

A Review on the Biasing Parameters of Ridge Regression Estimator in LRM

MADHULIKA DUBE^{1,*} AND ISHA^{2,†}

¹Professor & Head Department of Statistics, M. D. University, Rohtak

²Research Scholar Department of Statistics M. D. University, Rohtak

*Email: madhulikadube@gmail.com; †Email: ishahooda@gmail.com

Received: August 26, 2014| Revised: August 26, 2014| Accepted: August 26, 2014

Published online: September 20, 2014

The Author(s) 2014. This article is published with open access at www.chitkara.edu.in/publications

Abstract: Ridge regression is the one of the most widely used biased estimator in the presence of multicollinearity, preferred over unbiased ones since they have a larger probability of being closer to the true parametric value. Being the modification of the least squares method it introduces a biasing parameter to reduce the length of the parameter under study. As these biasing parameters depend upon the unknown quantities, extensive work has been carried out by several authors to work out the best one. Owing to the fact that over the years a large numbers of biasing parameters have been proposed and studied, this article presents an annotated bibliography along with the review on various biasing parameter available.

Keywords: Ordinary Least Squares Estimator, Multicollinearity, Ridge Regression, Biasing Parameter.

1. Introduction

Multicollinearity in linear regression models is a rule rather than exception. Near to strong collinearity is often seen in many economic, biological and other phenomena under study, clearly violating the usual assumption of independence of explanatory variable in linear regression models. Coined by Frisch [7], the term “multicollinearity” primarily referred to the existence of perfect or exact relationship among some or all of the explanatory variables in the regression model. However, over the years it was given a broader prospective by also considering the situations where a high degree of relationships among the explanatory variables exist [see; e.g. Johnston [13], Gujarati[8]]. The consequences of multicollinearity are often very serious from inability to

Mathematical Journal of
Interdisciplinary Sciences
Vol. 3, No. 1,
September 2014
pp. 73–82

Dube, M
Isha

estimate the unique effects of individual variables in the regression model to getting large sampling variances - all leading to erroneous inferences and consequently faulty predictions.

The problem of multicollinearity has been dealt and discussed in detail in several research articles and text books [see; e.g. Maddala [18] and references cited there in]. Once multicollinearity is detected in the data, a natural question is “how to estimate the coefficients in its presence?” Apart from several ad-hoc solutions, the most widely recognized and immensely used technique is the Ridge Regression proposed by Hoerl and Kennard [10,11]. The Ridge Regression estimator is the modification of the least squares method and allows biased estimator of the regression coefficients, introducing a biasing parameter to reduce the length of the parameter under study. These biased estimators are preferred over unbiased ones since they have a large probability of being closer to the true parametric value.

However, it is more simply said than done. The choice of biasing parameters, which are unknown in general, creates a lot of problems in selection of ridge regression estimator as these are invariably dependent on the sample data. A plethora of research articles suggesting different methods of selection of biasing parameters have been suggested in literature. This review article, presents an annotated bibliography on various choices of biasing parameters of ridge regression.

The plan of the article is as follows. Section 2, defines the model and estimator and Section 3 describes the properties of the estimators and presents a review of various choices of biasing parameters. Section 4 contains the final remarks on the studied biasing parameters.

2. The Model and the Estimator

Consider a linear regression model

$$y = X\beta + \epsilon \quad (2.1)$$

where y is an $n \times 1$ vector of observations on a response variable, X is a full column rank non stochastic matrix of n observations on p -predictor (or regressor) variables. β is a $p \times 1$ associated vector of unknown regression coefficients and ϵ is an $n \times 1$ vector of errors, the elements of which are identically and independently defined each having mean zero and variance σ^2 so that, $E(\epsilon) = 0$ and $V(\epsilon) = \sigma^2 I_n$

Application of least squares to (2.1) yields the ordinary least squares OLS estimator

$$b = (X'X)^{-1} X'y \quad (2.2)$$

which is well known to be minimum variance unbiased estimator of β . In the presence of multicollinearity, though the ordinary least squared(OLS) estimator remains unbiased, the variances of OLS estimates of the parameters of collinear variables are quite large. This consequently results in large confidence intervals and erroneous inferences about the parameter vector.

In view of this Hoerl and Kennard[11] proposed the ordinary ridge regression given by

$$\hat{\beta}_0 = (X'X + kI)^{-1} X'y \quad (2.3)$$

where k is a positive scalar characterizing the estimator. It may be noted that the ordinary ridge regression estimator uses a fixed biasing parameter for all the coefficients. It is interesting to note that when $k = 0$, $\hat{\beta}_0 \rightarrow b$ while as $k \rightarrow \infty$, the estimator $\hat{\beta}_0 \rightarrow 0$, the null vector which is termed by Judge *et al.* [14] as a desperation measure.

