Comparison of the r- (k,d) class estimator with some estimators for multicollinearity under the Mahalanobis loss function

Shalini Chandra and Nityananda Sarkar®

Banasthali University and Indian Statistical Institute

ABSTRACT

In the case of ill-conditioned design matrix in linear regression model, the r- (k,d) class estimator was proposed, including the ordinary least squares (OLS) estimator, the principal component regression (PCR) estimator, and the two-parameter class estimator. In this paper, we opted to evaluate the performance of the r- (k,d) class estimator in comparison to others under the weighted quadratic loss function where the weights are inverse of the variance-covariance matrix of the estimator, also known as the Mahalanobis loss function using the criterion of average loss. Tests verifying the conditions for superiority of the r- (k,d) class estimator have also been proposed. Finally, a simulation study and also an empirical illustration have been done to study the performance of the tests and hence verify the conditions of dominance of the r- (k,d) class estimator over the others under the Mahalanobis loss function in artificially generated data sets and as well as for a real data. To the best of our knowledge, this study provides stronger evidence of superiority of the r- (k,d) class estimator over the other competing estimators through tests for verifying the conditions of dominance, available in literature on multicollinearity.

Key words: *r- (k,d) class estimator, Principal component estimator, Two-parameter class estimator, Mahalanobis loss function, Risk criterion*

Mathematics Subject Classification: Primary 62J05; Secondary 62J07

1. INTRODUCTION

The presence of multicollinearity among the independent variables in a regression equation poses serious problems in regression analysis. A major consequence of multicollinearity on the OLS estimator is that the estimator becomes unstable. To circumvent this problem, several alternative estimators such as the ordinary ridge regression (ORR) estimator by Hoerl and Kennard (1970), PCR estimator by Massy (1965), the r- k class estimator by Baye and Parker (1984), the two-parameter class estimator by Özkale and Kaçiranlar (2007), have been suggested. Özkale (2012) also proposed the r- (k,d) class estimator by combining the two-parameter class estimator and the PCR estimator.

® Nityananda Sarkar, Indian Statistical Institute, 203 B.T. Road, Kolkata-700108, INDIA, (email: tanya@isical.ac.in).
Shalini Chandra, Banasthali University, Centre for Mathematical Sciences, Rajasthan-304022, INDIA.
This work was carried out when the second author was a visiting scientist at Indian Statistical Institute, North-EastCenter, Tezpur, Assam, India during July2011-June2012. The authors are grateful to Gopal K. Basak, Indian Statistical Institute, Kolkata, for helpful discussions. The authors are thankful to an anonymous referee for his useful comments on the paper. The authors are solely responsible for any remaining errors.

The mean squared error (MSE) criterion, or equivalently the criterion of average loss under quadratic loss function, is a criterion frequently used to evaluate the performance of estimators in linear regression models. The performance of these estimators have been evaluated by the mean squared error criterion, and conditions for superiority of one estimator over some others have also been derived. For instance, Nomura and Okhubo (1985) compared the $r-k$ class estimator with the ORR and OLS estimators in terms of MSE and Sarkar (1992) derived conditions under which the restricted ridge regression estimator is superior to the restricted least squares and the ORR estimators by the same criterion. Özkale (2007) also derived the conditions under which the $r-(k,d)$ class estimator dominates the OLS, PCR and the two-parameter class estimator by the MSE criterion. However, it may be noted that the weighted quadratic loss function is considered to be a stronger criterion than the quadratic loss function and hence it is also widely accepted as a criterion for evaluating the performance of estimators. In particular, if quadratic loss function is weighted by the inverse of the variance-covariance matrix of an estimator, it is called the Mahalanobis loss function. Peddada et al. (1989) showed the inadmissibility of ORR estimator when compared with the OLS estimator under the Mahalanobis loss function. Recently, Üstündağ-Şiray and Sakallioğlu (2012) derived a necessary and sufficient condition for the superiority of the $r-k$ class estimator over the OLS, PCR and ORR estimators under the Mahalanobis loss function using average loss criterion.

The purpose of this paper is to compare the performance of the $r-(k,d)$ class estimator with the OLS, PCR and two-parameter class estimator under the Mahalanobis loss function using average loss criterion, and propose tests for verifying the conditions for dominance. The structure of the paper is as follows: Section 2 describes the model and the estimators. Section 3 discusses the comparison between the $r-(k,d)$ class estimator and the other estimators. Section 4 provide tests for verifying the conditions for dominance. Sections 4 and 5 present the results of a simulation study as well as a numerical illustration for studying the performance of the estimators, respectively. Section 7 offers some concluding remarks.

2. THE MODEL AND THE ESTIMATORS

Let us consider the following regression model:

$$y = X\beta + u \tag{2.1}$$

where y is an $n\times 1$ vector of observations on the variable to be explained, X is an $n\times p$ matrix of n observations on p explanatory variables such that $X'X$ is ill conditioned, β is a $p\times 1$ vector of regression coefficients associated with the explanatory variables and u is an $n\times 1$ vector of disturbances, the elements of which are assumed to have mean zero and variance covariance matrix $\sigma^2 I$.

