ON THE PERFORMANCE OF SOME BIASED ESTIMATORS IN A MISSPECIFIED MODEL WITH CORRELATED REGRESSORS

Shalini Chandra[1], Gargi Tyagi[2]

ABSTRACT

In this paper, the effect of misspecification due to omission of relevant variables on the dominance of the $r - (k, d)$ class estimator proposed by Özkale (2012), over the ordinary least squares (OLS) estimator and some other competing estimators when some of the regressors in the linear regression model are correlated, have been studied with respect to the mean squared error criterion. A simulation study and numerical example have been demostrated to compare the performance of the estimators for some selected values of the parameters involved.

Key words: omission of relevant variables, multicollinearity, $r - (k, d)$ class estimator, mean squared error.

1. Introduction

In multiple linear regression, the presence of multicollinearity inflates sampling variance of the ordinary least squares estimator and may also produce wrong signs of the estimator. Many authors have witnessed the presence of multicollinearity in the various fields of application, including Hamilton (1972), Mahajan *et al.* (1977), Heikkila (1988), Graham (2003), among others. To cope up with the problem of multicollinearity several alternative methods to the OLS have been proposed, *viz.* ordinary ridge regression (ORR) by Hoerl and Kennard (1970); principal component regression (PCR) by Massy (1965). In the hope that combining two estimators will contain the properties of both gave rise to the development of the $r - k$ class, the two-parameter class and the $r - (k, d)$ class estimators (see Baye and Parker (1984); Kaciranlar and Sakallioglu (2001); Özkale and Kaciranlar (2007) and Özkale (2012)).

---

[1] Department of Mathematics & Statistics, Banasthali Vidyapith, Banasthali - 304022 India.
  E-mail: chandrshalini@gmail.com.
[2] Department of Mathematics & Statistics, Banasthali Vidyapith, Banasthali - 304022 India.
  E-mail: tyagi.gargi@gmail.com.

The performance of these estimators have been evaluated under various comparison criteria like mean squared error (MSE), matrix MSE, Pitman's closeness criterion and the Mahalanobis loss function. Nomura and Ohkubo (1985) derived the dominance conditions of the $r - k$ class estimator over the OLS and ORR estimators and Sarkar (1996) obtained conditions of the superiority of the $r - k$ class estimator over the other estimators under matrix MSE criterion. Özkale and Kaciranlar (2008) compared the $r - k$ class estimator with the OLS estimator under Pitman's closeness criterion. Özkale (2012) proposed the $r - (k, d)$ class estimator and compared this estimator with the other biased estimators under the MSE criterion. Sarkar and Chandra (2015) studied the performance of the $r - (k, d)$ class estimator over the OLS, PCR and the two-parameter class estimator under the Mahalanobis loss function and derived tests to verify the conditions.

In these studies, it has been assumed inherently that the model is correctly specified. However, in practice, some of the relevant regressors may get excluded from the model, i.e. the model does not remain correctly specified, known as misspecified model. The omission of relevant regressors causes biased and inconsistent estimation. The effect of the omission of relevant regressors on the performance of the estimators have been studied by several authors, for example, Kadiyala (1986); Trenkler and Wijekoon (1989) and Wijekoon and Trenkler (1989). Although not much work has been done when some of the regressors are omitted and multicollinearity is also present, Sarkar (1989) studied the performance of the $r - k$ class estimator and compared it with the OLS, ORR and PCR estimators under MSE criterion when the model is misspecified due to omission of relevant regressors.

In this paper, misspecification due to omission of relevant regressors and multicollinearity have been studied simultaneously and the effect of misspecification on the dominance of the $r - (k, d)$ class estimator over the other biased estimators has been studied under the MSE criterion. The plan of this paper is as follows: in Section 2, the model and the estimators under study are given. Section 3 provides the comparison of the estimators and a Monte Carlo simulation has been given in Section 4. A numerical example is given in Section 5 to see the effect of misspecification on the estimators, which in turn exhibits the utility of the estimators. The paper is concluded in Section 6.

2. Model structure and the estimators

Let us consider the regression model as:

$$y = X\beta + Z\gamma + \varepsilon, \tag{2.1}$$

where y is an $n \times 1$ vector of dependent variable, X and Z are $n \times p$ and $n \times q$ full column rank matrices of regressors respectively such that $X'X$ and $Z'Z$ are ill-conditioned, $p + q < n$, β and γ are the corresponding $p \times 1$ and $q \times 1$ vectors of

parameters associated with X and Z, respectively. ε is an $n \times 1$ vector of disturbance term, and it is assumed that $\varepsilon \sim N(0, \sigma^2 I_n)$. Suppose that an investigator has unknowingly excluded regressors of Z matrix, thus the misspecified model is given by:

$$y = X\beta + u, \tag{2.2}$$

where $u = Z\gamma + \varepsilon$. Misspecification occurs when the investigator assumes the disturbance vector u to be normally distributed with mean vector 0 and variance $\sigma^2 I_n$.

Let us consider the following transformation for the model in (2.2):

$$y = XTT'\beta + u = X^*\alpha + u, \tag{2.3}$$

where $X^* = XT$, $T'\beta = \alpha$, $T = (t_1, t_2, \dots, t_p)$ is a $p \times p$ orthogonal matrix with $T'X'XT = \Lambda$ and $\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_p)$ is a $p \times p$ diagonal matrix of eigen values of $X'X$ matrix such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Now, let $T_r = (t_1, t_2, \dots, t_r)$ be $p \times r$ orthogonal matrix after deleting last $p - r$ columns from T matrix, where $r \leq p$. Thus, $T_r'X'XT_r = \Lambda_r$ where $\Lambda_r = diag(\lambda_1, \lambda_2, \dots, \lambda_r)$ and $T_{p-r}'X'XT_{p-r} = \Lambda_{p-r}$, where $\Lambda_{p-r} = diag(\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_p)$. Also, $T'T = T_r'T_r + T_{p-r}'T_{p-r}$ and let $N = \{1, 2, \dots, r; r + 1, \dots, p\}$ be a set of first p integers such that $N = \{N_r; N_{p-r}\}$ where $N_r = \{1, 2, \dots, r\}$ and $N_{p-r} = \{r + 1, r + 2, \dots, p\}$.

