



Improving the Ordinary Least Squares Estimator by Ridge Regression

Ghadban Khalaf

Department of Business Administration, Mazaya University College, Nasiriyah, Iraq

Email: ac.ghadban@mpu.edu.iq

How to cite this paper: Khalaf, G. (2022) Improving the Ordinary Least Squares Estimator by Ridge Regression. *Open Access Library Journal*, 9: e8738. <https://doi.org/10.4236/oalib.1108738>

Received: April 17, 2022

Accepted: May 24, 2022

Published: May 27, 2022

Copyright © 2022 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In the presence of multicollinearity, ridge regression techniques result in estimated coefficients that are biased but have smaller variance than Ordinary Least Squares estimators and may, therefore, have a smaller Mean Squares Error (MSE). The ridge solution is to supplement the data by stochastically shrinking the estimates toward zero. In this study, we propose a new estimator to reduce the effect of multicollinearity and improve the estimation. We show by a simulation study that the MSE of the suggested estimator is lower than other estimators of the ridge and the OLS estimators.

Subject Areas

Mathematical Statistics

Keywords

OLS Estimator, Multicollinearity, Ridge Regression, Simulation

1. Introduction

The Ordinary Least Squares (OLS) method is one of the most frequently applied statistical procedures in application. It is well documented that the OLS method is extremely unreliable in parameter estimation while the independent variables are dependent (multicollinearity problem). Multicollinearity is the existence of a correlation between independent variables in modeled data. It can magnify the standard errors in the regression coefficients and reduce the efficiency of any t-tests. In addition, it can also produce deceiving results and p -values and increase the redundancy of a model, making its predictability inefficient and less reliable.

Consider the general linear regression model:

$$Y = X\beta + e, \quad (1)$$

where Y is an $(n \times 1)$ vector of observations on the dependent variable. X is $(n \times p)$ matrix of observations on p nonstochastic independent variables, β is the $(p \times 1)$ vector of parameters associated with the p regressors and e is an $(n \times 1)$ vector of disturbances having mean zero and variance-covariance matrix $\sigma^2 I_n$.

The OLS regression uses the following formula to estimate coefficients, given by:

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

It is well established that the OLS estimator given by (2), under the above assumptions about the error term, is Best Linear Unbiased Estimator (BLUE). However, multicollinearity can result in ill-conditioning of the matrix $X'X$ rendering the OLS estimator undesirable. In addition, least squares regression isn't defined at all when the number of predictors exceeds the number of observations; it doesn't differentiate "important" from "less-important" predictors in a model, so it includes all of them. This leads to over-fitting a model and failure to find unique solutions.

Ridge regression avoids all of these problems since it works with the advantage of not requiring unbiased estimators-rather, and it adds bias to estimators to reduce the standard error, making the estimates a reliable representation of the population of data.

The purpose of the research is to find a new estimator that improves the MSE to be smaller than that of the OLS estimator and it is organized as follows: The effect, detection and correction of multicollinearity are discussed in Section 2. The various types of ridge estimators and the proposed estimator are described in Section 3. Section 4 describes the simulation technique that we have adopted in our study to evaluate the performance of the new values of the ridge parameter as we suggested. The results of the simulation study, which appear in the tables, are presented in Section 5. Finally, some conclusions drawn from the present research are reported in Section 6.

2. Multicollinearity

Multicollinearity, or collinearity, is the existence of near-linear relationships among the independent variables. Effects of Multicollinearity can create inaccurate estimates of the regression coefficients, inflate the standard errors of the regression coefficients, deflate the partial t-tests for the regression coefficients, give false, non-significant, p -values, and degrade the predictability of the model (and that's just for starters). Sources of multicollinearity to deal with multicollinearity, one must be able to identify its source. The source of the multicollinearity impacts the analysis, the corrections, and the interpretation of the linear model. The detection of multicollinearity is key to the reduction of standard errors in models for predictability efficiency. There are five sources (see Montgomery

(1982) [1] for details).