In order to define the $r-(k,d)$ class estimator, let us consider an orthogonal matrix $T=(t_1,t_2\ldots,t_p)$ such that it diagonalizes $X'X$ i.e., $T'X'XT=\Lambda$, Λ being the diagonal matrix consisting of the eigenvalues of $X'X$ as its diagonal elements. Further, let $T_r=(t_1,t_2,\ldots,t_r)$, where $r\le p$. $T_r'X'XT_r=\Lambda_r=diag(l_1,l_2,\ldots,l_r)$, and $T_{p-r}'X'XT_{p-r}=\Lambda_{p-r}=diag(l_{r+1},l_{r+2},\ldots,l_p)$, where $T_{p-r}=(t_{r+1},t_{r+2},\ldots,t_p)$, and also $X'X=T_r\Lambda_rT_r'+T_{p-r}\Lambda_{p-r}T_{p-r}'$.

The $r-(k,d)$ class estimator as proposed by Özkale (2012) is given as

$$\hat{\beta}_r(k,d) = T_rS_r(k)^{-1}S_r(k,d)\Lambda_r^{-1}T_r'X'y \tag{2.2}$$

where $k\ge 0$, $0\le d<1$. It is worth mentioning that the $r-(k,d)$ class estimator encompasses a larger class of estimators, and, in particular, it includes the OLS, PCR, ORR, two-parameter class and the $r-k$ class estimators for specific values of r, k, and d. Thus,

1. $\hat{\beta}_p(0,0) = \hat{\beta}_p = (X'X)^{-1}X'y$ is the OLS estimator
2. $\hat{\beta}_r(0,0) = \hat{\beta}_r = T_r(T_r'X'XT_r)^{-1}T_r'X'y$ is the PCR estimator,
3. $\hat{\beta}_p(k,0) = \hat{\beta}_p(k) = (X'X + I_p)^{-1}X'y$ is the ORR estimator,
4. $\hat{\beta}_r(k,0) = \hat{\beta}_r(k) = T_r(T_r'X'XT_r + kI_r)^{-1}T_r'X'y$ is the r - k class estimator,
5. $\hat{\beta}_p(k,d) = \hat{\beta}_p(k,d) = (X'X + kI_p)^{-1}(X'y + kd\hat{\beta}_p)$ is the two-parameter class estimator.

3. RISK COMPARISONS

For any estimator $\hat{\beta}$ of β, the Mahalanobis loss function is defined as

$$L_M(\hat{\beta},\beta) = (\hat{\beta}-\beta)'(Cov(\hat{\beta}))^{-1}(\hat{\beta}-\beta) \tag{3.3}$$

Therefore, for the r - (k, d) class estimator, the Mahalanobis loss function is defined as

$$L_M(\hat{\beta}_r(k,d),\beta) = (\hat{\beta}_r(k,d)-\beta)'(Cov(\hat{\beta}_r(k,d)))^{-1}(\hat{\beta}_r(k,d)-\beta) \tag{3.4}$$

From (2.1) and (2.2), we find that

$$E(\hat{\beta}_r(k,d)) = T_r S_r(k)^{-1} S_r(k,d) T_r'\beta \tag{3.5}$$

and

$$Cov(\hat{\beta}_r(k,d) = \sigma^2 T_r S_r(k)^{-1} S_r(k,d)\Lambda_r^{-1} S_r(k,d) S_r(k)^{-1} T_r' \tag{3.6}$$

where $S_r(k,d) = \Lambda_r + kdI_r$ and $S_r(k) = \Lambda_r + kI_r$.

Substituting the expressions of $\hat{\beta}_r(k,d)$ and $Cov(\hat{\beta}_r(k,d))$ from (2.2) and (3.6) in (3.4), and then simplifying, we obtain the following:

$$\begin{aligned} L_M(\hat{\beta}_r(k,d),\beta) = \;& \sigma^{-2} y'XT_r\Lambda_r^{-1}T_r'X'y - 2\beta'T_r S_r(k) S_r(k,d)^{-1} T_r'X'y + \\ & \beta'T_r S_r(k) S_r(k,d)^{-1}\Lambda_r S_r(k,d)^{-1} S_r(k) T_r'\beta \end{aligned} \tag{3.7}$$

Suitable choices of the values of r, k and d in (3.7) give the loss functions of the corresponding estimators under the Mahalanobis loss function. Now, we compare the r - (k, d) class estimator with the others using the risk function defined as the average loss i.e.,viz, $E(L_M(\hat{\beta}_r(k,d),\beta))$ in case of the r - (k, d) class estimator.

3.1. Comparison of the r - (k, d) class estimator with the OLS estimator

We first consider comparing the risk associated with the r - (k, d) class estimator with that of the OLS estimator of β under the Mahalanobis loss function. It can easily be observed that by putting $r = p$ and $k = d = 0$ in (3.7), we get the loss function associated with the OLS estimator. We thus have:

$$\begin{aligned} L_M(\hat{\beta}_p,\beta) - L_M(\hat{\beta}_r(k,d),\beta) = \;& \sigma^{-2} y'XT_{p-r}\Lambda_{p-r}^{-1}T_{p-r}'X'y + \\ & + 2\sigma^{-2} y'X(k(1-d)T_r S_r(k,d)^{-1}T_r' - T_{p-r}T_{p-r}')\beta - \\ & - k(1-d)\sigma^{-2}\beta'T_r S_r(k,d)^{-1}\Lambda_r[k(1+d)I_r + 2\Lambda_r]S_r(k,d)^{-1}T_r'\beta + \\ & + \sigma^{-2}\beta'T_{p-r}\Lambda_{p-r}T_{p-r}'\beta \end{aligned} \tag{3.8}$$

