Journal of Modern Applied Statistical Methods

Volume 15 | Issue 2 Article 25

11-1-2016

Multicollinearity and A Ridge Parameter Estimation Approach

Ghadban Khalaf
King Khalid University, albadran50@yahoo.com

Mohamed Iguernane
King Khalid University, Saudi Arabia, mohamed.iguernane@gmail.com

Journal of Modern Applied Statistical Methods
November 2016, Vol. 15, No. 2, 400-410.
doi: 10.22237/jmasm/1478002980

ISSN 1538 − 9472

Multicollinearity and a Ridge Parameter Estimation Approach

Ghadban Khalaf
King Khalid University
Abha, Saudi Arabia

Mohammed Iguernane
King Khalid University
Abha, Saudi Arabia

One of the main goals of the multiple linear regression model, $Y = X\beta + u$, is to assess the importance of independent variables in determining their predictive ability. However, in practical applications, inference about the coefficients of regression can be difficult because the independent variables are correlated and multicollinearity causes instability in the coefficients. A new estimator of ridge regression parameter is proposed and evaluated by simulation techniques in terms of mean squares error (MSE). Results of the simulation study indicate that the suggested estimator dominates ordinary least squares (OLS) estimator and other ridge estimators with respect to MSE.

Keywords: OLS, ridge regression, multicollinearity, simulation; MSC 62J07, 62J05

Introduction

Consider the general linear regression model

$$Y = \beta_0 1 + X\beta + u \tag{1}$$

where Y is an $(n \times 1)$ vector of observations on the dependent variable, β_0 is a scalar intercept, 1 is an $(n \times 1)$ vector with all components equal to unity, X is an $(n \times p)$ matrix of regression variables of full rank p, β is the unknown parameter vector of regression coefficients, and $u \sim N(0, \sigma^2 I)$ is an $(n \times 1)$ vector of unobservable errors. Because the interest is in estimating β, omit the constant term β_0 in order to keep the notation simple.

The OLS estimator for the regression parameters is given by

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y \tag{2}$$

Dr. Khalaf is a Professor in the Department of Mathematics. Email him at albadran50@yahoo.com. Dr. Iguernane is an Assistant Professor in the Department of Mathematics. Email him at Mohamed.iguernane@gmail.com.

If any X's are highly correlated (or, multicollinear), the matrix becomes non-orthogonal, the inversion unstable and the inverse or estimated fractions highly sensitive to random error, and therefore, the OLS solution in (2) has inflated values of the coefficients of regression. Such a regression can be used for prediction, but is worthless in the analysis and interpretation of the individual predictors role in the model. In practice, multicollinearity almost always exists but is typically overlooked or ignored. The following overview stages the later proposed approaches.

Multicollinearity

Multicollinearity is a high degree of correlation among several independent variables. It commonly occurs when a large number of independent variables are incorporated in a regression model. Only existence of multicollinearity is not a violation of the OLS assumptions. However, a perfect multicollinearity violates the assumption that the X matrix is full ranked, making OLS, given by (2), impossible, because when the model, defined by (1), is not full ranked, then the inverse of X cannot be defined, there can be an infinite number of least squares solutions. Symptoms of multicollinearity may be observed in the following situations:

1. Small changes in the data produce wide swings in the parameters estimates.
2. Coefficients may have very high standard errors and low significance levels even though they are jointly significant and the R^2 for the regression is high.
3. Coefficients may have the wrong sign or implausible magnitude, Green (2000).

The consequences of multicollinearity are that the variance of the model (i.e. the error sum of squares) and the variances of coefficients are inflated. As a result, any inference is not reliable and the confidence interval becomes wide. Hence, even though the OLS estimator of β is the minimum variance unbiased estimator, its MSE will still be large if multicollinearity exists among the independent variables.