Data collection may cause multicollinearity when it is sourced using an inappropriate sampling procedure. The data may come from a smaller subset than expected, hence, the effect. Population or model constraints cause multicollinearity due to physical, legal, or political constraints, which are natural, regardless of the type of sampling method used. Over-defining a model will also cause multicollinearity due to the existence of more variables than observations. It is avoidable during the development of a model. The model's choice or specification can also cause multicollinearity due to the use of independent variables previously interacting in the initial variable set. Outliers are extreme variable values that can cause multicollinearity. The multicollinearity can be reversed by the elimination of the outliers before applying ridge regression.

The detection of multicollinearity is key to the reduction of standard errors in models for predictability efficiency. First, one can detect by investigating independent variables for correlation in pairwise scatter plots. High pairwise correlations of independent variables can mean the presence of multicollinearity. Secondly, one can detect multicollinearity through the consideration of Variance Inflation Factors (VIFs), given by:

$$\text{VIF} = (1 - R_j^2)^{-1}, \quad (3)$$

where R_j^2 is the coefficient of determination in the regression of explanatory variable X_j on the remaining explanatory variables of the model. A VIF score of 10 or more shows that variables are collinear. Thirdly, one can detect multicollinearity by checking the eigenvalues of $X'X$ in correlation form. When at least one eigenvalue is close to zero, then multicollinearity is existed.

Multicollinearity correction depends on the cause. When the source of collinearity is data collection, for example, the correction will involve collecting additional data from the proper subpopulation. If the cause is the linear model choice, the correction will include simplifying the model by the proper variable selection methods. If the causes of multicollinearity are based on certain observations, these observations should be eliminated. Ridge regression is also an effective eliminator of multicollinearity.

3. Ridge Regression

Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors, making the estimates reasonably reliable approximations to true population values. It is hoped that the net effect will be to give estimates that are more reliable.

In fact, there is another biased regression technique, called principal components regression but ridge regression is the more popular of the two methods.

For the sake of convenience, we assume that the matrix X is standardized in such a way that XX' is a non-singular correction matrix. Thus Equation (1) becomes:

$$Y = Z\alpha + e, \quad (4)$$

where $Z = XT$, $\alpha = T'\beta$ (Montgomery *et al.*, 2006, [2]). This implies that $Z'Z = \Delta = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, where λ_i being the i th eigenvalue of XX' and T be the matrix of the eigenvectors of XX' such that $T'T = TT' = I_p$. The OLS estimator of α is given by;

$$\hat{\alpha} = (Z'Z)^{-1} Z'Y \quad (5)$$

and the ordinary ridge regression of α is defined by:

$$\hat{\alpha}(k) = (Z'Z + kI_p)^{-1} Z'Y, \quad (6)$$

where k is the ridge parameter and I is the identity matrix. A tuning parameter (k) controls the strength of the penalty term. When $k = 0$, ridge regression equals least squares regression. If $k = \infty$, all coefficients are shrunk to zero. The ideal penalty is therefore somewhere in between 0 and ∞ .

The total MSE of the regression coefficient $\hat{\alpha}(k)$ in the presence of multicollinearity is given by:

$$\text{MSE}(\hat{\alpha}(k)) = \text{Variance of } \hat{\alpha}(k) + (\text{Bias } \hat{\alpha}(k))^2. \quad (7)$$

A main challenge in the literature has been finding the appropriate value of k , because choosing a value for k is not a simple task, which is perhaps one major reason why ridge regression isn't used as much as least squares. Several criteria have been proposed in the literature (see for example, Hoerl & Kennard (1970) [3], Hoerl *et al.* (1975) [4], Gibbons (1981) [5], Saleh & Kibria (1993) [6], Kibria (2003) [7], Khalaf & Shukur (2005) [8], Dorugade & Kashid (2010) [9], Khalaf (2013) [10], Khalaf & Iguernane (2014) [11], Fujii (2018) [12], Yuzbasi (2020) [13], Tsagris *et al.* (2021) [14]).