Using (3.8), the difference between the two risk functions is obtained as:

$$\begin{aligned} E\big(L_M(\hat{\beta}_p,\beta) - L_M(\hat{\beta}_r(k,d),\beta)\big) = \;& (p-r) - \\ & - k^2(1-d)^2\sigma^{-2}\beta'T_r S_r(k,d)^{-1}\Lambda_r S_r(k,d)^{-1}T_r'\beta \end{aligned} \tag{3.9}$$

The equation in (3.9) can also be stated as:

$$E\left(L_M(\hat{\beta}_p,\beta)-L_M(\hat{\beta}_r(k,d),\beta)\right)=(p-r)-k^2(1-d)^2\sigma^{-2}\beta'TPT'\beta \qquad (3.10)$$

where

$$P=diag\left(\frac{l_1}{(l_1+kd)^2},\frac{l_2}{(l_2+kd)^2},\ldots,\frac{l_r}{(l_r+kd)^2},0,0,\ldots,0\right) \qquad (3.11)$$

is a positive semi definite matrix for all $k>0$ and $0<d<1$, and $(p-r)$ is positive. Using the equation below (3.12) for the dominance of $r-(k,d)$ class estimator over the OLS estimator, we thus have the following theorem (Theorem 3.1):

$$E\left(L_M(\hat{\beta}_p,\beta)-L_M(\hat{\beta}_r(k,d),\beta)\right)\geq 0 \qquad (3.12)$$

Theorem 3.1. A necessary and sufficient condition for the dominance of the $r-(k,d)$ class estimator over the OLS estimator under the Mahalanobis loss function is given by:

$$\frac{p-r}{k^2(1-d)^2}\geq\frac{\beta'TPT'\beta}{\sigma^2}$$

Thus, the superiority of the $r-(k,d)$ class estimator over the OLS estimator upholds when the values of k and d are such that the given inequality holds. When $d=0$, the theorem reduces to the one obtained by Üstündağ and Sakallioğlu (2012) when the $r-k$ class estimator and the OLS estimator have been compared under the Mahalanobis loss function by the average loss criterion. Moreover, by substituting $r=p$ and $d=0$, the expression in equation (3.10) becomes the same as that obtained by Peddada et al. (1989), when the ORR estimator has been compared with the OLS estimator under the Mahalanobis loss function.

3.2. Comparison of the $r-(k,d)$ class estimator with the OLS estimator

The expression for loss function of $\hat{\beta}_r$, the PCR estimator, can easily be obtained by putting $k=0$ and $d=0$ in (3.7). We thus have,

$$L_M(\hat{\beta}_r,\beta)-L_M(\hat{\beta}_r(k,d),\beta)=\quad 2k(1-d)\sigma^{-2}y'XT_rS_r(k,d)^{-1}T_r'\beta- \qquad (3.13)$$
$$-k(1-d)\sigma^{-2}\beta'T_rS_r(k,d)^{-1}\Lambda_r[k(1+d)I_r+2\Lambda_r]S_r(k,d)^{-1}T_r'\beta$$

Hence, the difference between the risk functions is obtained as

$$E\left(L_M(\hat{\beta}_r,\beta)-L_M(\hat{\beta}_r(k,d),\beta)\right)=-k^2(1-d)^2\sigma^{-2}\beta'T_rS_r(k,d)^{-1}\Lambda_rS_r(k,d)^{-1}T_r'\beta$$

where $S_r(k,d)^{-1}\Lambda_r S_r(k,d)^{-1}$ is a diagonal and positive definite matrix with diagonal elements as

$$\frac{l_1}{(l_1+kd)^2},\frac{l_2}{(l_2+kd)^2},\ldots,\frac{l_r}{(l_r+kd)^2}$$

Thus,

$$E\left(L_M(\hat{\beta}_r,\beta)-L_M(\hat{\beta}_r(k,d),\beta)\right)\leq 0 \qquad (3.14)$$

since $k^2(1-\mathrm{d})^2\sigma^{-2}\beta'T_r\Lambda_rS_r(k,d)^{-2}T_r'\beta$ is nonnegative for all k and d. Hence, in general, the PCR estimator dominates, the $r-(k,d)$ class estimator. However, when:

$$k^2(1-d)^2\sigma^{-2}\beta'T_r\Lambda_rS_r(k,d)^{-2}T_r'\beta=0 \qquad (3.15)$$

these two estimators perform equally well by this criterion. Obviously, the equality in (3.15) holds iff $T_r'\beta=0$ as demonstrated in the following theorem.

Theorem 3.2. The PCR estimator and the $r-(k,d)$ class estimator perform equally well by the risk criterion under the Mahalanobis loss function iff $T_r'\beta=0$.

