

Superiority of the MCRR Estimator Over Some Estimators In A Linear Model

Dr.Feras Sh. M. Batah*

ABSTRACT

Modified (r, k) class ridge regression (MCRR) which includes unbiased ridge regression (URR), (r, k) class, principal components regression (PCR) and the ordinary least squares (OLS) estimators is proposed in regression analysis, to overcome the problem of multicollinearity. In this paper, we derive the necessary and sufficient conditions for the superiority of the MCRR estimator over each of these estimators under the Mahalanobis loss function by the average loss criterion. Then, we compare these estimators with each other using the same criterion. Finally, a numerical example is done to illustrate the theoretical results.

Key Words: Multicollinearity; Ordinary Ridge Regression Estimator; (r, k) Class Estimator; MCRR estimator; Mahalanobis loss function.

2000 Mathematics Subject Classification: 62J05; 62J07.

1. INTRODUCTION

According to the Gauss Markov theorem, this is based on the general linear regression model,

$$Y=XB+\epsilon,$$

where Y is an $n \times 1$ vector of responses, X is an $n \times p$ observed matrix of the regressor variables, assumed to have full rank, i.e., $\text{rank}(X) = p$, B is a $p \times 1$ vector of unknown parameters (to be estimated) and ϵ is an $n \times 1$ vector of error terms assumed to be multivariate normally distributed with mean 0 and variance $\sigma^2 I_p$. It is well known that the ordinary least squares (OLS) estimator of B , $\hat{B}_{LS} = (X'X)^{-1} X'Y$, is distributed normal $N(B, \sigma^2 (X'X)^{-1})$. The standard regression model assumes that the column vectors in X are not linearly dependent. In many practical applications, however, engineering in particular, we often find that these column vectors are nearly linearly dependent. We then say that the multicollinearity problem is present. This problem causes the diagonal elements of $(X'X)^{-1}$ to inflate implying that the estimated variance of \hat{B}_{LS} will be large. Multicollinearity could be present if small changes in the design matrix, X , causes the estimated coefficients to vary in sign. (see, e.g., Hoerl and Kennard (1970); Marquardt (1970); Mayer and Willke (1973); Swindel (1976); Batah and Gore (2008, 2009); Batah et al. (2008, 2009)). The OLS estimation is not stable in the existence of multicollinearity. Hence, alternative estimation techniques were designed to eliminate the instability in the estimates which

*Lecturer\ Education Collage for Pure Science\University of Al-anbar, Ramadi, Iraq

results in biased estimators and reduce the variance of the regression coefficients. Consequently, there is considerable interest in various biased estimators of B. The well known of these is the ordinary ridge regression (ORR) estimator of Hoerl and Kennard (1970). The ORR estimator proposed by Hoerl and Kennard (1970), is intended to overcome the problem of multicollinearity by adding a positive value k, usually $0 < k < 1$, to the diagonal elements of the matrix $X'X$. The ORR estimator of B

$$\hat{B}(k) = (X'X + kI_p)^{-1} X'Y, \quad k \geq 0$$

contains the OLS estimator when $k=0$. For increasing k, the ORR estimator approaches 0 which is stable but biased estimator of B. Swindel (1976) illustrated a technique for combining prior information with ridge regression, namely, modified ridge regression (MRR) estimator, extending Hoerl and Kennard's (1970) model as

$$\hat{B}(k,b) = (X'X + kI_p)^{-1} (X'Y + kb), \quad k \geq 0$$

where b being a fixed vector of prior estimate of B. Crouse et al. (1995) illustrated to incorporate prior information in the ORR, namely, the unbiased ridge regression (URR) estimator as follows:

$$\hat{B}(k,J) = (X'X + kI_p)^{-1} (X'Y + kJ), \quad k \geq 0$$

Where J is a random vector with $J \sim N\left(B, \frac{\sigma^2}{k} I\right)$ for $k > 0$. Another possibility for the

removal of information which is responsible for increase of impreciseness in estimation is offered by the principal components regression (PCR) estimator (see Massy (1965), Marquardt (1970) and Gunst and Mason (1977)). For this, let us consider the spectral decomposition of the matrix $X'X$ given as

$$X'X = (T_r, T_{p-r}) \begin{pmatrix} \Lambda_r & 0 \\ 0 & \Lambda_{p-r} \end{pmatrix} \begin{pmatrix} T_r' \\ T_{p-r}' \end{pmatrix}.$$