Sometimes, for the sake of convenience, the matrix X is standardized in such a way that $X'X$ is a non-singular correlation matrix. For this purpose, let Λ and T be the matrices of eigen values and eigen vectors of $X'X$ respectively, satisfying

$$T'X'XT = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p) \quad (2.4)$$

where λ_i is the i^{th} eigen value of $X'X$. Using it (2.1) can be written as

$$y = Z\alpha + \epsilon; Z = XT, \alpha = T'\beta \quad (2.5)$$

As $Z'Z = \Lambda$ and the OLS estimator of α is given by

$$\begin{aligned} a &= (Z'Z)^{-1} Z'y \\ &= \Lambda^{-1} Z'y \end{aligned} \quad (2.6)$$

giving the OLSE of β from the transformed model (2.5). Contrary to ordinary ridge regression(ORR), the generalized ridge regression (GRR) estimator of Hoerl and Kennard [10,11] uses distinct biasing parameters for all the coefficients so that the GRR estimator of β from (2.1) is given by

$$\hat{\beta}_G = (\Lambda + K)^{-1} X'y$$

which eventually leads to the estimator of α as

$$\begin{aligned} \hat{\alpha}_G &= (\Lambda + K)^{-1} X'y \\ \hat{\alpha}_G &= (I - KA^{-1})a \end{aligned} \quad (2.7)$$

where $A = (\Lambda + K)$ and $K = \text{diag}(k_1, k_2, \dots, k_p)$; $k_i \geq 0; i = 1, 2, \dots, p$.

3. Properties of Estimators and Choices of Biasing Parameters

The least squares estimator b of β is well known to be unbiased with variance $\sigma^2 (X'X)^{-1}$. The ridge regression estimators are however biased with bias of GRR estimator $\hat{\alpha}_G$ given by

$$\beta(\hat{\alpha}_G) = -K(\Lambda + K)^{-1} \alpha \quad (3.1)$$

and the mean squared error is given by

$$\begin{aligned} M(\hat{\alpha}_G) &= E(\hat{\alpha}_G - \alpha)(\hat{\alpha}_G - \alpha)' \\ &= \sigma^2 (I - KA^{-1})(Z'Z)^{-1} (I - KA^{-1}) + KA^{-1} \alpha \alpha' A^{-1} K \end{aligned} \quad (3.2)$$

Using (3.2), the risk under quadratic error loss may be computed as

$$\begin{aligned} R(\hat{\alpha}_G) &= \text{tr} M(\hat{\alpha}_G) \\ R(\hat{\alpha}_G) &= \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^p \frac{k_i^2 \alpha_i^2}{(\lambda_i + k_i)^2} \end{aligned} \quad (3.3)$$

Interestingly, the ordinary ridge regression estimator of α and its properties can be easily obtained by simply substituting $K = k I$. In above expressions in particular, the risk of ORR estimator of α is given by

$$R(\hat{\alpha}_0) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2} \quad (3.4)$$

The value of k , the biasing parameter, is chosen in such a manner that the risk of ridge estimator is smaller than that of the least squares estimator [see e.g. Hoerl and Kennard [10,11]]

Clearly, (2.7) is a generalization of (2.3), therefore, we concentrate only on the biasing parameters introduced in the former. Differentiating (3.3) with respect to k_i and substituting to zero gives the value of k_i for which the risk of GRR estimator is minimum. This suggests to use the value of k_i

$$k_i = \frac{\sigma^2}{\alpha_i^2}; i = 1, 2, \dots, p \quad (3.5)$$

where α_i is the i^{th} value of vector α .

As (3.5) involves unknown parameters, Hoerl and Kennard [11] suggested to use the unbiased estimators of σ^2 and α_i so that their estimator of k_i becomes

$$\hat{k}_i = \frac{\hat{\sigma}_i^2}{\hat{\alpha}_i^2}; i = 1, 2, \dots, p \quad (3.6)$$

It is pertinent to note that Hoerl and Kennard[11] worked with a fixed biasing parameter. They suggested choosing k using an iterative procedure in such a way that the ridge coefficients stabilize. Further they showed that

$$\hat{k}_i \leq \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2} \quad (3.7)$$

where $\hat{\alpha}_{\max}$ is the maximum value of the estimator.