Özkale (2012) introduced an estimator by grafting the two-parameter class estimator and the PCR estimator together, known as the $r - (k, d)$ class estimator to deal with the problem of multicollinearity. For the misspecified model in (2.2) the $r - (k, d)$ class estimator is given by:

$$\hat{\beta}_r(k, d) = T_r(T_r'X'XT_r + kI)^{-1}(T_r'X'y + kdT_r'\hat{\beta}_r) \quad k \geq 0, \ 0 < d < 1 \tag{2.4}$$

which can be rewritten as:

$$\hat{\beta}_r(k, d) = T_r S_r(k)^{-1} \Lambda_r^{-1} S_r(kd) T_r'X'y \, k \geq 0, \ 0 < d < 1, \tag{2.5}$$

where $S_r(k) = \Lambda_r + kI_r$ and $S_r(kd) = \Lambda_r + kdI_r$. This is a general estimator which includes the OLS, ORR, PCR, $r - k$ class and the two-parameter class estimators as its special cases as:

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

2.1. Properties of the estimator

From (2.5), the bias and the variance of $\hat{\beta}_r(k,d)$ can be obtained as:

$$\text{Bias}(\hat{\beta}_r(k,d)) = \left(k(d-1)T_r S_r(k)^{-1}T_r' - T_{p-r}T_{p-r}'\right)\beta + T_r S_r(k)^{-1}\Lambda_r^{-1}S_r(kd)T_r'\delta \qquad (2.6)$$

where $\delta = X'Z\gamma$, and

$$\text{Var}(\hat{\beta}_r(k,d)) = \sigma^2 T_r S_r(k)^{-2}\Lambda_r^{-1}S_r(kd)^2 T_r' \qquad (2.7)$$

respectively.

It is clear from (2.6) and (2.7) that the bias of the $r - (k,d)$ class estimator increases due to omission of relevant regressors whereas the variance of the estimator is not affected by the misspecification.

Further, the MSE for an estimator $\tilde{\beta}$ of β is defined as:

$$MSE(\tilde{\beta}) = E(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta) = tr(\text{Var}(\tilde{\beta})) + \left[\text{Bias}(\tilde{\beta})\right]'\left[\text{Bias}(\tilde{\beta})\right] \qquad (2.8)$$

By substituting (2.6) and (2.7) in (2.8) and on simplification, we get:

$$MSE(\hat{\beta}_r(k,d)) = \sigma^2 tr[S_r(k)^{-1}S_r(kd)\Lambda_r^{-1}S_r(kd)S_r(k)^{-1}]$$
$$+\beta'(k(1-d)T_r S_r(k)^{-1}T_r' + T_{p-r}T_{p-r}')(k(1-d)\,T_r S_r(k)^{-1}T_r'$$
$$+ T_{p-r}T_{p-r}')\beta$$
$$-2\beta'(k(1-d)T_r S_r(k)^{-1}T_r' + T_{p-r}T_{p-r}')T_r S_r(k)^{-1}\Lambda_r^{-1}S_r(kd)T_r'\delta$$
$$+\delta'T_r S_r(k)^{-1}\Lambda_r^{-1}S_r(kd)S_r(k)^{-1}\Lambda_r^{-1}S_r(kd)T_r'\delta \qquad (2.9)$$

which can be rewritten as:

$$MSE(\hat{\beta}_r(k,d)) = \underbrace{\sum_{i=1}^{r}\frac{\sigma^2(\lambda_i+kd)^2+k^2(d-1)^2\lambda_i\alpha_i^2}{\lambda_i(\lambda_i+k)^2} + \sum_{i=r+1}^{p}\alpha_i^2}$$

$$+\underbrace{\sum_{i=1}^{r}\frac{(\lambda_i+kd)^2\eta_i^2-2k(1-d)\lambda_i(\lambda_i+kd)\alpha_i\eta_i}{\lambda_i^2(\lambda_i+k)^2}} \qquad (2.10)$$

where $T'\delta = \eta = \{\eta_1, \eta_2, \ldots, \eta_p\}$. Following Özkale(2012), the first under-bracket is the MSE obtained when there is no misspecification and the second under-bracket is the contribution of omission of relevant regressors.

The MSE of other estimators can be obtained by substituting the suitable values of r, k and d in (2.10). From the risk expression in (2.10), it can be seen that the effect of omission of relevant regressors on the MSE values will depend on the sign of the second term. If $\alpha_i\eta_i$ is negative for all values of $i \in N_r$, the second term in (2.10) will be negative and thus the MSE of the $r - (k,d)$ class

estimator will increase due to omission of relevant regressors. However, if $\alpha_i \eta_i$ is non-negative for some values of $i \in N_r$ no definite conclusion can be made regarding the effect of misspecification.

2.2. Optimum values of k and d

The selection of the unknown biasing parameters k and d in the $r - (k, d)$ class estimator is an important problem. The optimum values of k and d in $r - (k, d)$ class estimator can be obtained by minimizing the MSE of the estimator with respect to k and d. To find a pair (k, d) of optimum values of k and d, we will use the technique of maxima and minima in calculus.

Let the two-dimensional function $MSE(\hat{\beta}_r(k, d))$ have its minimum value at (k_0, d_0) and have a continuous partial derivative at this point, then $\frac{\partial MSE(\hat{\beta}_r(k_0,d_0))}{\partial k} = 0$ and $\frac{\partial MSE(\hat{\beta}_r(k_0,d_0))}{\partial d} = 0$. The points k_0 and d_0 can be found as follows:

On differentiating $MSE(\hat{\beta}_r(k, d))$ in (2.10) with respect to d keeping r and k fixed, we obtain

$$\frac{\partial MSE(\hat{\beta}_r(k,d))}{\partial d} = 2k \sum_{i=1}^{r} \frac{\sigma^2(\lambda_i + kd) - k(1-d)\lambda_i\alpha_i^2 + (\lambda_i + 2kd - k)\alpha_i\eta_i + (\lambda_i + kd)\eta_i^2/\lambda_i}{\lambda_i(\lambda_i + k)^2} \quad (2.11)$$

and equating (2.11) to zero, we get:

$$d_0 = \frac{\sum_{i=1}^{r} \frac{k\alpha_i^2 - \sigma^2}{(\lambda_i + k)^2} - \sum_{i=1}^{r} \frac{(\lambda_i - k)\alpha_i\eta_i + \eta_i^2}{\lambda_i(\lambda_i + k)^2}}{\sum_{i=1}^{r} \frac{k(\sigma^2 + \lambda_i\alpha_i^2)}{\lambda_i(\lambda_i + k)^2} + \sum_{i=1}^{r} \frac{k(2\alpha_i\eta_i + \eta_i^2/\lambda_i)}{\lambda_i(\lambda_i + k)^2}}. \quad (2.12)$$

Assuming that $\alpha_i\eta_i > 0$ for all $i \in N_r$, if $k\alpha_i^2 - \sigma^2 > 0$ and $\lambda_i > k$ for all $i \in N_r$, the upper bound of d_0 is given by

$$\frac{\sum_{i=1}^{r}(k\alpha_i^2 - \sigma^2)/(\lambda_i + k)^2}{\sum_{i=1}^{r} k(\sigma^2 + \lambda_i\alpha_i^2)/\lambda_i(\lambda_i + k)^2} \quad (2.13)$$

which is the optimum value of d when there is no misspecification due to omission of relevant regressors. Thus, if $k\alpha_i^2 - \sigma^2 > 0$ and $\lambda_i > k$ for all $i \in N_r$, the optimum value of d in the misspecified model is less than that in the case of no misspecification. Moreover, for d_0 to be a positive value $\lambda_i(k\alpha_i^2 - \sigma^2) - (\lambda_i - k)\alpha_i\eta_i + \eta_i^2$ should be positive for $i \in N_r$.