Here Λ_r and Λ_{p-r} are diagonal matrices such that the main diagonal elements of the $r \times r$ matrix Λ_r are the r largest eigenvalues of $X'X$, while the main diagonal elements of the $(p - r) \times (p - r)$ matrix Λ_{p-r} are the remaining $p - r$ eigenvalues. The $p \times p$ matrix $T = (T_r, T_{p-r})$ is orthogonal with $T_r = (t_1, t_2, \dots, t_r)$ consisting of its first r columns and $T_{p-r} = (t_{r+1}, t_{r+2}, \dots, t_p)$ consisting of the remaining $p - r$ columns of the matrix T. The PCR estimator for B can be written as

$$\hat{B}_{PC} = T_r (T_r' X' X T_r)^{-1} T_r' X' Y = T_r \Lambda_r^{-1} T_r' X' Y.$$

Baye and Parker (1984) proposed the application of ridge methods to improve the PCR estimator, namely (r, k) class estimator as

$$\hat{B}_r(k) = T_r (T_r' X' X T_r + kI_r)^{-1} T_r' X' Y = T_r (\Lambda_r + kI_r)^{-1} T_r' X' Y, \quad k \geq 0.$$

Batah et al.(2009) suggested a new class of estimators for the regression parameter by modifying the URR estimator in the line of the PCR estimator is called the modified (r, k) class ridge regression (MCRR) estimator denoted by :

$$\hat{B}_r(k, J) = T_r (T_r' X' X T_r + k I_r)^{-1} T_r' (X' Y + k J), \quad k \geq 0$$

Where J is a random vector with $J \sim N\left(B, \frac{\sigma^2}{k} I\right)$ for $k > 0$. It may be noted the proposed estimator is obtained as the case of the (r, k) class estimator but with the URR estimator instead of the ORR estimator. The MCRR estimator $\hat{B}_r(k, J)$ has the following properties:

1. $\hat{B}_p(0, J) = \hat{B}_{LS}$ is the OLS estimator.
2. $\hat{B}_p(k, J) = \hat{B}(k, J)$ is the URR estimator.
3. $\hat{B}_r(0, J) = \hat{B}_{PC}$ is the PCR estimator.
4. $\lim_{k \rightarrow 0} \hat{B}_r(k, J) = \hat{B}_{PC}$.
5. $\lim_{k \rightarrow \infty} \hat{B}_r(k, J) = J$.
6. $\hat{B}_r(k, 0) = \hat{B}_r(k)$ is the (r, k) class estimator.

Thus, for suitable choices of the incorporated principal components, r , the ridge parameter, k , and the prior information, J , the MCRR estimator is a general estimator which includes the URR, the PCR, the (r, k) class and the OLS estimators as special cases. It is interesting to the note that the studies on the biased estimators use the mean square error (MSE) criterion, or equivalently the quadratic loss function as a measure of estimators performance. This article extends these studies by choice of the loss function used to decide on a preferred estimator of B . We considered the loss function is Mahalanobis loss function in order to comparing the MCRR estimator the OLS estimator, the PCR estimator and the ORR estimator. Mahalanobis loss function is previously used by Peddada et al. (1989) for comparing generalized ridge regression (GRR) estimator and the OLS estimator. In this article we compare the MCRR estimator to the OLS estimator, the PCR estimator and the ORR estimator by the average loss criterion. The average loss criterion may be defined as follows: Let \hat{B}_1 and \hat{B}_2 be two estimators for a parameter B . The estimator \hat{B}_1 is superior to \hat{B}_2 iff

$$E(L(\hat{B}_1, B)) < E(L(\hat{B}_2, B))$$

where L denotes the loss function. Clearly, when using mean squares as loss function, this reduces to the MSE criterion. For an estimator \hat{B}_1 of B , Mahalanobis loss function is defined as

$$L(\hat{B}_1) = (\hat{B}_1 - B)' \left[\text{Cov}(\hat{B}_1) \right]^{-1} (\hat{B}_1 - B)$$

Where $\text{Cov}(\hat{B}_1)$ is the covariance matrix of the \hat{B}_1 .

In this paper we compare the MCRR estimator to the OLS estimator, the PCR estimator and the ORR estimator. In special case, we get the comparison of the ORR estimator to the OLS estimator by the average loss criterion. We can also get the comparison of the PCR estimator to the OLS estimator and comparison of the PCR estimator to the ORR estimator under the Mahalanobis loss function by the average loss criterion. Finally, we consider a numerical example to justify the superiority of the mentioned estimator.