Hoerl and Kennard (1970) [3] found that the best method for achieving a better estimator $\hat{\alpha}(k)$ is to use $k_i = k$ for all i , and they suggested k to be:

$$\hat{k}_{\text{HK}} = \frac{\hat{\sigma}^2}{\max(\hat{\alpha}_i^2)}. \quad (8)$$

They showed that the estimator, given by (8), is sufficient to give ridge estimators with smaller MSEs than the OLS estimator. This estimator will be denoted by HK.

Hoerl *et al.* (1975) [4] argued that a reasonable choice of k is:

$$\hat{k}_{\text{HKB}} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}, \quad (9)$$

Khalaf and Shukur (2005) [8] suggested a new method to estimate the ridge parameter k , given by:

$$\hat{k} = \frac{\lambda_{\max} \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_{\max} \hat{\alpha}_{\max}^2}, \quad (10)$$

which guaranteed lower MSE, where λ_{\max} is the maximum eigenvalue of the matrix $Z'Z$.

In this article, we propose a modification of the Hoerl and Kennard (1970) [3] estimator shown in (8) to obtain a new estimator, given by:

$$\hat{k}_{\text{GK}} = \frac{\hat{\sigma}^2}{\max(\hat{\alpha}_i^2)} + \sqrt{\frac{1}{\frac{\lambda_{\max} + \lambda_{\min}}{2}}}$$

$$\hat{k}_{\text{GK}} = \frac{\hat{\sigma}^2}{\max(\hat{\alpha}_i^2)} + \sqrt{\frac{2}{\lambda_{\max} + \lambda_{\min}}}, \quad (11)$$

where $\lambda_{\max}, \lambda_{\min}$ are the largest and smallest eigenvalues of the matrix $Z'Z$, respectively. For this estimator we will use the acronym GK. Because

$\sqrt{\frac{2}{\lambda_{\max} + \lambda_{\min}}} > 0$, then GK is greater than HK.

4. The Simulation Study

In this section, we present our simulation study regarding the properties of the OLS estimator, the Hoerl and Kennard estimator given by (8) and our suggested estimator defined by (11). These properties of these estimates will be compared in terms of MSEs. To compare between these three methods, we prefer those/that who give the smallest MSE. A number of factors can affect these properties. The sample size (n), degree of correlation (ρ) between the explanatory variables and the error variance (σ^2) are three such factors. In this article, we will study the consequences of varying n , degree of correlation and the error variances, while the value of the parameters are chosen to be ones.

Now our primary interest lies in investigating the properties of our proposed approach to minimize the MSE and thus the different degree of correlation between the variables included in the models has been used. We put these values equal to 0.7, 0.9, 0.95, and 0.99. These values will cover a wide range of moderate and strong correlation between the variables.

The variance of the error form can take an infinite number of forms. We, however, enforced the errors to have low and moderate variances equal to 0.1, 0.5 and 1.

To investigate the effect of sample sizes, we used samples of the sizes 10, 20, 50 and 80, which may cover situations of small, moderate and large samples with 4 and 6 explanatory variables.

To investigate the performance of Hoerl and Kennard estimator, our suggested estimator and the OLS, we calculate the MSE using the following equation:

$$\text{MSE} = \frac{1}{R} \sum_{i=1}^R (\hat{\alpha} - \alpha)_i^2, \quad (12)$$

where $\hat{\alpha}$ is the estimate of α obtained from the OLS, HK or GK and R equals 5000 which corresponds to the number of replication used in the simulation study.

5. The Results of Simulation

In this section, we present the result of our simulation study concerning the properties of our proposed approach for choosing the ridge parameter k , when multicollinearity among the columns of the design matrix exists. Our primary interest lies in comparing the MSEs of two suggested method for choosing the ridge parameter k that is used in this study, *i.e.*, the HK and GK. The results of our study are presented in **Table 1** and **Table 2**.