To detect multicollinearity, in fact there is no clear-cut criterion for evaluating multicollinearity of linear regression models. We may compute correlation coefficients of independent variables. But high correlation coefficients do not necessarily imply multicollinearity. We can make a judgment by checking related statistics, such as variance inflation factor (VIF) and condition number (CN), where

Variance Inflation Factor

The VIF is given by

$$VIF = \frac{1}{1-R_i^2},\ i = 1, 2, \ldots, p \tag{3}$$

and R_i^2 represents the squared multiple correlation coefficients when X_i (the i^{th} column of X) is regressed on the remaining $(p - 1)$ regressor variables.

The VIF shows how multicollinearity has increased the instability of the coefficient estimates (Freund and Littell, 2000). In other words, it tells us how inflated the variance of the coefficient is, compared to what it would be if the variable were uncorrelated with any other variable in the model (Allison, 1999). However, there is no formal criterion for determining the bottom line of the VIF. Some argue that VIF greater than 10 roughly indicates significant multicollinearity. Others insist that magnitude of model's R^2 be considered determining significance of multicollinearity. Klein (1962) suggested an alternative criterion that R_i^2 (the coefficient of determination for regression of the i^{th} independent variable) exceeds R^2 of the regression model. In this vein, if VIF is greater than $1/(1 - R^2)$, then multicollinearity can be considered statistically significant.

Condition Number

To quantify the seriousness of multicollinearity, computation of the eigenvalues, λ_i, of the matrix $X'X$ is recommended, because the degree of collinearity of any data set is indicated the CN, which is given by

$$CN = \frac{\lambda_1}{\lambda_p} \tag{4}$$

where λ_1 is the largest eigenvalue of the matrix $X'X$ and λ_p is the smallest eigenvalue of $X'X$.

A set of eigenvalues of relatively equal magnitudes indicates that there is little multicollinearity (Freund and Littell, 2000). A zero eigenvalue means perfect collinearity among independent variables and very small eigenvalues implies severe multicollinearity. In other words, an eigenvalue close to zero (less than 0.01, say) or CN greater than 50 indicates significant multicollinearity. Belsley et al. (1980) insist 10 to 100 as a beginning, and maintains that collinearity affects estimates.

There are several ways to solve the problem of multicollinearity. Some of them are

1. Changing specification by omitting or adding independent variables.
2. Obtaining more data (observations) if problems arise because of a shortage of information.
3. Transforming independent variables by taking logarithmic or exponential.
4. Trying biased estimated methods such as ridge regression estimation. The ridge regression estimator has a covariance matrix smaller than that of OLS (Judge, et al., 1985)

Ridge Regression and a New Proposed Ridge Parameter

Although the OLS estimator is BLUE, it is not necessarily closest to β, because linearity and unbiasedness are not irrelevant for closeness, particularly when the input matrix of the design is multicollinear. For orthogonal data, the OLS estimator for β in the linear regression model is strongly efficient (getting estimates with minimum MSE). But in the presence of multicollinearity, the OLS efficiency can be reduced and hence an improvement upon it would be necessary and desirable. Thus it is natural to look at biased estimator for an improvement over the OLS estimator because it is meaningful to focus on small MSE as the relevant criterion, if a major reduction in variance can be obtained as a result of allowing a little bias. This is precisely what the ridge regression estimator can accomplish.

Ridge regression, due to Hoerl and Kennard (1970), amounts to adding a small positive quantity, say k, to each of the diagonal elements of the matrix $X'X$. The resulting estimator is

$$\hat{\beta}(k) = (X'X + kI)^{-1} X'Y \tag{5}$$

where k is a positive scalar. When $k = 0$, (5) reduces to the unbiased OLS estimator given by (2).

Considering $\hat{\beta}(k)$ with regards to MSE

$$MSE\left(\hat{\beta}(k)\right) = Var\left(\hat{\beta}(k)\right) + Bias^2\left(\hat{\beta}(k)\right) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^2} + \sum_{i=1}^{p} \frac{k^2 \beta_i^2}{(\lambda_i + k)^2}$$

It is known that, as k increases from zero, the MSE initially decreases to a minimum, and then increases with increasing k. Hence, there always exists a minimum. Thus it is quite helpful allowing a small bias in order to achieve the main criterion of keeping the MSE small.