2. THE PERFORMANCE OF THE MCRR ESTIMATOR UNDER THE AVERAGE LOSS CRITERION

MCRR estimator is biased (Batah et al.(2009)) and the appropriate criterion for gauging the performance of this estimator is the quadratic loss function criterion. We considered the loss function is Mahalanobis loss function in order to comparing the MCRR estimator the OLS, PCR, ORR estimators. Let us denote covariance matrix of the MCRR estimator is $Cov(\hat{B}_r(k, J))$. It is known that the covariance matrix of the MCRR estimator is same to r- k class estimator when $J = 0$ as follows:

$$Cov(\hat{B}_r(k, 0)) = Cov(\hat{B}_r(k)) = \sigma^2 T_r (T_r' X' X T_r + k I_r)^{-1} \Lambda_r (T_r' X' X T_r + k I_r)^{-1} T_r'$$

Now, we can write the Mahalanobis loss function of the MCRR estimator as follows:

$$\begin{aligned} L(\hat{B}_r(k, 0)) &= (\hat{B}_r(k, 0) - B)' [Cov(\hat{B}_r(k, 0))]^{-1} (\hat{B}_r(k, 0) - B) \\ &= \sigma^{-2} Y' X T_r \Lambda_r^{-1} T_r' X' Y - \sigma^{-2} Y' X T_r \Lambda_r^{-1} (T_r' X' X T_r + k I_r) T_r' B \\ &\quad - \sigma^{-2} B' T_r (T_r' X' X T_r + k I_r) \Lambda_r^{-1} T_r' X' Y + \sigma^{-2} B' T_r (T_r' X' X T_r + k I_r) \Lambda_r^{-1} (T_r' X' X T_r + k I_r) T_r' B \end{aligned}$$

Theorem 1: For $k > 0$, the difference $\Delta = L(\hat{B}_{LS}) - L(\hat{B}_r(k, 0))$ is n.d. if and only if the inequality $p - r < k^2 \sigma^{-2} B' T C T' B$, is satisfied.

Proof: We have The difference $\Delta = L(\hat{B}_{LS}) - L(\hat{B}_r(k, 0))$ is given as

$$\begin{aligned} \Delta &= \sigma^{-2} Y' X T_{p-r} \Lambda_{p-r}^{-1} T_{p-r}' X' Y + \sigma^{-2} Y' X (T_r \Lambda_r^{-1} (T_r' X' X T_r + k I_r) T_r' - I_p) B \\ &\quad + \sigma^{-2} B' (T_r \Lambda_r^{-1} (T_r' X' X T_r + k I_r) T_r' - I_p) X' Y \\ &\quad + \sigma^{-2} B' T_r (\Lambda_r^{-1} - (T_r' X' X T_r + k I_r)) \Lambda_r^{-1} (\Lambda_r^{-1} - (T_r' X' X T_r + k I_r)) T_r' B + \sigma^{-2} B' T_{p-r} \Lambda_{p-r}^{-1} T_{p-r}' B \end{aligned}$$

The expectation mean of Δ is equal to $E(\Delta) = p - r - k^2 \sigma^{-2} B' T C T' B$, where

$$C = \begin{pmatrix} \Lambda_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}. \text{ It can be seen that } p - r \text{ is positive. Also, Since } C \text{ is nonnegative}$$

definite matrix. We obtain that the necessary and sufficient condition for the superiority of the OLS estimator to the MCRR estimator under the Mahalanobis loss function as follow:

$$E(\Delta) \leq 0 \Leftrightarrow p - r \leq k^2 \sigma^{-2} B' T C T' B,$$

Then the proof is completed.

Also, we give an estimate for k in order that the OLS estimator cab be superior to the MCRR estimator.

Theorem 2: For $k > 0$, the OLS estimator is better than the MCRR estimator under Mahalanobis loss function iff

$$k \geq \sqrt{\frac{\sigma^2 (p-r)}{B' T C T' B}}, \text{ with } B' T C T' B > 0$$

Similarly, by using same proof of Theorem 1, we can compare between MCRR and other estimators as special cases from MCRR which is maintained above under the same criterion.

Theorem 3: For $k > 0$, the MCRR estimator is better than the ORR estimator under Mahalanobis loss function.

Theorem 4: For $k > 0$, the PCR estimator is better than the MCRR estimator under Mahalanobis loss function.