Further, differentiating (2.10) with respect to k keeping r and d fixed, we obtain:

$$\frac{\partial MSE(\hat{\beta}_r(k,d))}{\partial k} = -2(1-d) \sum_{i=1}^{r} \frac{\sigma^2(\lambda_i + kd) - k(1-d)\lambda_i\alpha_i^2}{(\lambda_i + k)^3}$$
$$-2(1-d) \sum_{i=1}^{r} \frac{(\lambda_i - k + 2kd)\alpha_i\eta_i + (\lambda_i + kd)\eta_i^2/\lambda_i}{(\lambda_i + k)^3} \quad (2.14)$$

From (2.14) and (2.12), we have:

$$\frac{\partial MSE(\hat{\beta}_r(k,d_0))}{\partial k} = \frac{-2\sum_{i=1}^{r}(\sigma^2+\alpha_i\eta_i+\eta_i^2/\lambda_i)/(\lambda_i(\lambda_i+k))}{k\left[\sum_{i=1}^{r}(\sigma^2+\lambda_i\alpha_i^2+2\alpha_i\eta_i+\eta_i^2/\lambda_i)/(\lambda_i(\lambda_i+k)^2)\right]^2}$$

$$\times\left[\sum_{i=1}^{r}\frac{(\sigma^2+\lambda_i\alpha_i^2+2\alpha_i\eta_i+\eta_i^2/\lambda_i)}{\lambda_i(\lambda_i+k)^2}\sum_{i=1}^{r}\frac{\lambda_i(\sigma^2+\alpha_i\eta_i+\eta_i^2/\lambda_i)-k(\lambda_i\alpha_i^2+\alpha_i\eta_i)}{(\lambda_i+k)^3}\right.$$

$$\left.+\sum_{i=1}^{r}\frac{k(\lambda_i\alpha_i^2+\alpha_i\eta_i)-\lambda_i(\sigma^2+\alpha_i\eta_i+\eta_i^2/\lambda_i)}{\lambda_i(\lambda_i+k)^2}\sum_{i=1}^{r}\frac{\sigma^2+\lambda_i\alpha_i^2+2\alpha_i\eta_i+\eta_i^2/\lambda_i}{(\lambda_i+k)^3}\right]. \quad (2.15)$$

Clearly, $\frac{\partial MSE(\hat{\beta}_r(k,d_0))}{\partial k}$ is zero, when

$$k_0 = \frac{\sigma^2+\alpha_i\eta_i+\eta_i^2/\lambda_i}{\alpha_i^2+\alpha_i\eta_i/\lambda_i}, \quad for\, i = 1,2,\dots,r. \quad (2.16)$$

Then (k_0, d_0) is the expected point which minimizes $MSE(\hat{\beta}_r(k,d))$ where k_0 and d_0 are given as (2.16) and (2.12) respectively. However, when we substitute $k = k_0$ in (2.12) d_0 becomes zero. Therefore, a point (k_0, d_0) which satisfies $k > 0$, $0 < d < 1$ and minimizes $MSE(\hat{\beta}_r(k,d))$ cannot be found (see Fig.1). In order to find an appropriate value of k and d, the behaviour of the MSE of the estimator at boundary points can be studied. This conclusion has been illustrated through the graph reported below:
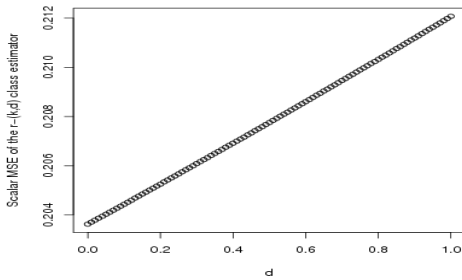
(a) In the case of no misspecification when $k = 0.1$
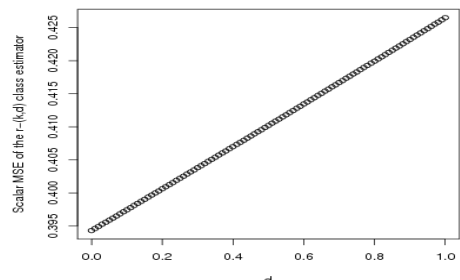
(b) In the case of misspecification when $k = 0.1$
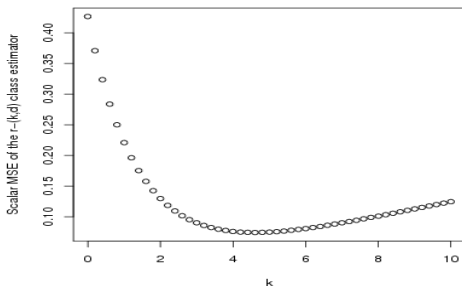
(c) In the case of no misspecification when $d = 0.1$

(d) In the case of misspecification when $d = 0.1$

(e) In the case of no misspecification when
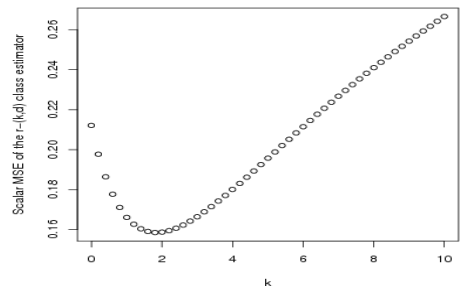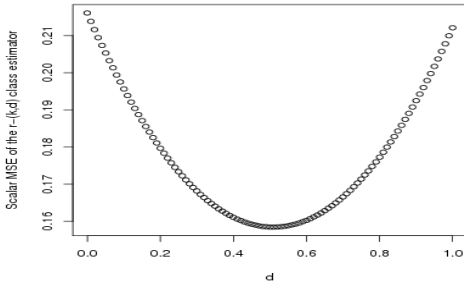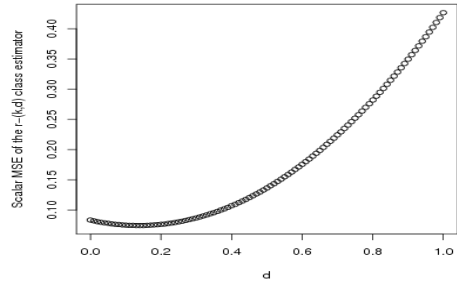$k = 5$