Using empirical Bayesian approach, Dempster[3] suggested to compute the value of k by solving

$$\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{\hat{\sigma}^2 \left[\frac{1}{k} + \frac{1}{\lambda_i} \right]} = p \quad (3.8)$$

While Sclove[23] suggested to compute k from the following equation

$$\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{\hat{\sigma}^2 \left[\frac{1}{k} + \frac{1}{\lambda_i} \right]} = p \hat{\sigma}^2 \left[\frac{n-p}{n-p-2} \right] \quad (3.9)$$

Using Monte Carlo simulation, McDonald & Galarrneau[19] devised a technique based on largest and smallest eigen value of design matrix X for choosing an optimum value of the biasing parameter. They observed that if the coefficient based on largest eigenvalue was used, the ridge estimator based on selected biasing parameter performs at least as well as the OLS estimates for all collinearity and residual variance levels. However, their rules were a poor prediction of optimal values of biasing parameter.

For a GRR, Hoerl *et al.* [12] suggested to use the following estimator of k_i 's

$$k_{HKB} = \frac{p \hat{\sigma}^2}{\sum_{i=1}^p \hat{\alpha}_i^2} \quad (3.10)$$

Thisted[24] found that (3.10) seems to over shrink the estimator towards zero and hence came out with an estimator of k_i as

$$k_T = \frac{(p-2) \hat{\sigma}^2}{\sum_{i=1}^p \hat{\alpha}_i^2} \quad (3.11)$$

Also working with GRR, Hocking *et al.*[9] showed that for known optimal k_p , it is superior to all other estimators within the class of biased estimators they considered. They also proposed to take the biasing parameter as

$$k_{HSL} = \frac{\hat{\sigma}^2 \sum_{i=1}^p (\lambda_i \hat{\alpha}_i)^2}{\sum_{i=1}^p (\lambda_i \hat{\alpha}_i^2)^2} \quad (3.12)$$

Dube, M
Isha

Forwarding the ideas of (3.12), Lawless and Wang [17] proposed the biasing parameter to be

$$k_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \lambda_i \hat{\alpha}_i^2} \quad (3.13)$$

Using the concept of predictive error sum of squares (PRESS), Wahba *et. al* [25] suggested generalized cross validation as a method for choosing a good ridge parameter. Interestingly, this estimate does not require an estimate of error variance.

In an interesting paper, Nomura [22] proposed the biasing parameters for both, the GRR as well as ORR, for GRR it is given by

$$k_{iNO} = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \left\{ 1 + \left[1 + \lambda_i \left(\frac{\hat{\alpha}_i^2}{\hat{\sigma}^2} \right)^{\frac{1}{2}} \right] \right\} ; i = 1, 2, \dots, p \quad (3.14)$$

Interestingly, these choices provide smaller MSE than that obtained by using (3.6), the initial choice proposed by Hoerl and Kennard [11].

Using a new approach, Firinguetti [6] proposed the biasing parameter to be computed as

$$k_{iF} = \frac{\lambda_i \hat{\sigma}^2}{\lambda_i \hat{\alpha}_i^2 + (n-p) \hat{\sigma}^2} ; i = 1, 2, \dots, p \quad (3.15)$$

which are interestingly bounded. Firinguetti [6] demonstrated that such a choice of biasing parameter ensures that the corresponding GRR estimator is better behaved than that proposed by Hoerl and Kennard [11].

Kibria[16] proposed three new estimators of the biasing parameters in GRR using A.M, G.M and median of those given by Hoerl and Kennard[11]. These are given by

$$k_{KAM} = \frac{1}{p} \sum_{i=1}^p \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \quad (3.16)$$

$$k_{KGM} = \frac{\hat{\sigma}^2}{\left(\prod_{i=1}^p \hat{\alpha}_i^2 \right)^{1/p}} \quad (3.17)$$

and

$$k_{KMed} = \text{Median} \left(\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \right) ; i = 1, 2, \dots, p \quad (3.18)$$

respectively.

Using simulation, he demonstrated that when signal to noise ratio is large the performance of OLS is reasonably better than all the proposed estimators for all degree of multicollinearity condition. Otherwise, all of the proposed

estimators have smaller MSE than the OLS estimator. K_{KAM} and K_{KGM} perform equally well and they are slightly better than (3.10). He also finds that K_{KGM} , the one which uses geometric mean of parameters, is the most preferred biasing parameter among those proposed by him.

Modifying the ridge parameter (3.15), Khalaf and Shukur [15] studied the properties of a new estimator by choosing the ridge parameter as

$$K_{KS} = \frac{\lambda_{\max} \hat{\sigma}^2}{[(n - \rho - 1) \hat{\sigma}^2 + \lambda_{\max} \hat{\alpha}_{\max}^2]} \quad (3.19)$$

The investigation has been done using Monte Carlo methods, where in addition to the different multicollinearity levels, the number of observations and the error variances have been varied. Comparing the results with (3.6) they showed that their estimator performs better in the situations where the error variance is high.