When using ridge estimates, the choice of k in (5) is important and several methods have been proposed for this purpose (see, e.g., Hoerl & Kennard, 1970; McDonald & Galarneau, 1975; Nomura, 1988; Hag & Kibria, 1996; Khalaf & Shukur, 2005; Muniz & Kibria, 2009; Khalaf, 2011; Khalaf, 2013; Khalaf & Iguernane, 2014).

Hoerl and Kennard (1970) suggested that the best method for achieving an improved estimate (with respect to MSE) is by choosing

$$\hat{k} = \frac{\hat{\sigma}^2}{\hat{\beta}_{\max}^2} \tag{6}$$

where $\hat{\beta}_{\max}$ denote the maximum of β_i and σ^2 is the usual estimate of σ^2, defined by

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

and referred to henceforth as the HK estimator. They proved that there exists a $k > 0$ such that the sum of the MSEs of all $\hat{\beta}_i(k)$ is smaller than the corresponding term of $\hat{\beta}_i$, the OLS estimator, i.e.

$$MSE\left(\hat{\beta}(k)\right) < MSE\left(\hat{\beta}\right) = \hat{\sigma}^2 \sum_{i=1}^{p} \lambda_i^{-1}$$

Khalaf and Shukur (2005) suggested a new method of estimating k as a modification of equation (6), as follows

$$\hat{k}_{KS} = \frac{\lambda_{\max}\hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_{\max}\hat{\beta}_{\max}^2} \tag{7}$$

where $\lambda_{\max}$ is the largest eigenvalue of the matrix $X'X$. They concluded the ridge estimator using (7) performed very well and was substantially better than any estimators included in their study.

In the light of above, which indicates the satisfactory performance of $\hat{k}_{KS}$ with the potential for improvement, modification of the ridge estimator using $\hat{k}_{KS}$ (the KS estimator) by taking its square root is suggested. This proposed estimator (the KSM estimator) is

$$\hat{k}_{KSM} = \sqrt{\hat{k}_{KS}} \tag{8}$$

To investigate the performance, relative to the OLS and other ridge estimators given by (6) and (7), of the new ridge estimator given by (8), we calculate the MSE using the following equation

$$MSE = \frac{\sum_{i=1}^{R}\left(\hat{\beta}-\beta\right)_i'\left(\hat{\beta}-\beta\right)}{R} \tag{9}$$

where $\hat{\beta}$ is the estimator of β obtained from OLS or other ridge estimators, and R equals 5000 which corresponds to the number of replicates used in the simulation.

Simulations

Consider the true model $Y = X\beta + u$. Here $u \sim N(0,\sigma^2 I)$ and the independent variables are generated from

$$x_{ij} = \left(1-\rho^2\right)^{\frac{1}{2}} z_{ij} + \rho z_{ip},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, p \tag{10}$$

where z_{ij} are generated using the standard normal distribution. Here, we consider four values of ρ corresponding to 0.7, 0.9, 0.95 and 0.99. The dependent variable is then determined by

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + u_i,\ i = 1, 2, \ldots, n \tag{11}$$

where n is the number of observations, u_i are *i.i.d*. pseudo-random numbers, and β_0 is taken to be zero. Parameter values are chosen such that $\sum_{j=1}^{p} \beta_j^2 = 1$, which is a common restriction in simulation studies (McDonald and Galarneau, 1975; Muniz and Kibria, 2009). Sample sizes selected are n = 10, 25, 50, 85, 200 and 1000, with 4 or 7 independent variables. The variance of the error terms is taken as σ^2 = 0.01, 0.1, and 0.5. Ridge estimates are computed using the different ridge parameters given in (6) and (7). Because the proposed estimator (8) is a modification of (7), this estimator is included for purposes of comparison. The MSE of the ridge regression parameters is obtained using (9). This experiment is repeated 5000 times.