(f) In the case of misspecification when
$k = 5$
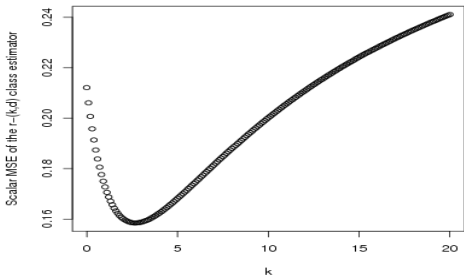
(g) In the case of no misspecification when
$d = 0.3$
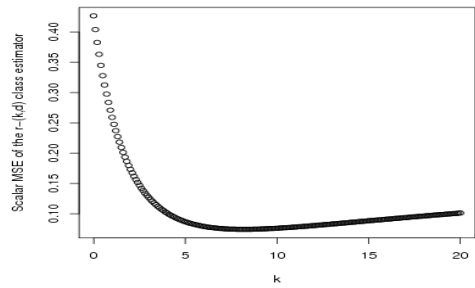
(h) In the case of misspecification when
$d = 0.3$

Figure 1. MSE of the $r - (k, d)$ class estimator for the true and misspecified model

From Figure 1 the effect of misspecification on the optimum values of k and d for fixed values of d and k respectively can be observed, and, also the pair of values of k and d may not be found out for which the $r - (k, d)$ class estimator has minimum MSE. Further, we note that for the fixed values of d, the MSE of $\hat{\beta}_r(k, d)$ takes the minimum value for smaller value of k in the misspecified model when compared with the true model. However, for small value of k (see Fig. (a) and Fig. (b)), no variations are observed in the MSE values of $\hat{\beta}_r(k, d)$ for both the models, whereas for $k = 5$, the MSE of $\hat{\beta}_r(k, d)$ takes the minimum value for a smaller value of d in the misspecified model.

3. Comparison of the estimators under mse criterion

In this section, we compare the $r - (k, d)$ class estimator with other biased estimators when the model is misspecified due to omission of relevant regressors, and also study the effect of misspecification on the dominance conditions.

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

The MSE of the OLS estimator in the misspecified model can be obtained by substituting $r = p, k = 0$ in (2.10), as:

$$MSE(\hat{\beta}) = \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i} + \sum_{i=1}^{p} \frac{\eta_i^2}{\lambda_i^2}. \tag{3.1}$$

The difference of MSEs of the $r - (k, d)$ class estimator and the OLS estimator, say Δ_1, can be written as:

$$\Delta_1 = MSE(\hat{\beta}) - MSE(\hat{\beta}_r(k, d))$$
$$= \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i} - \sum_{i=1}^{r} \frac{\sigma^2(\lambda_i + kd)^2 + k^2(1-d)^2 \lambda_i \alpha_i^2}{\lambda_i(\lambda_i + k)^2} - \sum_{i=r+1}^{p} \alpha_i^2$$
$$+ \sum_{i=1}^{p} \frac{\eta_i^2}{\lambda_i^2} - \sum_{i=1}^{r} \frac{-2k(1-d)(\lambda_i + kd)\alpha_i \eta_i + (\lambda_i + kd)^2 \eta_i^2 / \lambda_i}{\lambda_i(\lambda_i + k)^2}. \tag{3.2}$$

On further simplification, the difference can be rewritten as:

$$\Delta_1 = k(1-d) \sum_{i=1}^{r} \frac{[2\lambda_i(\sigma^2 + \alpha_i \eta_i + \eta_i^2 / \lambda_i) + k((\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2 / \lambda_i) + d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2 / \lambda_i)]}{\lambda_i(\lambda_i + k)^2}$$
$$+ \sum_{i=r+1}^{p} \frac{\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2 / \lambda_i}{\lambda_i}. \tag{3.3}$$

It is clear from the above expression that $\Delta_1 \geq 0$ that is, the $r - (k, d)$ class estimator dominates the OLS estimator, for all $k > 0$, $0 < d < 1$ if $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2 / \lambda_i \geq 0$, for all $i \in N$. From (3.3), it can also be observed that when there is no misspecification due to omission of relevant regressors (i.e. $\eta_i = 0$ for all $i \in N$) the condition reduces to $\sigma^2 - \lambda_i \alpha_i^2 \geq 0$ for all $i \in N$, which is the same as that of obtained by Ozkale (2012)). It is evident that due to addition of a positive term η_i^2 / λ_i the odds for $\Delta_1 \geq 0$ are higher in the misspecified model.

Further, if $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2 / \lambda_i < 0$, $i \in N_r$ and $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2 / \lambda_i \geq 0, \forall\ i \in N_{p-r}$ then for a fixed k there exists a d in the range:

$$\frac{\sum_{i=1}^{r} \frac{k(\alpha_i^2 \lambda_i - \sigma^2) - 2\lambda_i \sigma^2}{\lambda_i(\lambda_i + k)^2} - \sum_{i=1}^{r} \frac{2\lambda_i \alpha_i \eta_i + (2\lambda_i + k)\eta_i^2 \lambda_i}{\lambda_i(\lambda_i + k)^2}}{\sum_{i=1}^{r} \frac{k(\alpha_i^2 \lambda_i + \sigma^2)}{\lambda_i(\lambda_i + k)^2} + \sum_{i=1}^{r} \frac{k(2\alpha_i \eta_i + \eta_i^2 / \lambda_i)}{\lambda_i(\lambda_i + k)^2}} < d < 1 \tag{3.4}$$

such that the $r - (k, d)$ class estimator dominates the OLS estimator. If $\alpha_i \eta_i > 0$ for all $i \in N_r$, then the lower limit of d decreases due to omission of relevant regressors and thus the dominance range of the $r - (k, d)$ class estimator over the OLS estimator increases.

Furthermore, if $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2 / \lambda_i < 0$, for some $i = r + 1, r + 2, ..., p$, no definite conclusion can be drawn regarding dominance of one over the other. The results obtained are reported in the form of the following theorem.

Theorem 3.1

(i) If $\sigma^2 - \lambda_i\alpha_i^2 + \eta_i^2/\lambda_i \geq 0$, for all $i \in N$, the $r - (k,d)$ class estimator dominates the OLS estimator for all $k > 0$ and $0 < d < 1$. The odds for superiority of $r - (k,d)$ class estimator over the OLS estimator increases in the misspecified model.

(ii) If $\sigma^2 - \lambda_i\alpha_i^2 + \eta_i^2/\lambda_i < 0$, $i \in N_r$ and $\sigma^2 - \lambda_i\alpha_i^2 + \eta_i^2/\lambda_i \geq 0$, for all $i \in N_{p-r}$, the $r - (k,d)$ class estimator dominates the OLS estimator for all $k > 0$ and d such that it satisfies (3.4). The range of dominance of $r - (k,d)$ class estimator increases in the misspecified model provided $\alpha_i\eta_i$ is positive for all $i \in N_r$.

(iii) If $\sigma^2 - \lambda_i\alpha_i^2 + \dfrac{\eta_i^2}{\lambda_i} < 0$, for some $i = r+1, r+2, \dots, p$, no definite conclusion can be drawn regarding their dominance.