In a recent work, Alkhamisi and Shukur [2] proposed the biasing parameter for GRR as

$$k_{AS} = \max \left(\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} + \frac{1}{\lambda_i} \right) ; i = 1, 2, \dots, p \quad (3.20)$$

Inspired by proposal of several biasing parameters by Kibria[16], Muniz and Kibria[20] and most recently Muniz *et al.*[21] proposed several estimators of k_i . While Muniz and Kibria[20] first forwarded the following estimator

$$k_{KM1} = \left(\prod_{i=1}^p \frac{\lambda_i \hat{\sigma}^2}{(n - p) \hat{\sigma}^2 + \lambda_i \hat{\alpha}_i^2} \right)^{1/2} \quad (3.21)$$

and then letting $m_i = \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}$, they proposed to use various values of the biasing parameter among which the following were recommended to the practitioners

$$k_{KM4} = \left(\prod_{i=1}^p \left(\frac{1}{m_i} \right) \right)^{1/p} \quad (3.22)$$

$$k_{KM5} = \left(\prod_{i=1}^p m_i \right)^{1/p} \quad (3.23)$$

$$k_{KM6} = \text{median} \left(\frac{1}{m_i} \right) \quad (3.24)$$

Continuing with the similar ideas Muniz *et al.* [21] again forwarded several ideas for selection of k_i 's. However they recommended to use

Dube, M
Isha

$$k_{KM8} = \max \left(\frac{1}{\sqrt{\frac{\lambda_{\max} \sigma^2}{(n-p)\sigma^2 + \lambda_{\max} \alpha_i^2}}} \right) \quad (3.25)$$

$$k_{KM12} = \text{median} \left(\frac{1}{\sqrt{\frac{\lambda_{\max} \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_{\max} \hat{\alpha}_i^2}}} \right) \quad (3.26)$$

to work with the models with large residual variances since these performs better in such conditions.

Combining the concept of (3.12), Hassan and Yazid [1] proposed to use

$$k_{NHSL} = k_{HSL} + \frac{1}{\lambda_{\max}} \quad (3.27)$$

and demonstrated that the corresponding GRR estimator uniformly dominates those based on (3.6) and (3.12).

Another innovative idea for the choice of biasing parameter is forwarded by Dorugade and Kashid [5] using variance inflation factor, which is given by

$$k_{DK} = \max \left(0, \frac{p\sigma^2}{\sum_{i=1}^p \alpha_i^2} - \frac{1}{n(VIF_j)_{\max}} \right) \quad (3.28)$$

Where $VIF_j = \frac{1}{1-R_j^2}$; $j=1,2,\dots,p$ is the variance inflation factor of j^{th} regressor

This estimator is based on number data points and strength of multicollinearity in the data. Through simulation study the authors compares the ratio of average MSE with ridge parameter proposed in (3.10) and (3.19) and have demonstrated the superiority of their estimator.

The most recent work is by Dorugade[4] who has listed several choices of k_i 's and has also proposed to use

$$k_{AD} = \frac{2\hat{\sigma}^2}{\lambda_{\max} \hat{\alpha}_i^2} \quad (3.29)$$

in case of GRR while several alternative estimators using different averages are suggested for ORR.

4. Conclusion

The history of the choice of biasing parameters in Ridge Regression initiated by Hoerl and Kennard [11] is long and checkered. Over the years, researchers have proposed several alternative estimators and have evaluated the performance of resulting estimators using Monte Carlo Simulation. In fact, all the basic choices of biasing parameters centre around the estimated error variance, the estimated coefficient vector using least squares, number and amount of correlation between predictor variables and the sample size. It is also noted that the eigen values of the correlation matrix of the explanatory variables play an important role in determination of the biasing parameters. The review of literature is indicative of the fact that the biasing parameter based on Geometric Mean proposed by Muniz *et al.*[21] worked best.