3.3. Comparison of the r - (k, d) class estimator with the two-parameter class estimator

Finally, we compare the r - (k, d) class estimator and the two-parameter class estimator under the Mahalanobis loss function. The expression for the loss function of $\hat{\beta}(k,d)$, the two-parameter class estimator, can easily be obtained by putting $r = p$ in (3.7). The difference of the two estimators in terms of the Mahalanobis loss function is:

$$L_M(\hat{\beta}(k,d),\beta) - L_M(\hat{\beta}_r(k,d),\beta) = \sigma^{-2} y'XT'_{p-r}\Lambda^{-1}_{p-r}T'_{p-r}X'y - \\ -2\sigma^{-2}y'XT_{p-r}S_{p-r}(k,d)^{-1}S_{p-r}(k)T'_{p-r}\beta + \\ +\sigma^{-2}\beta'T_{p-r}S_{p-r}(k)S_{p-r}(k,d)^{-1}\Lambda_{p-r}S_{p-r}(k,d)^{-1}S_{p-r}(k)T'_{p-r}\beta \quad (3.16)$$

The difference between the two risk functions is given by:

$$E\left(L_M(\hat{\beta}(k,d),\beta) - L_M(\hat{\beta}_r(k,d),\beta)\right) = (p-r) + \\ + k^2(1-d)^2\sigma^{-2}\beta'T_{p-r}S_{p-r}(k,d)^{-1}\Lambda_{p-r}S_{p-r}(k,d)^{-1}T'_{p-r}\beta \quad (3.17)$$

The expression in (3.17) can be restated as:

$$E\left(L_M(\hat{\beta}(k,d),\beta) - L_M(\hat{\beta}_r(k,d),\beta)\right) = (p-r) + k^2(1-d)^2\sigma^{-2}\beta'TQT'\beta$$

where

$$Q = diag\left(0,0,\ldots,0,\frac{l_{r+1}}{(l_{r+1}+kd)^2},\frac{l_{r+2}}{(l_{r+2}+kd)^2},\ldots,\frac{l_p}{(l_p+kd)^2},\right)$$

is a diagonal and positive semi definite matrix for all $k \geq 0$ and $0 < d < 1$. Therefore, since $p - r$ is positive, $(p - r) + k^2(1 - d)^2\sigma^{-2}\beta'TQT'\beta$ cannot be negative: implying that the r - (k, d) class estimator always dominates the two-parameter class estimator under the Mahalanobis loss function stated in the form of the following theorem:

Theorem 3.3. The r - (k, d) class estimator always outperforms the two-parameter class estimator under the Mahalanobis loss function using average loss criterion.

4. TESTS FOR VERIFYING THE CONDITIONS

The conditions stated in Theorems 3.1 and 3.2 in the previous section are not directly verifiable due to unknown parameters β and σ^2. This Section proposes tests in order to be able to infer whether the conditions are satisfied or not for a given dataset. We provide two test statistics for testing the restrictions for the dominance of the r - (k, d) class estimator over the PCR and OLS, by the criterion of average loss under the Mahalanobis loss function. Under the assumption of normality of the disturbance term, it can be easily verified that $\hat{\beta}_r(k,d)$ has a normal distribution with covariance matrix $\sigma^2 T_r S_r(k)^{-1} S_r(k,d)\Lambda_r^{-1} S_r(k,d) S_r(k)^{-1} T_r'$ and mean $T_r S_r(k)^{-1} S_r(k,d) T_r'\beta$.

4.1. The r - (k, d) class versus the PCR estimators

The first null hypothesis considered is H_{01}: $T_r'\beta = 0$. The test statistic for this hypothesis, $T_r'\beta = 0$, was first obtained by Sarkar (1996) in the context of MSE comparisons of the r - k class estimator over the PCR estimator. Along the same lines, below we obtain the test statistic for comparing the r - (k, d) class estimator with the PCR estimator based on average loss under the criterion of the Mahalanobis loss function. Now, under the assumption of normality of the disturbance term, it can be easily verified from (3.5) and (3.6) that $\hat{\beta}_r(k,d)$ has a normal distribution with covariance matrix $\sigma^2 T_r S_r(k)^{-1} S_r(k,d)\Lambda_r^{-1} S_r(k,d) S_r(k)^{-1} T_r'$ and mean $T_r S_r(k)^{-1} S_r(k,d) T_r'\beta$. Hence, that estimator can be defined as:

$$T_r'\hat{\beta}_r(k,d) \sim N(S_r(k)^{-1} S_r(k,d) T_r'\beta, \sigma^2 S_r(k)^{-1} S_r(k,d)\Lambda_r^{-1} S_r(k,d) S_r(k)^{-1})$$

Since, $T_r'\hat{\beta}_r(k,d)$ is an unbiased estimator of $T_r'\beta$ under H_{01}, the test statistic under H_{01} is given by:

$$F_1 = \frac{\hat{\beta}_r(k,d)' T_r S_r(k) S_r(k,d)^{-1} \Lambda_r S_r(k,d)^{-1} S_r(k) T_r' \hat{\beta}_r(k,d)/r}{e'e/(n-p)} \quad (4.18)$$

where $e'e/(n-p)$ is the usual OLS estimator of σ^2, with e as the vector of residuals. F_1 follows a F distribution with d.f. r and $n-p$ under H_{01}.