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

The MSE of the ORR estimator can be obtained by substituting $r = p$ and $d = 0$ in (2.10), given as:

$$MSE(\hat{\beta}(k)) = \sum_{i=1}^{p} \frac{\sigma^2\lambda_i + k^2\alpha_i^2}{(\lambda_i+k)^2} + \sum_{i=1}^{p} \frac{\eta_i^2 - 2k\alpha_i\eta_i}{(\lambda_i+k)^2}. \tag{3.5}$$

Using (2.10) and (3.5), the difference between the MSEs,; say Δ_2, is given by:

$$\Delta_2 = MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_r(k,d))$$
$$= \sum_{i=1}^{p} \frac{\sigma^2\lambda_i + k^2\alpha_i^2}{(\lambda_i+k)^2} - \sum_{i=1}^{r} \frac{\sigma^2(\lambda_i+kd)^2 + k^2(1-d)^2\lambda_i\alpha_i^2}{\lambda_i(\lambda_i+k)^2} - \sum_{i=r+1}^{p} \alpha_i^2$$
$$+ \sum_{i=1}^{p} \frac{-2k\alpha_i\eta_i + \eta_i^2}{(\lambda_i+k)^2} - \sum_{i=1}^{r} \frac{-2k(1-d)(\lambda_i+kd)\alpha_i\eta_i + (\lambda_i+kd)^2\eta_i^2/\lambda_i}{\lambda_i(\lambda_i+k)^2}. \tag{3.6}$$

On further simplification, we obtain:

$$\Delta_2 = kd \sum_{i=1}^{r} \frac{[k[\lambda_i\alpha_i^2 + 2\alpha_i\eta_i - d(\lambda_i\alpha_i^2 + 2\alpha_i\eta_i + \sigma^2 + \eta_i^2)] - 2\lambda_i(\sigma^2 + \alpha_i\eta_i + \eta_i^2)]}{\lambda_i(\lambda_i+k)^2}$$
$$+ \sum_{i=r+1}^{p} \frac{(\sigma^2 - \lambda_i\alpha_i^2 + \eta_i^2/\lambda_i) - 2k(\alpha_i^2 + \alpha_i\eta_i/\lambda_i)}{(\lambda_i+k)^2}. \tag{3.7}$$

From (3.7), it can be noticed that the $r - (k,d)$ class estimator dominates the ORR estimator if both summations are positive, that is:

$$k(2\lambda_i\alpha_i^2 + \alpha_i\eta_i) - kd(\sigma^2 + \lambda_i\alpha_i^2 + \alpha_i\eta_i + \lambda_i\eta_i^2) - 2\lambda_i(\sigma^2 + \alpha_i\eta_i + \lambda_i\eta_i^2) > 0 \quad \text{for all } i \in N_r \tag{3.8}$$

and

$$\left(\sigma^2 - \lambda_i\alpha_i^2 + \frac{\eta_i^2}{\lambda_i}\right) - 2k\left(\alpha_i^2 + \frac{\alpha_i\eta_i}{\lambda_i}\right) > 0 \quad \text{for all } i \in N_{p-r}. \tag{3.9}$$

If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_{p-r}$ then (3.9) holds when:

$$k < \frac{\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i}{2(\alpha_i^2 + \alpha_i \eta_i/\lambda_i)} \quad \text{for all } i \in N_{p-r}. \tag{3.10}$$

If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i < 0$ for all $i \in N_{p-r}$, then (3.9) does not hold true. However, if $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i < 0$ for some $i \in N_{p-r}$, a positive k can be found such that the second summation in Δ_2, i.e. $\sum_{i=r+1}^{p} \frac{(\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i) - 2k(\alpha_i^2 + \alpha_i \eta_i/\lambda_i)}{(\lambda_i + k)^2}$ is positive.

Further, from (3.8) we obtain:
$$d < \frac{k(2\lambda_i \alpha_i^2 + \alpha_i \eta_i) - 2\lambda_i(\sigma^2 + \alpha_i \eta_i + \lambda_i \eta_i^2)}{k(\sigma^2 + \lambda_i \alpha_i^2 + \alpha_i \eta_i + \lambda_i \eta_i^2)} \quad \text{for all } i \in N_r. \tag{3.11}$$

For d to be a positive in (3.11) , $k(2\lambda_i \alpha_i^2 + \alpha_i \eta_i) - 2\lambda_i(\sigma^2 + \alpha_i \eta_i + \lambda_i \eta_i^2)$ should be a positive value for all $i \in N_r$, i.e.

$$k > \frac{\sigma^2 + \alpha_i \eta_i + \lambda_i \eta_i^2}{\alpha_i^2 + \alpha_i \eta_i/2\lambda_i} \quad \text{for all } i \in N_r. \tag{3.12}$$

If upper bound of d in (3.11) is greater than 1, any value smaller than 1 can be taken, which satisfies (3.11) and $0 < d < 1$.

The conditions of dominance of the $r - (k, d)$ class estimator over the ORR estimator under MSE criterion is stated below in the form of the following theorem:

Theorem 3.2

(i) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_r$ and $k > \frac{\sigma^2 + \alpha_i \eta_i + \lambda_i \eta_i^2}{\alpha_i^2 + \alpha_i \eta_i/2\lambda_i}$ for all $i \in N_r$, the $r - (k, d)$ class estimator dominates the ORR estimator if $k < \min_{i \in N_{p-r}} \frac{\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i}{2(\alpha_i^2 + \alpha_i \eta_i/\lambda_i)}$ and $0 < d < \min\left\{1, \min_{i \in N_r} \frac{k(2\lambda_i \alpha_i^2 + \alpha_i \eta_i) - 2\lambda_i(\sigma^2 + \alpha_i \eta_i + \lambda_i \eta_i^2)}{k(\sigma^2 + \lambda_i \alpha_i^2 + \alpha_i \eta_i + \lambda_i \eta_i^2)}\right\}$.

(ii) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for some $i \in N_{p-r}$ and $k > \frac{\sigma^2 + \alpha_i \eta_i + \lambda_i \eta_i^2}{\alpha_i^2 + \alpha_i \eta_i/2\lambda_i}$ for all $i \in N_r$, the $r - (k, d)$ class estimator dominates the ORR estimator for a value of k such that $\sum_{i=r+1}^{p} \frac{\lambda_i((\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i) - 2k(\alpha_i^2 + \alpha_i \eta_i/\lambda_i))}{(\lambda_i + k)^2}$ is positive and $0 < d < \min\left\{1, \min_{i \in N_r} \frac{k(2\lambda_i \alpha_i^2 + \alpha_i \eta_i) - 2\lambda_i(\sigma^2 + \alpha_i \eta_i + \lambda_i \eta_i^2)}{k(\sigma^2 + \lambda_i \alpha_i^2 + \alpha_i \eta_i + \lambda_i \eta_i^2)}\right\}$.