5. References

- [1] Al-Hassan, Y.: Performance of New Ridge Regression Estimators. *Journal of the Association of Arab Universities for Basic and Applied Science* **9**, 23-26 (2010).
- [2] Alkhamisi, M.A., Shukur, G.: A Monte Carlo Study of Recent Ridge Parameters. *Communication in Statistics – Simulation and Computation* **36**(3), 535-547 (2007). <http://dx.doi.org/10.1080/03610910701208619>.
- [3] Dempster, Arthur P.: Alternatives to Least Squares in Multiple Regression in D. Kabe and R.P. Gupta (eds.), *Multivariate Statistical Inference*. Amsterdam North-Holland Publishing Co. 25-40 (1973).
- [4] Dorugade, A. V.: New Ridge parameters for Ridge Regression, *Journal of the Association of Arab Universities for Basic and Applied Sciences*, 1-6(2013).
- [5] Dorugade, A. V., Kashid, D.N. : Alternative Method for Choosing Ridge Parameter for Regression. *Applied Mathematical Sciences* **4**(9), 447-456 (2010).
- [6] Firinguetti, L: A Generalized Ridge Regression Estimator and its Finite Sample Properties. *Communications in Statistics-Theory and Methods* **28**(5), 1217-1229. (1999). <http://dx.doi.org/10.1080/03610929908832353>
- [7] Frisch, R.: *Statistical Confluence Analysis by Means of Complete Regression Systems*, Institute of Economics, Oslo University, publ. no. 5(1934).
- [8] Gujarati, D.N.: *Basic Econometrics*, McGraw Hill, New York (1995).
- [9] Hocking, R.R., Speed, F.M., Lynn, M.J.: A Class of Biased Estimators in Linear Regression. *Technometrics* **18** (4), 425-437 (1976). <http://dx.doi.org/10.1080/00401706.1976.10489474>
- [10] Hoerl, A.E. and Kennard, R.W.: Ridge Regression Applications to Non-orthogonal Problems. *Technometrics* **12**, 69-82 (1970b). <http://dx.doi.org/10.1080/00401706.1970.10488634>

Dube, M
Isha

- [11] Hoerl, A.E. and Kennard, R.W.: Ridge Regression Biased Estimation for Non-orthogonal Problems. *Technometrics* **12**, 55-67 (1970a).
<http://dx.doi.org/10.1080/00401706.1970.10488634>
- [12] Hoerl, A. and Kennard, R., Baldwin, K.: Ridge Regression: Some Simulation. *Commun. Statist. Theory.Meth.* **4**,105-123 (1975).
<http://dx.doi.org/10.1080/03610927508827232>
- [13] Johnston, J. : Econometrics Methods, McGraw-Hill, Auckland (1987).
- [14] Judge G., W. Griffiths, R. Hill, H. Liitkepohl, and T. Lee.: The Theory and Practice of Econometrics, (2d ed.), New York: Wiley (1985).
<http://dx.doi.org/10.1017/S0266466600011294>
- [15] Khalaf, G., Shukur, G. :Choosing Ridge Parameter for Regression Problem. *Communication in Statistics-Theory and Methods* **34**, 1177-1182 (2005).
<http://dx.doi.org/10.1081/STA-200056836>
- [16] Kibria, B.M.G.: Performance of Some New Ridge Regression Estimators. *Communications in Statistics-Simulation and Computation.* **32**(2), 419-435 (2003). <http://dx.doi.org/10.1081/SAC-120017499>
- [17] Lawless, J.F., Wang, P.: A Simulation Study of Ridge and Other Regression Estimators. *Communications in Statistics- Theory and Methods* **14**, 1589-1604 (1976).
- [18] Maddala, G.S: Introduction to Econometrics, John Wiley & Sons (Asia) Pte. Ltd., Singapore (2005).
- [19] McDonald, G.C., Galarneau, D.I.: A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association* **70** (350), 407-412 (1975). <http://dx.doi.org/10.1080/01621459.1975.10479882>
- [20] Muniz, G., Kibria, B.M.G. :On Some Ridge Regression Estimators –An Empirical Comparison. *Comm. Stat. Simul. Computation*, **38**(3), pp 621-630 (2009).
<http://dx.doi.org/10.1080/03610910802592838>
- [21] Muniz, G., Kibria, B.M.G., Mansson, K., Shukur, G. : On Developing Ridge Regression Parameters: A Graphical Investigation. *SORT.* **36**(2), 115-138 (2012).
- [22] Nomura, Masuo : On the Almost Unbiased Ridge Regression Estimation. *Communications in Statistics-Simulation and Computation* **17** (3), 729-743 (1988). <http://dx.doi.org/10.1080/03610918808812690>
- [23] Sclove, Stanley: Least Squares with Random Regression Coefficient. Technical Report 87, Department of Economics, Stanford University (1973).
- [24] Thisted, Ronald, A. : Ridge Regression, Minimax Estimation and Empirical Bayes Methods. Technical Report 87, Division of Biostatistics, Stanford University (1976).
- [25] Wahba, G., Golub, G.H., Heath, M: Generalized Cross Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* **21** (2), 215-223 (1979).
<http://dx.doi.org/10.1080/00401706.1979.10489751>
-