4.2. The *r* - (*k*, *d*) class and the OLS estimators

For testing the condition stated in Theorem 3.1 for the superiority of the $r-(k,d)$ class estimator over the OLS estimator, we state the null and alternative hypotheses as:

$$H_{02} = \frac{c\beta' TPT'\beta}{\sigma^2} \leq 1 \text{ and } H_{12} = \frac{c\beta' TPT'\beta}{\sigma^2} > 1$$

where $c = k^2(1-d)^2/(p-r)$. Under the assumption of normality for the disturbances, $c^{½}P^{½}T'\hat{\beta}_p$ follows a normal distribution with mean $c^{½}P^{½}T'\beta$ and the covariance matrix $c\sigma^2 P\Lambda^{-1}$. The test statistic is:

$$F_2 = \frac{\hat{\beta}_p' T\Lambda T'\hat{\beta}_p / p}{e'e/(n-p)} \quad (4.19)$$

F_2 follows a non-central F distribution with d.f. p and $n-p$, and the non-centrality parameter is $\lambda = \beta' T\Lambda T'\beta/\sigma^2$ under both the null (H_{02}) and alternative (H_{12}) hypotheses where for H_{02}, $\lambda \leq 1/c$ and for H_{12}, $\lambda > 1/c$.

Hence, depending on the outcomes of these tests for a given data set, conclusions can be drawn regarding the dominance of the $r-(k,d)$ class estimator over the OLS and PCR. If, for instance, H_{01}: $T_r'\beta = 0$ cannot be rejected against the alternative H_{11}: $T_r'\beta \neq 0$ in a given sample, the conclusion is that the PCR estimator and the $r-(k,d)$ class estimator performs equally well by the risk criterion under the Mahalanobis loss function.

In the next two sections, we report the results of a simulation study and a numerical illustration. These were carried out to study the performance of the tests for verifying the conditions for dominance of the $r-(k,d)$ class estimator over the others as obtained in Section 4.

5. A SIMULATION STUDY

We conduct the simulation study for different degrees of collinearity to verify the conditions of dominance of the $r-(k,d)$ class estimator over others given in Section 4. All computations have been done using R codes. Following Newhouse and Oman (1971), we derive the required results, using the explanatory variables generated by the method below:

$$X_{ij} = U + aV_{ij}, \; i = 1, 2, ..., p \text{ and } j = 1, 2, ..., n \quad (5.20)$$

where Vij and U are independent sequence of standard normal pseudo-random numbers. Here, α is so chosen as to result in a desired sample correlation (ρ) among explanatory variables, and this is given by $\alpha = \sqrt{(1-\rho)/\rho}$. For the design matrix X, the normalized eigenvector corresponding to the largest eigenvalue is chosen as a coefficient vector. We generate observations on the dependent variable by:

$$y_{ij} = \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + u_j,\ j = 1, 2, ..., n$$

where u_j's are independent normal random variables with mean zero and variance σ^2. In this study, we have taken $n = 50$; $p = 3$; and $\sigma^2 = 100$. The values of ρ are considered as 0.7, 0.8, 0.9, 0.95. Further, we take r to be the number of eigenvalues greater than unity.

In the simulation study, the tests for the various hypothesis discussed in Section 4, have been carried out for $k = 0.001, 0.01, 0.05, 0.1, 0.5, 0.7, 0.9, 1$ and $d = 0.001, 0.01, 0.1, 0.5, 0.75, 0.9$. We calculate the values of the test statistics for testing the conditions stated in Theorem 3.1 and 3.2 and repeat the process for 1000 times. The proportions of the cases when a relevant null hypothesis is not rejected, are reported in the following tables (Tables 5.1-5.3).

5.1. The r - (k, d) class estimator and the PCR estimator

The value of the test statistic given in (4.18) for testing the null hypothesis H_{01} is calculated and the proportion of the cases, say $P_{1(k,d)}$, when we fail to reject H_{01}, is calculated for the different values of k and d considered here. We note that the test statistic follows F distribution with degrees of freedom 2 and 47. The results are given in Table 5.1.

k/d	**0.001**	**0.01**	**0.10**	**0.50**	**0.75**	**0.90**
0.001	0.909	0.909	0.909	0.894	0.936	0.910
0.01	0.898	0.920	0.905	0.905	0.912	0.907
0.05	0.905	0.936	0.908	0.912	0.900	0.906
0.10	0.922	0.921	0.929	0.909	0.897	0.916
0.50	0.920	0.915	0.920	0.902	0.919	0.895
0.70	0.925	0.904	0.905	0.917	0.920	0.910
0.90	0.924	0.899	0.911	0.910	0.901	0.901
1.00	0.910	0.937	0.911	0.892	0.913	0.904

Table 5.1 Values of $P_{1(k,d)}$ at 5% level of significance for $\rho = 0.80$.

The values of $P_{1(k,d)}$ at 5% level of significance, are high for all values of k and d considered here. In fact, it is clear from the results in Table 5.1 that the actual size of the test is more or less the same as that of the nominal size. Therefore, we conclude that the r - (k, d) class estimator and the PCR estimator perform equally well under the Mahalanobis loss function i.e., the condition stated in Theorem 3.2 holds true for all combinations of chosen k and d at 5% level of significance. In fact, the conclusion remains the same at a 1% level of significance as well.