Result

All factors chosen to vary in the design of the experiment affect the estimated MSE. As expected, increasing the degree of correlation leads to a higher estimated MSE, especially when n is small and σ^2 = 0.01. This increase is much greater for OLS than for ridge regression estimators.

Table 1a. Estimated MSE when p = 4 and ρ = 0.7

	σ^2=0.01				σ^2=0.1				σ^2=0.5			
n	*OLS*	*HK*	*KS*	*KSM*	*OLS*	*HK*	*KS*	*KSM*	*OLS*	*HK*	*KS*	*KSM*
10	16114	5236	6140	31	156.00	52.00	60.00	7.00	6.320	3.030	3.220	1.850
25	3799	1242	2153	27	39.00	15.00	23.00	5.90	1.560	1.170	1.240	0.990
50	1722	597	1248	32	17.00	7.00	12.00	5.00	0.690	0.600	0.620	0.560
85	988	344	806	36	9.70	4.60	8.00	4.10	0.390	0.360	0.370	0.340
200	399	141	363	42	4.00	2.40	3.60	2.60	0.161	0.156	0.157	0.153
1000	77	28	76	35	0.77	0.67	0.75	0.70	0.032	0.031	0.031	0.031

Table 1b. Estimated MSE when $p = 4$ and $\rho = 0.9$

	σ^2=0.01				σ^2=0.1				σ^2=0.5			
n	***OLS***	***HK***	***KS***	***KSM***	***OLS***	***HK***	***KS***	***KSM***	***OLS***	***HK***	***KS***	***KSM***
10	46391	14512	15254	41	478.0	149.0	156.0	8.0	18.000	6.700	7.000	2.500
25	11854	3692	4695	29	114.0	37.0	46.0	5.7	4.700	2.500	2.700	1.600
50	5179	1678	2607	27	52.0	18.0	27.0	5.3	2.120	1.480	1.560	1.170
85	2967	969	1778	25	29.0	11.0	18.0	4.9	1.190	0.950	0.990	0.820
200	1184	380	885	26	12.0	5.1	9.2	4.0	0.482	0.439	0.446	0.410
1000	233	75	216	36	2.3	1.6	2.2	1.7	0.094	0.092	0.093	0.090

Table 1c. Estimated MSE when $p = 4$ and $\rho = 0.95$

	σ^2=0.01				σ^2=0.1				σ^2=0.5			
n	***OLS***	***HK***	***KS***	***KSM***	***OLS***	***HK***	***KS***	***KSM***	***OLS***	***HK***	***KS***	***KSM***
10	99744	29610	30311	51	957.00	282.00	289.00	9.00	39.000	12.000	13.000	3.000
25	24979	7538	8527	32	240.00	74.00	84.00	6.00	9.000	4.100	4.400	2.000
50	10642	3290	4305	26	108.00	36.00	46.00	5.40	4.330	2.380	2.570	1.570
85	6109	1945	2925	23	60.00	20.00	29.00	5.00	2.480	1.650	1.760	1.250
200	2498	802	1543	22	24.00	9.00	15.00	4.60	1.010	0.830	0.858	0.724
1000	494	163	426	31	4.82	2.60	4.21	2.64	0.192	0.185	0.186	0.179

Table 1d. Estimated MSE when $p = 4$ and $\rho = 0.99$

	σ^2=0.01				σ^2=0.1				σ^2=0.5			
n	***OLS***	***HK***	***KS***	***KSM***	***OLS***	***HK***	***KS***	***KSM***	***OLS***	***HK***	***KS***	***KSM***
10	533881	156406	157056	84	5352.0	1605.0	1612.0	12.0	218.0	67.0	67.3	5.0
25	130105	39322	40154	46	1325.0	417.0	425.0	7.4	54.0	16.0	17.0	3.0
50	59142	18290	19221	32	593.0	189.0	199.0	6.5	23.0	8.0	8.4	2.5
85	33685	10461	11481	25	330.0	105.0	160.0	5.7	13.0	5.1	5.4	2.1
200	13727	4394	5464	17	137.0	43.0	54.0	5.1	5.4	2.7	3.0	1.6
1000	2637	814	1575	16	26.0	9.0	16.0	4.4	1.0	0.8	0.9	0.7