3.3. Comparison of the $r - (k, d)$ class estimator with the PCR estimator

On substituting $k = 0$ in (2.10), the MSE of the PCR estimator can be obtained as:

$$MSE(\hat{\beta}_r) = \sum_{i=1}^{r} \frac{\sigma^2}{\lambda_i} + \sum_{i=r+1}^{p} \alpha_i^2 + \sum_{i=1}^{r} \frac{\eta_i^2}{\lambda_i^2}. \tag{3.13}$$

From (2.10) and (3.13), the difference in MSEs, say Δ_3, is given by:

$$\Delta_3 = MSE(\hat{\beta}_r) - MSE(\hat{\beta}_r(k, d))$$
$$= \sum_{i=1}^{r} \frac{\sigma^2}{\lambda_i} + \sum_{i=r+1}^{p} \alpha_i^2 - \sum_{i=1}^{r} \frac{\sigma^2(\lambda_i + kd)^2 + k^2(d-1)^2 \lambda_i \alpha_i^2}{\lambda_i(\lambda_i + k)^2} -$$
$$\sum_{i=r+1}^{p} \alpha_i^2$$
$$+ \sum_{i=1}^{r} \frac{\eta_i^2}{\lambda_i^2} - \sum_{i=1}^{r} \frac{2k(1-d)\lambda_i(\lambda_i + kd)\alpha_i\eta_i - (\lambda_i + kd)^2\eta_i^2}{\lambda_i^2(\lambda_i + k)^2}.$$

On further simplifying it, we get:

$$\Delta_3 = k(1-d) \sum_{i=1}^{r} \frac{\left[2\lambda_i\left(\sigma^2 + \alpha_i\eta_i + \frac{\eta_i^2}{\lambda_i}\right) + k\left\{\left(\sigma^2 - \lambda_i\alpha_i^2 + \frac{\eta_i^2}{\lambda_i}\right) + d\left(\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i + \frac{\eta_i^2}{\lambda_i}\right)\right\}\right]}{\lambda_i(\lambda_i + k)^2}. \tag{3.14}$$

It can be observed from the above expression that if $\sigma^2 - \lambda_i\alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_r$, the $r - (k, d)$ class estimator dominates the PCR estimator for all $k > 0$ and $0 < d < 1$. Evidently, the odds for superiority of the $r - (k, d)$ class estimator over the PCR estimator increases due to misspecification.

If $\sigma^2 - \lambda_i\alpha_i^2 + \eta_i^2/\lambda_i < 0$ for all $i \in N_r$, then Δ_3 is positive when $2\lambda_i(\sigma^2 + \alpha_i\eta_i + \eta_i^2/\lambda_i) + k((\sigma^2 - \lambda_i\alpha_i^2 + \eta_i^2/\lambda_i) + d(\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i + \eta_i^2/\lambda_i)$ is positive for all $i \in N_r$, which can be rewritten as

$$2\lambda_i(\sigma^2 + \alpha_i\eta_i + \eta_i^2/\lambda_i) - k((\lambda_i\alpha_i^2 - \sigma^2 - \eta_i^2/\lambda_i) - d(\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i + \eta_i^2/\lambda_i) > 0 \text{ for all } i \in N_r$$
$$\tag{3.15}$$

If $\left(\lambda_i\alpha_i^2 - \sigma^2 - \frac{\eta_i^2}{\lambda_i}\right) - d\left(\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i + \frac{\eta_i^2}{\lambda_i}\right) < 0$, forall $i \in N_r$, i.e.

$$d > \frac{(\lambda_i\alpha_i^2 - \sigma^2 - \eta_i^2\lambda_i)}{(\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i + \eta_i^2/\lambda_i)} \quad \text{for all} i \in N_r \tag{3.16}$$

then Δ_3 is positive for all $k > 0$. It is noticeable that the lower limit of d decreases due to misspecification, thus a wider range for the dominance of the $r - (k, d)$ class estimator over the PCR estimator is obtained as compared with no misspecification.

Further, if

$$d < \frac{(\lambda_i \alpha_i^2 - \sigma^2 - \eta_i^2/\lambda_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \quad \text{for all } i \in N_r \tag{3.17}$$

holds, then Δ_3 is positive for k such that

$$k < \frac{2\lambda_i(\sigma^2 + \alpha_i \eta_i + \eta_i^2/\lambda_i)}{((\lambda_i \alpha_i^2 - \sigma^2 - \eta_i^2/\lambda_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \quad \text{for all } i \in N_r. \tag{3.18}$$

By rewriting (3.18), it is observed that the upper limit of k increases due to misspecification. Additionally, the upper limit of d decreases. Thus, due to misspecification a wider range of k for a shorter range of d in which the $r - (k, d)$ class estimator dominates the PCR estimator is obtained.

The comparisons can be concluded in the following theorem.

Theorem 3.3

(i) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_r$, the $r - (k, d)$ class estimator dominates the PCR estimator for all $k > 0$ and $0 < d < 1$. The odds for superiority of the $r - (k, d)$ class estimator over the PCR estimator increases due to misspecification.

(ii) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i < 0$ for all $i \in N_r$ and $\max_{i \in N_r}\{(\lambda_i \alpha_i^2 - \sigma^2 - \eta_i^2/\lambda_i)/(\lambda_i \alpha_i^2 + \sigma^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i) < d < 1$, then the $r - (k, d)$ class estimator dominates the PCR estimator for all $k > 0$. The dominance range of the $r - (k, d)$ class estimator over the PCR estimator increases due to misspecification.

(iii) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i < 0$ for all $i \in N_r$ and $0 < d < \min_{i \in N_r}\{(\lambda_i \alpha_i^2 - \sigma^2 - \eta_i^2/\lambda_i)/(\lambda_i \alpha_i^2 + \sigma^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)$, then the $r - (k, d)$ class estimator dominates the PCR estimator if $k < \min_{i \in N_r}\{2\lambda_i(\sigma^2 + \alpha_i \eta_i + \eta_i^2/\lambda_i)/((\lambda_i \alpha_i^2 - \sigma^2 - \eta_i^2/\lambda_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i))\}$. The range of k increases while the range of d decreases due to misspecification, in which the $r - (k, d)$ class estimator dominates the PCR estimator.

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

The MSE of the $r - k$ class estimator can be obtained by substituting $d = 0$ in (2.10), given as:

$$MSE(\hat{\beta}_r(k)) = \sum_{i=1}^{r} \frac{\lambda_i \sigma^2 + k^2 \alpha_i^2}{(\lambda_i + k)^2} + \sum_{i=r+1}^{p} \alpha_i^2 + \sum_{i=1}^{r} \frac{\eta_i^2 - 2k\alpha_i \eta_i}{(\lambda_i + k)^2}. \tag{3.19}$$

From (2.10) and (3.19), the difference between the MSEs, say Δ_4, is given by:

$$\Delta_4 = MSE(\hat{\beta}_r(k)) - MSE(\hat{\beta}_r(k,d))$$
$$= \sum_{i=1}^{r} \frac{\lambda_i \sigma^2 + k^2 \alpha_i^2}{(\lambda_i + k)^2} - \sum_{i=1}^{r} \frac{\sigma^2 (\lambda_i + kd)^2 + k^2 (d-1)^2 \lambda_i \alpha_i^2}{\lambda_i (\lambda_i + k)^2}$$
$$+ \sum_{i=1}^{r} \frac{\eta_i^2 - 2k\alpha_i \eta_i}{(\lambda_i + k)^2} - \sum_{i=1}^{r} \frac{(\lambda_i + kd)^2 \eta_i^2 - 2k(1-d)\lambda_i (\lambda_i + kd)\alpha_i \eta_i}{\lambda_i^2 (\lambda_i + k)^2}. \quad (3.20)$$