5.2. The r - (k, d) class and the OLS estimators

We have note in Section (4.2) that for testing the condition for superiority of the r - (k, d) class estimator over the OLS estimator, the test statistic is F_2 and it follows a non-central F distribution with degrees of freedom p and $n - p$ and non-centrality parameter λ, with $\lambda \leq 1/c$ under the null hypothesis H_{02} and $\lambda > 1/c$ under the alternative hypothesis H_{12}. Following Johnson et al. (2004), we can approximate such a non-central distribution, denoted as $F_{p,n-p}(\lambda)$, by $(1 + \lambda/p)F_{p^*,n-p}$, where $F_{p^*,n-p}$ is the central F distribution with degrees of freedom p^* and $n - p$ where $p^* = (p + \lambda)^2/(p + 2\lambda) = p + \lambda^2/(p + 2\lambda)$ is always greater than p and approximated to the nearest integer value.

Since, $dp^*/d\lambda = 2\lambda(p+\lambda)/(p+2\lambda)^2$ is always strictly positive, and hence p^* is an increasing function of λ. Further, writing $F_{p^*,n-p}$ as

$$F_{p^*,n-p} = \frac{Y_1 / p^*}{Y_2 / n-p}$$

where Y_1 and Y_2 are independent central chi square distributions, we have

$$F_{p,n-p}(\lambda) = \frac{Y_1}{Y_2}\frac{(n-p)(p+2\lambda)}{p(p+\lambda)}$$

Now, it is easy to check that $(n-p)(p+2\lambda)/(p(p+\lambda))$ is positive and its derivative with respect to λ is $(n-\mathrm{p})/(p+\lambda)^2$, which is also positive, making $(n-p)(p+2\lambda)/(p(p+\lambda))$ an increasing function of λ. Thus, for a test of size α i.e., if

$$prob_{\lambda=1/c}(F_{p,n-p}(\lambda) > C_r) = prob_{\lambda=1/c}(\frac{Y_1}{Y_2}\frac{(n-p)(p+2\lambda)}{p(p+\lambda)} > C_r) = \alpha$$

then $prob_{\lambda=1/c}(F_{p,n-p}(\lambda) > C_r) \le \alpha$ for all $\lambda \le 1/c$ i.e., under H_{02}, where C_r is the critical value at α% level of significance. Hence, we carry out the test with the value of the non-centrality parameter $\lambda = 1/c$.

	d=0.001				d=0.01			
k	***λ***	***p******	***F**$_{p^*,n-p}$*	**critical value**	***λ***	***p******	***F**$_{p^*,n-p}$*	**critical value**
0.001	999999.00	500001.80	F(∞,47)	1.47	999900.00	499952.30	F(∞,47)	1.47
0.01	9999.90	5002.245	F(∞,47)	1.47	9999.00	5001.75	F(∞,47)	1.47
0.05	400.00	202.253	F(∞,47)	1.47	399.96	202.23	F(∞,47)	1.47
0.10	100.00	52.261	F(52,47)	1.59	99.99	52.26	F(52,47)	1.59
0.50	4.00	4.455	F(4,47)	2.57	4.00	4.45	F(4,47)	2.57
0.70	2.04	3.588	F(4,47)	2.57	2.04	3.59	F(4,47)	2.57
0.90	1.24	3.279	F(3,47)	2.80	1.23	3.28	F(3,47)	2.80
1.00	1.00	3.20	F(3,47)	2.80	1.00	3.20	F(3,47)	2.80
	d=0.10				**d=0.50**			
k	***λ***	***p******	***F**$_{p^*,n-p}$*	**critical value**	***λ***	***p******	***F**$_{p^*,n-p}$*	**critical value**
0.001	990000.00	495002.30	F(∞,47)	1.47	750000.00	375002.30	F(∞,47)	1.47
0.01	9900.00	4952.25	F(∞,47)	1.47	7500.00	3752.25	F(∞,47)	1.47
0.05	396.00	200.25	F(∞,47)	1.47	300.00	152.254	F(∞,47)	1.47
0.10	99.00	51.76	F(52,47)	1.59	75.00	39.765	F(40,47)	1.63
0.50	3.96	4.44	F(4,47)	2.57	3.00	4.000	F(4,47)	2.57
0.70	2.02	3.58	F(4,47)	2.57	1.53	3.387	F(3,47)	2.80
0.90	1.22	3.27	F(3,47)	2.80	0.93	3.177	F(3,47)	2.80
1.00	0.99	3.20	F(3,47)	2.80	0.75	3.125	F(3,47)	2.80
	d=0.75				**d=0.90**			
k	***λ***	***p******	***F**$_{p^*,n-p}$*	**critical value**	***λ***	***p******	***F**$_{p^*,n-p}$*	**critical value**
0.001	437500.00	218752.30	F(∞,47)	1.47	190000.00	95002.25	F(∞,47)	1.47
0.01	4375.00	2189.75	F(∞,47)	1.47	1900.00	952.25	F(∞,47)	1.47
0.05	175.00	89.756	F(90,47)	1.55	76.00	40.26	F(40,47)	1.63
0.10	43.75	24.150	F(24,47)	1.73	19.00	11.805	F(11,47)	2.00
0.50	1.75	3.471	F(4,47)	2.57	0.76	3.13	F(3,47)	2.80
0.70	0.89	3.167	F(3,47)	2.80	0.39	3.04	F(3,47)	2.80
0.90	0.54	3.071	F(3,47)	2.80	0.24	3.02	F(3,47)	2.80
1.00	0.44	3.049	F(3,47)	2.80	0.19	3.01	F(3,47)	2.80

Table 5.2 Values of λ, p^* and $F_{p^*,n-p}$ for selected k and d.