Table 2a. Estimated MSE when $p = 7$ and $\rho = 0.7$

	σ^2=0.01				σ^2=0.1				σ^2=0.5			
n	***OLS***	***HK***	***KS***	***KSM***	***OLS***	***HK***	***KS***	***KSM***	***OLS***	***HK***	***KS***	***KSM***
10	74818	24592	25042	110	768.00	238.00	242.00	19.00	29.00000	10.00000	11.00000	4.20000
25	8804	3457	4423	46	89.00	37.00	46.00	10.00	3.54000	2.76000	2.81000	2.13000
50	3618	1508	2367	48	36.00	17.00	24.00	8.70	1.44000	1.31000	1.32000	1.17000
85	1998	848	1506	52	19.00	10.00	15.00	7.40	0.78300	0.74400	0.74800	0.69900
200	795	337	691	63	7.90	5.50	7.00	4.80	0.31700	0.31100	0.31200	0.30300
1000	152	67	148	60	1.52	1.39	1.48	1.35	0.06110	0.06094	0.06096	0.06060

Table 2b. Estimated MSE when p = 7 and ρ = 0.9

	σ^2=0.01				σ^2=0.1				σ^2=0.5			
n	*OLS*	*HK*	*KS*	*KSM*	*OLS*	*HK*	*KS*	*KSM*	*OLS*	*HK*	*KS*	*KSM*
10	235966.0	68291.0	68644.0	136.0	2224.0	658.0	661.0	27.0	91.0000	28.1000	28.2000	6.4000
25	26871.0	10240.0	11090.0	49.0	273.0	105.0	113.0	12.0	10.0000	6.2000	6.3000	3.5000
50	10990.0	4275.0	5224.0	39.0	110.0	45.0	54.0	10.0	4.3800	3.2900	3.3400	2.3900
85	6112.0	2430.0	3321.0	38.1	59.0	25.0	33.0	8.8	2.4200	2.0500	2.0700	1.6700
200	2430.0	966.0	1624.0	40.0	23.0	11.0	16.0	7.0	0.9790	0.9120	0.9170	0.8300
1000	466.0	185.0	410.0	57.0	4.6	3.5	4.2	3.1	0.1878	0.1852	0.1854	0.1816

Table 2c. Estimated MSE when p = 7 and ρ = 0.95

	σ^2=0.01				σ^2=0.1				σ^2=0.5			
n	*OLS*	*HK*	*KS*	*KSM*	*OLS*	*HK*	*KS*	*KSM*	*OLS*	*HK*	*KS*	*KSM*
10	516796	152429	152764	171	4818.0	1430.0	1434.0	35.0	192.000	62.400	62.600	9.300
25	57214	21072	21887	55	582.0	219.0	227.0	15.0	23.000	10.000	11.000	4.500
50	22961	8791	9736	41	231.0	91.0	100.0	12.0	9.200	5.600	5.800	3.300
85	12508	4916	5857	35	126.0	50.0	59.0	10.0	5.000	3.600	3.700	2.500
200	5037	1977	2795	34	50.0	21.0	29.0	8.4	2.010	1.730	1.740	1.430
1000	985	396	771	49	9.8	6.1	8.0	4.7	0.389	0.377	0.378	0.361