On further simplification, we get:

$$\Delta_4 = kd \sum_{i=1}^{r} \frac{\left[k(2(\lambda_i \alpha_i^2 - \alpha_i \eta_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \eta_i^2/\lambda_i)) - 2\lambda_i(\sigma^2 + \alpha_i \eta_i + \eta_i^2/\lambda_i)\right]}{\lambda_i(\lambda_i + k)^2}. \quad (3.21)$$

From (3.21), the $r - (k,d)$ class estimator dominates the $r - k$ class estimator when

$$k\left(2(\lambda_i \alpha_i^2 - \alpha_i \eta_i) - d\left(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \frac{\eta_i^2}{\lambda_i}\right)\right) - 2\lambda_i\left(\sigma^2 + \alpha_i \eta_i + \frac{\eta_i^2}{\lambda_i}\right) >$$
$$0 \quad \text{forall } i \in N_r. \quad (3.22)$$

If $\lambda_i \alpha_i^2 - \alpha_i \eta_i > 0$ for all $i \in N_r$ and $2(\lambda_i \alpha_i^2 - \alpha_i \eta_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \eta_i^2/\lambda_i) > 0$ for all $i \in N_r$, that is:

$$d < \frac{2(\lambda_i \alpha_i^2 - \alpha_i \eta_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \eta_i^2/\lambda_i)} \quad \text{forall } i \in N_r \quad (3.23)$$

then Δ_4 is positive when k is such that:

$$k > \frac{2\lambda_i(\sigma^2 + \alpha_i \eta_i + \eta_i^2/\lambda_i)}{(2(\lambda_i \alpha_i^2 - \alpha_i \eta_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \eta_i^2/\lambda_i))} \quad \text{forall } i \in N_r. \quad (3.24)$$

If $\lambda_i \alpha_i^2 - \alpha_i \eta_i > 0$ for all $i \in N_r$ and $2(\lambda_i \alpha_i^2 - \alpha_i \eta_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \eta_i^2/\lambda_i) < 0$ for all $i \in N_r$, that is:

$$d > \frac{2(\lambda_i \alpha_i^2 - \alpha_i \eta_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \eta_i^2/\lambda_i)} \quad \text{forall } i \in N_r \quad (3.25)$$

then Δ_4 is negative for all $k > 0$.

If $\lambda_i \alpha_i^2 - \alpha_i \eta_i < 0$ for all $i \in N_r$, then Δ_4 is negative for all $k > 0$ and $0 < d < 1$.

The results obtained are given in the following theorem.

Theorem 3.4

(i) If $\lambda_i \alpha_i^2 - \alpha_i \eta_i > 0$ for all $i \in N_r$ and $0 < d < \min_{i \in N_r}\left\{\frac{2(\lambda_i \alpha_i^2 - \alpha_i \eta_i)}{\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \eta_i^2/\lambda_i}\right\}$, then the $r - (k,d)$ class estimator

dominates the $r - k$ class estimator for values of k such that

$$k > \max_{i \in N_r} \left\{ \frac{2\lambda_i(\sigma^2 + \alpha_i\eta_i + \eta_i^2/\lambda_i)}{(2(\lambda_i\alpha_i^2 - \alpha_i\eta_i) - d(\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i - \eta_i^2/\lambda_i))} \right\}.$$

(ii) If $\lambda_i\alpha_i^2 - \alpha_i\eta_i > 0$ for all $i \in N_r$ and $\max_{i \in N_r} \left\{ \frac{2(\lambda_i\alpha_i^2 - \alpha_i\eta_i)}{\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i - \eta_i^2/\lambda_i} \right\} < d < 1$, then the $r - (k, d)$ class estimator dominates the $r - k$ class estimator for all values of $k > 0$.

(iii) If $\lambda_i\alpha_i^2 - \alpha_i\eta_i < 0$ for all $i \in N_r$, the $r - k$ class estimator dominates the $r - (k, d)$ class estimator for all values of $k > 0$ and $0 < d < 1$.

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

The MSE of the two-parameter class estimator can be obtained by substituting $r = p$ in (2.10), given as:

$$MSE(\hat{\beta}(k, d)) = \sum_{i=1}^{p} \frac{\sigma^2(\lambda_i + kd)^2 + k^2(1-d)^2\lambda_i\alpha_i^2}{\lambda_i(\lambda_i + k)^2}$$
$$+ \sum_{i=1}^{p} \frac{(\lambda_i + kd)^2\eta_i^2 - 2k(1-d)\lambda_i(\lambda_i + kd)\alpha_i\eta_i}{\lambda_i^2(\lambda_i + k)^2}. \tag{3.26}$$

From (2.10) and (3.26), the difference in the MSEs, denoted as Δ_5, is given by:

$$\Delta_5 = MSE(\hat{\beta}(k, d)) - MSE(\hat{\beta}_r(k, d))$$
$$= \sum_{i=1}^{p} \frac{\sigma^2(\lambda_i + kd)^2 + k^2(1-d)^2\lambda_i\alpha_i^2}{\lambda_i(\lambda_i + k)^2} - \sum_{i=1}^{r} \frac{\sigma^2(\lambda_i + kd)^2 + k^2(d-1)^2\lambda_i\alpha_i^2}{\lambda_i(\lambda_i + k)^2} - \sum_{i=r+1}^{p} \alpha_i^2$$
$$+ \sum_{i=1}^{p} \frac{(\lambda_i + kd)^2\eta_i^2 - 2k(1-d)\lambda_i(\lambda_i + kd)\alpha_i\eta_i}{\lambda_i^2(\lambda_i + k)^2} - \sum_{i=1}^{r} \frac{(\lambda_i + kd)^2\eta_i^2 - 2k(1-d)\lambda_i(\lambda_i + kd)\alpha_i\eta_i}{\lambda_i^2(\lambda_i + k)^2}.$$

which can be further simplified as:

$$\Delta_5 = \sum_{i=r+1}^{p} \frac{(\lambda_i + kd)[\lambda_i(\sigma^2 - \lambda_i\alpha_i^2 + \eta_i^2/\lambda_i) + k(d(\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i + \eta_i^2/\lambda_i) - 2(\lambda_i\alpha_i^2 + \alpha_i\eta_i))]}{\lambda_i(\lambda_i + k)^2}. \tag{3.27}$$

From (3.27), it is evident that Δ_5 is positive if

$$\lambda_i\left(\sigma^2 - \lambda_i\alpha_i^2 + \frac{\eta_i^2}{\lambda_i}\right) + k\left(d\left(\sigma^2 + \lambda_i\alpha_i^2 + 2\alpha_i\eta_i + \frac{\eta_i^2}{\lambda_i}\right) - 2(\lambda_i\alpha_i^2 + \alpha_i\eta_i)\right) > 0 \quad \text{for all } i \in N_{p-r} \tag{3.28}$$

If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_{p-r}$, Δ_5 is positive when $d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i) - 2(\lambda_i \alpha_i^2 + \alpha_i \eta_i) > 0$ for all $i \in N_{p-r}$, i.e.

$$d > \frac{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \quad \text{forall } i \in N_{p-r} \tag{3.29}$$

for all values of $k > 0$.