Finally, $c = k(1-d)^2/(p-r)$ is always a positive fraction for our choices of the values of k and d, and hence $1/c$ is always a relatively positive large number. Hence, the loss in terms of size and power of the test due to approximation of the first degree of freedom, p^*, to the nearest integer, is minimal.

The values of the non-centrality parameter, λ, the first degree of freedom of the F distribution, p^*, and the critical values of $F_{p*,n-p}$ distribution at 5% level of significance are provided in Table 5.2. For testing the null hypothesis H_{02}, the values of the statistic $F_2/(1+\lambda/p)$, say F_2^*, is approximated by central $F_{p*,n-p}$. The proportion of number of times when we fail to reject the null hypothesis H_{02}, say $P_{2(k,d)}$, is reported in Table 5.3.

k/d	**0.001**	**0.01**	**0.10**	**0.50**	**0.75**	**0.90**
0.001	1.00	1.00	1.00	1.00	1.00	1.00
0.01	1.00	1.00	1.00	1.00	1.00	1.00
0.05	1.00	1.00	1.00	1.00	1.00	1.00
0.10	1.00	1.00	1.00	1.00	1.00	1.00
0.50	1.00	1.00	1.00	0.997	0.998	0.978
0.70	0.994	0.994	0.994	0.993	0.984	0.970
0.90	0.989	0.989	0.988	0.985	0.973	0.964
1.00	0.985	0.985	0.985	0.978	0.971	0.959

Table 5.3 Values of $P_{2(k,d)}$ at 5% level of significance for $\rho = 0.80$.

From Table 5.3, it can be clearly seen that the values of $P_{2(k,d)}$ at 5% level of significance are either 1 or close to 1 for all values of k and d considered here. The results render that the r-(k,d) class estimator a more suitable estimator than the OLS estimator for large proportions of replications. Subsequently, we may conclude that the r-(k,d) class estimator outperforms the OLS estimator under the Mahalanobis loss function. The same also holds for 1% level of significance. The test results do not show much variation when carried out for $\rho = 0.7, 0.8, 0.9, 0.95$. Therefore, in the interest of brevity, we report results only for the case when $\rho = 0.8$.

6. A NUMERICAL ILLUSTRATION

In this section, we provide an example with a real data set is provided to demonstrate the performance of the test statistics obtained in Section 4, along with the evaluation of the r-(k,d) class estimator as compared to the others. The data set is taken from Hald (1952), in which the response variable y represents the heat evolved in a cement mix measured in calorie/gm, and four explanatory variables, which are ingredients of the mix, viz, X_1: tricalcium aluminate, X_2: tricalcium silicate, X_3: tetracalcium alumino ferrite and X_4: dicalcium silicate. These four variables are measured as percentage weights in clinkers of their respective chemical compounds. This data set has been widely used to illustrate collinearity and variable selection (see, for instance, Draper and Smith, 1981; Montgomery and Peck, 1982; Piepel and Redgate 1998; Özkale, 2012). Here, we use this data set to compute the test statistics in order to verify if the conditions for the superiority of the r-(k,d) class estimator over the others stated in Theorems 3.1 and 3.2 holds for this data set. The data set is given in Table 6.4:

Based on these 13 observations the variables X_1 and X_3, and X_2 and X_4 are highly correlated, with the correlation coefficients of -0.824 and -0.975, respectively. The condition index number of the matrix $X'X$ is found to be 20.585, which shows that the data set has moderate collinearity of structure. For this study, we have chosen r to be 2. The values of the test statistic for dominance conditions stated in Theorem 3.1 and 3.2 are given below for $k = 0.001, 0.01, 0.02, 0.03, 0.07, 0.1, 0.5$ and $d = 0.001, 0.01, 0.1, 0.5, 0.75, 0.9$.

X_1	X_2	X_3	X_4	y
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

Table 6.4 Data set for the numerical illustration

	d=0.001			d=0.01		
k	λ	p^*	F_2^*	λ	p^*	F_2^*
0.001	2000002.000	1000006	**0.010**	2000200.000	1000102	**0.010**
0.01	20000.020	10002	**1.035**	20002.000	10003	**1.035**
0.02	5000.005	2502	4.138	5000.500	2503	4.137
0.03	2222.224	1113	9.301	2222.444	1113	9.300
0.07	408.164	206	50.237	408.204	206	50.232
0.10	200.000	102	101.499	200.020	102	101.490
0.50	8.000	6	1725.490	8.001	6	1725.376
	d=0.10			**d=0.50**		
k	λ	p^*	F_2^*	λ	p^*	F_2^*
0.001	2020202.000	1010103	**0.01025**	2666667	1333336	**0.00776**
0.01	20202.020	10103	**1.025**	26667	13335	**0.078**
0.02	5050.505	2528	4.097	6667	3336	3.104
0.03	2244.669	1125	9.208	2963	1484	6.979
0.07	412.286	208	49.740	544	274	37.769
0.10	202.020	103	100.504	267	136	76.500
0.50	8.081	6	1713.949	11	8	1411.765
	d=0.75			**d=0.90**		
k	λ	p^*	F_2^*	λ	p^*	F_2^*
0.001	4571429.00	2285717	**0.00452**	10526320.00	5263160	**0.00196**
0.01	45714.29	22860	**0.05**	105263.20	52634	**0.02**
0.02	11428.57	5717	**1.81**	26315.79	13160	**0.08**
0.03	5079.37	2542	4.07	11695.91	5850	**1.77**
0.07	932.94	469	22.1	2148.23	1076	9.62
0.10	457.14	231	44.9	1052.63	529	19.6
0.50	18.29	11	929.11	42.11	23	449.1