3. NUMERICAL EXAMPLE

Recently, Batah and Gore (2009) used the data generated by Hoerl and Kennard (1981) to illustrate the comparisons among OLS, URR, ORR, (r, k) class and MCRR estimators by using mean squares error criterion. We now use this data to illustrate the performance of the MCRR estimator when $J = 0$ to other estimators under the Mahalanobis loss criterion. Table 1 shows the $E(L(\hat{B}_r(k,J)))$, $E(L(\hat{B}(k)))$ for various k values. It is obvious that the mean of the Mahalanobis loss function of the ORR and MCRR when $J = 0$ estimators for $k = 0$ are equal to the mean of the Mahalanobis loss function of the OLS estimator and PCR estimator $E(L(\hat{B}_{LS}))=5$, and $E(L(\hat{B}_{PC}))=1$, respectively. The result given in Theorem 1 is obtained, for k values bigger than 0.2736 as $k \geq 0.2736$. Therefore, MCRR estimator is better than the OLS estimator under the Mahalanobis loss criterion as seen from Table 1.

Table 1 Values of $E(L(\hat{B}_r(k,J)))$, $E(L(\hat{B}(k)))$ for various k values.

k	$E(L(\hat{B}_r(k,J)))$	$E(L(\hat{B}(k)))$
0	1.0000	5.0000
0.01	1.0053	5.1122
0.02	1.0214	5.4487
0.03	1.0481	6.0096
0.04	1.0856	6.7949
0.05	1.1337	7.8045
0.06	1.1925	9.0385
0.07	1.2620	10.4968
0.08	1.3423	12.1795
0.09	1.4331	14.0866
0.10	1.5348	16.2180
0.15	2.2032	30.2405
0.20	3.1390	49.8719
0.25	4.3422	75.1124
0.2736	5.0030	88.9744
0.30	5.8128	105.9618
0.40	9.5560	184.4877
0.50	14.3688	285.4495
0.60	20.2510	408.8473
0.70	27.2028	554.6811
0.80	35.2240	722.9508
0.90	44.3148	913.6565
1.00	54.4750	1126.798

4. CONCLUSIONS

In this paper, we have obtained necessary and sufficient conditions for the superiority of the MCRR estimator over the OLS, PCR and ORR estimators under the Mahalanobis loss function, which is a different loss function from quadratic loss function, using the average loss criterion. We have seen that the conditions obtained on the comparisons of the MCRR estimator to the OLS, PCR, ORR estimators depend on the unknown parameters. Also, we have mentioned that if the comparisons depend on the unknown parameters, using unbiased estimates or a priori information about these parameters will lead to the practicable results which is illustrated with a numerical example.

REFERENCES

- Batah, F. and Gore, S. (2008). Improving precision for jackknifed ridge type estimation. *Far East Journal of Theoretical Statistics* 24:157-174.
- Batah, F., Gore, S. and Verma, M. (2008). Effect of jackknifing on various ridge type estimators. *Model Assisted Statistics and Applications* 3: 111-121.
- Batah, F., Gore, S. and Ozkale, M (2009). Combining Unbiased Ridge and Principal Component Regression Estimators, *Communication in Statistics - Theory and Methods* 38: 1 - 9.
- Batah F. and Gore, S., (2009). Ridge regression estimator: combining unbiased and ordinary ridge regression methods of estimation, *Surveys in Mathematics and its Applications* 4: 99 – 109
- Baye, M., and Parker, D. (1984). Combining ridge and principal component regression: a money demand illustration. *Communication in Statistics. Theory and Method* 13:197-205.
- Crouse, R., Jin, C., and Hanumara, R. (1995). Unbiased ridge estimation with prior information and ridge trace. *Communication in Statistics. Theory and Method* 24: 2341-2354.
- Gunst, R. F., and Mason, R. L. (1977). Biased estimation in regression: an evaluation using mean squared error. *Journal of the American Statistical Association* 72:616-628.
- Hoerl, A., and Kennard, R. (1970). Ridge regression : biased estimation for nonorthogonal problems. *Technometrics* 12:55-67.
- Hoerl, A., Kennard, R. (1981). Ridge regression-1980 advances, algorithms, and applications. *American. J. Mathematical Management. Sci.* 1:5-83.
- Marquardt, D. (1970). Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics* 12:591-612.
- Massy, W. F. (1965). Principal component regression in explanatory statistical research. *Journal of the American Statistical Association* 60:234-256.
- Mayer, L. and Willke, T. (1973). On biased estimation in linear models. *Technometrics* 15:497-508.
- Ozkale, M.,R., and Kaciranlar, S. (2007). Comparisons of the unbiased ridge estimation to the other estimations. *Communication in Statistics. Theory and Method* 36:707-723.
- Swindel, B. (1976). Good ridge estimators based on prior information. *Communication in Statistics. Theory and Method* A5:1065-1075.