The comparison is mainly done by calculating the MSEs and we consider the method that leads to the minimum MSE to be the best from the MSE point of view. Looking at **Table 1**, *i.e.*, when $p = 4$, we can see that both of HK and GK are better than the OLS estimator and that the proposed estimator GK produces lower MSEs than the HK estimator, especially when ρ is high, *i.e.*, when $\rho = 0.95$ and 0.99 .

The results also reveal that the GK estimator performs extremely better than the HK estimator. On the other hand, when $\rho = 0.7$, the performance of the OLS estimator becomes better but never superior to the *GK* suggested estimator except when $n = 50, 80$ and $\sigma^2 = 1$. When looking at **Table 2**, *i.e.*, when $p = 6$, we see that our suggested method performs much better than both of the others.

This result may, however, show the good performance of our method and its

Table 1. Estimated MSE when $p = 4$.

		$\sigma^2 = 0.1$			$\sigma^2 = 0.5$			$\sigma^2 = 1$		
ρ	n	OLS	HK	GK	OLS	HK	GK	OLS	HK	GK
0.7	10	168	112	26.50	3.94	3.55	0.75	1.43	1.39	0.48
	20	46.25	24.80	72.70	1.22	1.05	0.49	0.49	0.47	0.40
	50	12.40	8.09	4.39	0.55	0.53	0.38	0.19	0.183	0.36
	80	5.50	3.98	1.05	0.41	0.42	0.36	0.113	0.102	0.354
0.9	10	471	309	389	16.8	14.07	0.43	3.43	3.31	0.32
	20	155	93	169	4.13	3.67	0.39	1.35	1.28	0.23
	50	49.4	31.4	25.7	1.99	1.63	0.26	0.45	0.41	0.246
	80	27	19	0.71	1.22	1.19	0.22	0.202	0.201	0.163
0.95	10	1002	619	4830	39.71	35.59	0.40	8.55	8.31	0.30
	20	308	179	255	10.61	9.87	0.26	3.15	3.11	0.23
	50	105	65.5	8.34	3.92	3.16	0.22	1.09	1.08	0.22
	80	59	44	0.59	1.99	1.95	0.21	0.58	0.54	0.19
0.99	10	5560	3200	3450	210	189	0.89	51.25	49.66	0.18
	20	1730	1060	25.6	70.69	68.16	0.19	16.18	16.09	0.16
	50	590	390	18	22.17	21.89	0.16	4.67	4.61	0.15
	80	355	265	109	12.80	12.15	0.169	3.44	3.43	0.151

Table 2. Estimated MSE when $p = 6$.

ρ	n	$\sigma^2 = 0.1$			$\sigma^2 = 0.5$			$\sigma^2 = 1$		
		OLS	HK	GK	OLS	HK	GK	OLS	HK	GK
0.7	10	415	225	1081	16.32	15.12	1.13	4.41	4.35	0.65
	20	93	45	1889	2.89	3.81	0.69	0.98	0.91	0.48
	50	25.14	16.35	189.2	1.09	1.01	0.50	0.273	0.271	0.35
	80	14.78	11.43	2.51	0.63	0.62	0.41	0.192	0.190	0.29
0.9	10	1320	690	2231	52.11	49.79	0.66	13.35	13.34	0.43
	20	290	149	256	10.67	10.12	0.53	3.11	3.05	0.33
	50	88.25	59.61	4.58	3.58	3.44	0.32	0.79	0.77	0.28
	80	50.66	39.18	1.16	2.11	2.09	0.29	0.54	0.536	0.267
0.95	10	2650	1480	1580	105	101	0.53	25.17	24.89	0.27
	20	619	329	6440	23.18	22.08	0.29	5.19	5.11	0.23
	50	189	125	6.80	6.17	6.09	0.22	1.61	1.59	0.19
	80	104	79	0.79	3.89	3.81	0.19	1.15	1.07	0.17
0.99	10	14235	7690	2066	578	567	0.24	141	136	0.15
	20	3234	1656	6314	129	121	0.17	33.23	32.66	0.14
	50	1044	725	29.80	39.35	38.11	0.14	9.16	9.06	0.13
	80	611	466	0.35	23.55	23.13	0.139	5.98	5.71	0.139

robustness against some situations where the other methods behave badly.