Table 6.5 Values of λ, p^* and F_2^* for the given dataset

6.1. The r - (k, d) class estimator and the PCR estimator

The value of the statistic in (4.18) for testing H_{01} comes out to be around 10264.27, for all the chosen values of k and d, thereby showing hardly any variation in the values of the test statistic for the values of k and d considered here. The values of the statistic are evidently very high when compared with the critical value of $F(2,9)$ at a 5% level of significance suggesting that the null hypothesis H_{01}: $T_r'\beta = 0$ is rejected at a 5% level of significance. Therefore, it can be concluded that the r - (k, d) class estimator and the PCR estimator do not perform equally well under the Mahalanobis loss function for this data set.

6.2. The r - (k, d) class and the OLS estimators

A statistic in (4.19) follows a non-central F distribution with degrees of freedom (4, 9) and the non-centrality parameter λ. The values of the non-centrality parameter λ, the first degree of freedom of F distribution, p^*, and the values of the statistic F_2^* for chosen values of k and d are given in Table 6.5.

The values of the F_2^* for $k = 0.001, 0.01$ and for all values of d considered here, and also for $k = 0.02$; $d = 0.75, 0.9$, and $k = 0.03$; $d = 0.9$, implies that the null hypothesis is not rejected at 5% level of significance when compared with the critical value 2.71 of $F_{0.05}(\infty, 9)$. For $k = 0.02, 0.03, 0.07, 0.1, 0.5$ and for all the chosen values of d, except for the cases mentioned above, the values of the statistic are quite high, suggesting the rejection of H_{02}. For this data set, the r - (k, d) class estimator outperforms the OLS estimator for $k = 0.001, 0.01$ and for all the chosen values of d, as well as for $k = 0.02$; $d = 0.75, 0.9$, and $k = 0.03$; $d = 0.9$ under the Mahalanobis loss function.

7. CONCLUDING REMARKS

In this paper, we compare the performance of the r - (k, d) class estimator with the OLS, PCR and the two-parameter class estimator using the Mahalanobis loss function as the risk criterion. The conditions for the dominance of the r - (k, d) class estimator over the OLS and PCR estimators have been derived, and tests for verifying these conditions have also been suggested. A numerical illustration and a simulation study have been conducted to study the performance of the tests to evince the superiority of the r - (k, d) class estimator over the others in artificially generated data sets as well as for a real data set.

REFERENCES

Baye, M.R. and D.F. Parker (1984). Combining ridge and principal components regression: a money demand illustration. *Communications in Statistics-Theory and Methods*, 13 (2), 197-205.

Draper, N.R. and A. Smith (1981). *Applied Regression Analysis*. (II edition) New York: Wiley.

Hald, A. (1952). *Statistical Theory with Engineering Applications*. New York: Wiley, 647.

Hoerl, A.E. and R.W. Kennard (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12, 55-67.

Johnson, N.L., S. Kotz and N. Balakrishnan (2004). *Continuous Univariate Distributions*. Vol 2 (II edition) New York: Wiley.

Massy, M.F. (1965). Principal component regression in explanatory research. *Journal of the American Statistical Association*, 60, 234-266.

Montgomery, D.C. and E.A. Peck (1982). *Introduction to linear regression analysis*. New York: Wiley.

Newhouse, J.P. and S.D. Oman (1971). An evaluation of ridge estimators. *Rand corporation*, 1-29.

Nomura, M. and T. Okhuba (1985). A note on combining ridge and principal component regression. *Communications in Statistics-Theory and Methods*, 14, 489-2493.

Özkale, M.R. and S. Kaçiranlar (2007). The restricted and unrestricted two parameter estimators. *Communications in Statistics-Theory and Methods*i 36 (15), 2707-2725.

Özkale M.R. (2012). Combining the unrestricted estimators into a single estimator and a simulation study on the unrestricted estimators. *Journal of Statistical Computation and Simulation*, 62 (4), 653-688.

Peddada, S.D., A.K. Nigam and A.K. Saxena (1989). On the inadmissibility of ridge estimator in a linear model. *Communications in Statistics-Theory and Methods*, 18 (10), 3571-3585.

Piepel, G. and T. Redgate (1998). A mixture experiment analysis with Hald Cement data. *Journal of the American Statistical Association*, 52 (1), 23-30.

Sarkar, N. (1992). A new estimator combining the ridge regression and the restricted least squares methods of estimation. *Communications in Statistics-Theory and Methods*, 21, 1987-2000.

Sarkar, N. (1996). Mean square error matrix comparison of some estimators in linear regression with multicollinearity. *Statistics and Probability Letters*, 30, 133-138.

Üstündağ-Şiray, G. and S. Sakallioğlu (2012). Superiority of the r-k class estimator over some estimators in a linear model. *Communications in Statistics-Theory and Methods*, 41 (15), 2819-2832.