18	0.92

$\sigma^2 = 0.1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	2059	836	455	9.90	7.50	9.15	4.66
	100	899	371	201	9.57	4.73	7.10	1.85

$\sigma^2 = 0.5$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	4.60	3.85	2.26	5.57	1.28	0.77	1.11
	100	2.03	1.88	1.33	5.59	1.02	0.54	0.77

$\sigma^2 = 0.5$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	14	9.50	4.83	3.84	076	0.63	1.12
	100	6.51	5.09	2.76	3.86	0.63	0.61	1.10

$\sigma^2 = 0.5$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	83	63	19	2.36	0.32	0.10	0.30
	100	35	17	9.50	1.99	0.28	0.27	0.71

$\sigma^2 = 1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	1.15	1.09	0.84	5.39	0.80	0.41	0.53
	100	0.51	0.50	0.44	5.34	0.57	0.327	0.325

$\sigma^2 = 1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	3.73	3.14	1.83	3.76	0.49	0.46	0.84
	100	1.63	1.51	1.05	3.86	0.38	0.27	0.55

$\sigma^2 = 1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	20	11	5.84	2.01	0.29	0.48	0.99
	100	9.09	6.39	3.28	2.03	0.25	0.55	1.07

If we focus on these values of σ^2, we find that among the ridge estimators considered, KI_1, KI_2 and KI_3 are the best followed by AS, HKB then HK. Further, the MSEs decrease as σ^2 increases, especially when $\sigma^2 = 1$ and $\rho = 0.85$.

In comparing models, exhibiting high multicollinearity and where $p = 10$ and 15, respectively, we notice that the MSEs are lowest for $p = 10$ in case of KI_3 followed by KI_2 and KI_1. This is to say that the ridge estimators are more helpful when high multicolliearity exists, especially when σ^2 is not too small and *n* is large.

Table 2. The Estimated MSE when $p = 15$

$\sigma^2 = 0.01$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	20748	2992	4581	14.9997	14.9787	14.9978	14.9375
	100	8304	4105	1936	14.9987	14.9333	14.9907	14.8201

$\sigma^2 = 0.01$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	65721	29824	13617	14.9999	14.9870	14.9991	14.9556
	100	27106	12627	5925	14.9996	14.9598	14.9961	14.8727

$\sigma^2 = 0.01$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	365230	159770	73413	15.00	14.9946	14.9998	14.9791
	100	146910	65668	30841	14.9999	14.9831	14.9992	14.9388

$\sigma^2 = 0.05$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	525	398	184	14.85	12.79	13.76	9.62
	100	334	168	80	14.35	9.84	10.84	5.29

$\sigma^2 = 0.05$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	2664	1204	561	14.95	13.54	14.54	10.77
	100	1072	508	240	14.77	11.12	12.88	6.51

$\sigma^2 = 0.05$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	14701	6509	2940	14.99	14.35	14.87	12.69
	100	5935	2641	1253	14.95	13.19	14.48	9.51

$\sigma^2 = 0.1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	206	101	48	12.33	6.85	6.18	2.40
	100	84	46	22	10.99	4.46	1.94	1.30

$\sigma^2 = 0.1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	660	303	141	14.31	7.93	9.18	2.82
	100	269	128	61	12.37	3.88	3.40	0.84

$\sigma^2 = 0.1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	3649	1579	725	14.87	10.85	13.22	5.32
	100	1489	667	312	14.40	6.40	9.28	1.73

$\sigma^2 = 0.5$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	8.23	6.95	3.63	9.49	1.89	0.97	1.61
	100	3.33	3.11	2.05	9.59	1.51	0.66	1.13

$\sigma^2 = 0.5$								
ρ	*N*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	26	17	7.67	6.69	1.09	0.82	1.63
	100	10.76	8.57	4.22	6.74	0.81	0.79	1.59

$\sigma^2 = 0.5$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	143	65	30	4.20	0.48	0.12	0.43
	100	58	31	14	3.70	0.37	0.38	1.07

$\sigma^2 = 1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.85	50	2.04	1.94	1.38	9.19	1.12	0.51	0.80
	100	0.83	0.81	0.69	9.31	0.91	0.38	0.49

$\sigma^2 = 1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.95	50	6.53	5.50	2.85	6.66	0.64	0.59	1.23
	100	2.74	2.55	1.62	6.85	0.49	0.34	0.81

$\sigma^2 = 1$								
ρ	*n*	*OLS*	*HK*	*HKB*	*AS*	KI_1	KI_2	KI_3
0.99	50	36	21	9.28	3.52	0.33	0.65	1.46
	100	14.78	10.65	4.96	3.67	0.28	0.76	1.58

5. Summary and Conclusions

Several procedures for constructing ridge estimators have been proposed in the literature. These procedures were aiming at a rule for selecting the constant k in equation (4).

The results of our simulation indicate that the estimators KI_1, KI_2 and KI_3, suggested by us, performed well in this study. They outperform the estimator AS and they are also considerably better than both HK and HKB. Also, they appeared to offer an opportunity for large reduction in MSE, especially when the degree of multicollinearity is high. Since the potential reduction using the ridge estimators is measured by the MSE, then the performance of KI_1, KI_2 and KI_3, in comparison with the other estimators included in our simulation study, is very good from this point of view, see the Tables 1 and 2.

References

[1] M. Alkhamisi and G. Shukur, Developing ridge parameters for SUR model, Commun. Stat. Theory Methods 37 (2008), 544-564.

[2] M. S. Hag and B. M. G. Kibria, A shrinkage estimator for the restricted linear regression ridge regression approach, J. Appl. Stat. Sci. 3 (1996), 301-316.

[3] A. E. Hoerl and R. W. Kennard, Ridge regression: biased estimation for non-orthogonal problems, Technometrics 12 (1970), 55-67.

[4] A. E. Hoerl, R. W. Kennard and K. F. Baldwin, Ridge regression: some simulation, Commun. Stat. Theory Methods 4 (1975), 105-124.

[5] G. Khalaf, Ridge regression: an evaluation to some new modifications, Int. J. Stat. Anal. 1(4) (2011), 325-342.

[6] G. Khalaf, A comparison between biased and unbiased estimators, J. Modern Appl. Stat. Methods 12(2) (2013), 293-303.

[7] G. Khalaf and M. Iguernane, Ridge regression and ill-conditioning, J. Modern Appl. Stat. Methods 13(2) (2014), 355-363.

[8] G. Khalaf and G. Shukur, Choosing ridge parameters for regression problems, Commun. Stat. Theory Methods 34 (2005), 1177-1182.

[9] G. C. McDonald and D. I. Galarneau, A Monte Carlo evaluation of some ridge-type estimators, J. Amer. Stat. Assoc. 70 (1975), 407-416.

[10] M. Nomura, On the almost unbiased ridge regression estimation, Commun. Stat. Theory Methods 17 (1988), 729-743.

[11] G. Muniz and B. M. G. Kibria, On some ridge regression estimators: an empirical comparison, Commun. Stat. Simul. Comput. 38 (2009), 621-630.