6. Conclusion

This paper considers several estimators for estimating the biasing parameter (k) in the study of linear models in the presence of multicollinearity. After exhibiting the MSE of the ridge estimator, a simulation study has been conducted to compare the performance of the estimators. The investigation has been done using the simulation where in addition to the different multicollinearity levels, the number of observations and the error variances have been varied. For each combination, we have used 5000 replications. The evaluation of our method has been done by comparing the MSEs between our proposed method and that of the OLS and the HK estimators. Results show that the proposed estimator, given by (11), uniformly dominates the other estimators.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Montgomery, D.C. (1982) Economic Design of an \bar{X} Control Chart. *Journal of Quality Technology*, **14**, 40-43. <https://doi.org/10.1080/00224065.1982.11978782>
- [2] Montgomery, D.C., Peck, E.A. and Vining, G.G. (2006) Introduction to Linear Regression Analysis. John Wiley & Sons, Hoboken.
- [3] Hoerl, A.E. and Kennard, R.W. (1970) Ridge Regression: Biased Estimation for Non-Orthogonal Problems. *Technometrics*, **12**, 55-67. <https://doi.org/10.1080/00401706.1970.10488634>
- [4] Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. (1975) Ridge Regression: Some Simulation. *Communications in Statistics—Theory and Methods*, **4**, 105-124. <https://doi.org/10.1080/03610917508548342>
- [5] Gibbons, D.G. (1981) A Simulation Study of Some Ridge Estimators. *Journal of the American Statistical Association*, **76**, 131-139. <https://doi.org/10.1080/01621459.1981.10477619>
- [6] Saleh, A.K. and Kibria, B.M. (1993) Performance of Some New Preliminary Test Ridge Regression Estimators and Their Properties. *Communications in Statistics—Theory and Methods*, **22**, 2747-2764. <https://doi.org/10.1080/03610929308831183>
- [7] Kibria, B.M.G. (2003) Performance of Some New Ridge Regression Estimators. *Communications in Statistics—Theory and Methods*, **32**, 419-435. <https://doi.org/10.1081/SAC-120017499>
- [8] Khalaf, G. and Shukur, G. (2005) Choosing Ridge Parameters for Regression Problems. *Communications in Statistics—Theory and Methods*, **34**, 1177-1182. <https://doi.org/10.1081/STA-200056836>
- [9] Dorugade, A.V. and Kashid, D.N. (2010) Alternative Method for Choosing Ridge Parameter for Regression. *International Journal of Applied Mathematical Sciences*, **4**, 447-456.
- [10] Khalaf, G. (2013) A Comparison between Biased and Unbiased Estimators. *Journal of Modern Applied Statistical Methods*, **12**, 293-303. <https://doi.org/10.22237/jmasm/1383279360>
- [11] Khalaf, G. and Iguernane, M. (2014) Ridge Regression and Ill-Conditioning. *Journal of Modern Applied Statistical Methods*, **13**, 355-363. <https://doi.org/10.22237/jmasm/1414815420>
- [12] Fujii, K. (2018) Least Squares Method from the View Point of Deep Learning. *Advances in Pure Mathematics*, **8**, 485-493. <https://doi.org/10.4236/apm.2018.85027>
- [13] Yuzbasi, B., Arashi, M. and Ejaz Ahmed, S. (2020) Shrinkage Estimation Strategies in Generalized Ridge Regression Models: Low/High-Dimension Regime. *International Statistical Review*, **88**, 229-251. <https://doi.org/10.1111/insr.12351>
- [14] Tsagris, M. and Pandis, N. (2021) Multicollinearity. *American Journal of Orthodontics and Dentofacial Orthopedics*, **159**, 695-696. <https://doi.org/10.1016/j.ajodo.2021.02.005>