Table 2d. Estimated MSE when p = 7 and ρ = 0.99

	σ^2=0.01				σ^2=0.1				σ^2=0.5			
n	*OLS*	*HK*	*KS*	*KSM*	*OLS*	*HK*	*KS*	*KSM*	*OLS*	*HK*	*KS*	*KSM*
10	2501132	764126	764446	235	25773	7976	7979	62	1019.0	289.3	289.4	18.0
25	314693	115277	116046	72	3077	1107	1115	21	126.0	48.4	48.7	8.7
50	128529	48265	49173	48	1259	475	484	17	50.0	20.4	20.7	6.0
85	67913	25511	26492	38	691	262	272	15	28.0	12.8	13.0	5.0
200	27914	10645	11673	31	271	102	112	11	11.0	6.3	6.5	3.6
1000	5479	2117	2922	32	53	22	29	8	2.1	1.7	1.8	1.4

Conclusion

Based on the result from the simulation study, some recommendations are warranted. The KSM is usually among the estimators with the lowest estimated MSE, especially when $\rho = 0.95$ and $p = 7$. Also, regardless of the degree of correlations, KSM is the best among the considered ridge estimators, followed by HK, and then KS, specifically when the sample size is high, $n = 1000$, and $\sigma^2 = 0.5$.

Several procedures for constructing ridge estimators have been proposed in the literature. These procedures aim at establishing a rule for selecting the constant k in equation (5). Nevertheless, to date there is no rule for choosing k that assures that the corresponding ridge estimator is better than OLS estimator.

The proposed choice of k, the ridge regression parameter defined by (8), was shown through simulation to yield a lower MSE than $\hat{\beta}$ for all β, as noted in Tables 1 and 2. The estimators HK and KS, which were evaluated in other simulation studies, also performed well. However, the superiority of the suggested estimator KSM over the estimators HK and KS was observed, especially at the large values of n and σ^2. In general, the OLS estimator has larger estimated MSE values than all estimators considered, and the proposed estimator given by (8) performs very well and has the lowest MSE when compared with the other ridge estimators. This is to say that ridge estimators are more helpful when high multicollinearity exists, especially when σ^2 is not too small.

References

Allison, P. D. (1999). *Logistic regression using the SAS System: Theory and application*. Cary, NC: SAS Institute.

Belsley, D. A., Edwin, K. and Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley and Sons.

Freund, R. J. and Littell, R. C. (2000). *SAS System for regression* (3rd ed.). Cary, NC: SAS Institute.

Green, W. H. (2000). *Econometric Analysis* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hag, M. S. and Kibria, B. M. G. (1996). A shrinkage estimator for the restricted linear regression ridge regression approach. *Journal of Applied Statistical Science, 3*, 301–316.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics, 12*(1), 55–67. doi: 10.1080/00401706.1970.10488634

Judge, G. G., Griffiths, W. E., Hill, R. C. and Lee, T. C. (1985). *The theory and practice of econometrics* (2nd ed.) New York: John Wiley and Sons.

Klein, L. (1962). *An introduction to econometrics*. New York: Prentice Hall.

Khalaf, G. (2011). Ridge regression: An evaluation to some new modifications. *International Journal of Statistics and Analysis*, *1*(4), pp. 325–342.

Khalaf, G. (2013). A comparison between biased and unbiased estimators. *Journal of Modern Applied Statistical Methods*, *12*(2), 293–303.

Khalaf, G. and Iguernane, M. (2014). Ridge regression and ill-conditioning. *Journal of Modern Applied Statistical Methods*, *13*(2), 355–363.

Khalaf, G. and Shukur, G. (2005). Choosing ridge parameter for regression problems. *Communication in Statistics – Theory and Methods*, *34*(5), 1177–1182. doi: 10.1081/STA-200056836

McDonald, G. C., and Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, *70*(350), 407–416. doi: 10.1080/01621459.1975.10479882

Nomura, M. (1988). On the almost unbiased ridge regression estimation. *Communication in Statistics – Simulation and Computation*, *17*(3), 729–743. doi: 10.1080/03610918808812690

Muniz, G. and Kibria, B. M. G. (2009). On some ridge regression estimators: An empirical comparison. *Communication in Statistics – Simulation and Computation*, *38*(3), 621–630. doi: 10.1080/03610910802592838