However, when

$$d < \frac{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \quad \text{forall } i \in N_{p-r} \tag{3.30}$$

Δ_5 is positive for the values of k such that

$$k < \frac{\lambda_i(\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i)}{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \quad \text{forall } i \in N_{p-r} \tag{3.31}$$

Furthermore, if $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i < 0$ for all $i \in N_{p-r}$ and d satisfies (3.29), Δ_5 is positive for the values of k, which satisfies

$$k > \frac{\lambda_i(\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i)}{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \quad \text{forall } i \in N_{p-r}. \tag{3.32}$$

And, if d satisfies (3.30), Δ_5 is negative for all values of $k > 0$ and $0 < d < 1$. The comparisons can be concluded in the following theorem.

Theorem 3.5

(i) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_{p-r}$ and $\max_{i \in N_{p-r}} \left\{ \frac{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \right\} < d < 1$, then $\Delta_5 > 0$ for all $k > 0$.

(ii) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_{p-r}$ and $0 < d < \min_{i \in N_{p-r}} \left\{ \frac{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \right\}$, then $\Delta_5 > 0$ when k is such that

$$k < \min_{i \in N_{p-r}} \left\{ \frac{\lambda_i(\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i)}{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \right\}$$

(iii) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_{p-r}$ and $\max_{i \in N_{p-r}} \left\{ \frac{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \right\} < d < 1$, then $\Delta_5 > 0$ when k is such that

$$k > \max_{i \in N_{p-r}} \left\{ \frac{\lambda_i(\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i)}{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i) - d(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \right\}.$$

(iv) If $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i > 0$ for all $i \in N_{p-r}$ and $0 < d < \min_{i \in N_{p-r}} \left\{ \frac{2(\lambda_i \alpha_i^2 + \alpha_i \eta_i)}{(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i + \eta_i^2/\lambda_i)} \right\}$, then $\Delta_5 < 0$ for all values of $k > 0$.

In this section, conditions for dominance of the $r - (k, d)$ class estimator over the OLS, ORR, PCR, $r - k$ class estimator and the two-parameter class estimator under the MSE criterion in the misspecified model have been obtained. However, the range of dominance does not remain the same in the misspecified model as it is in the model assumed to be correct. Moreover, the depletion or enlargement of the dominance range for the $r - (k, d)$ class estimator over the other competing estimators depend on certain parametric conditions. For instance, if $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i \geq 0$, for all $i \in N$, the range of dominance of the $r - (k, d)$ class estimator over the OLS estimator increases in the misspecified model. Furthermore, a Monte Carlo study has been conducted to understand the effect of misspecification on the dominance of the $r - (k, d)$ class estimator over the other competing estimators.

4. Monte Carlo simulation

To compare the dominance of the estimators in true (when there is no misspecification in the model) and misspecified model, the regressors have been generated by the method given in McDonald and Galarneau (1975) and Gibbons(1981), which is defined as:

$$X_i = (1 - \rho^2)^{\frac{1}{2}} w_i + \rho w_{p+1}, \qquad i = 1,2, \dots, p,$$
$$Z_j = (1 - \rho^2)^{1/2} w_j + \rho w_{q+1}, \qquad j = 1,2, \dots, q.$$

where w_i and w_j are $n \times 1$ vectors of independent standard normal pseudo-random numbers, ρ is specified so that the correlation between any two regressors is given by ρ^2. The dependent variable y has been generated as follows:

$$y = D\zeta + u = X\beta + Z\gamma + u; \qquad u \sim N(0, \sigma^2 I) \qquad (4.1)$$

where $\zeta = [\beta\gamma]$. u is a vector of normal pseudo-random numbers with standard deviation σ. Following McDonald Galarneau (1975), Gibbons (1981), Kibria (2003) and others, ζ has been chosen as the normalized eigenvector corresponding to the largest eigenvalue of the $D'D$ matrix. As this study is aimed at studying the effect of the omission of relevant regressors on the performance of some competing estimators of β, the following is estimated for the model (4.1) and the misspecified model: when there is no misspecification, both X and Z have been used in estimation and when there is misspecification due to the omission of relevant regressors, information in Z matrix has not been used to estimate β. For example, the OLS estimator for the misspecified model is obtained by:

$$\hat{\beta}_M = (X'X)^{-1}X'y$$

and the OLS estimate of β in the case of no misspecification is obtained by taking first p components of the OLS estimate of ζ, given as:

$$\hat{\zeta} = (D'D)^{-1}D'y = [\hat{\beta}_T \quad \hat{\gamma}_T]'.$$

In this study, simulation is done for some selected values of n, p, q, ρ, σ^2, k and d to compare the performance of the estimators. The values of the parameters are taken as: $n = 50$; $p = 5$; $q = 3$; $\rho = 0.95, 0.99$; $\sigma = 0.5, 1$; $k = 0.1, 0.5, 0.9, 1.5, 5$ and $d = 0.1, 0.5, 0.9$. The value of r is decided by a scree plot, which is drawn between eigenvalues and components (see Johnson and Wichern (2007)). For each parametric combination, the simulation process has been repeated 2500 times and the estimated MSE (EMSE) is calculated by the following formula

$$EMSE(\hat{\beta}) = \frac{1}{2500} \sum_{i=1}^{2500} (\hat{\beta}_{(i)} - \beta)' (\hat{\beta}_{(i)} - \beta), \qquad (4.2)$$

where $\hat{\beta}_{(i)}$ is the estimated value of β in i^{th} iteration. The results of the simulation are shown in Tables 1 to 4, where EMSE of estimators in true model and in misspecified model are denoted by $EMSE_T$ and $EMSE_M$, respectively, and $\tilde{\beta}$, $\tilde{\beta}(k)$, $\tilde{\beta}_r$, $\tilde{\beta}_r(k)$, $\tilde{\beta}(k,d)$ and $\tilde{\beta}_r(k,d)$ denote the OLS, ORR, PCR, $r - k$ class, two-parameter class and $r - (k,d)$ class estimators respectively. The following remarks are made from simulation results:

Table 1. Estimated MSE of the estimators for true and misspecified model when $\rho = 0.95$ and $\sigma = 0.5$

		$d = 0.1$		$d = 0.5$		$d = 0.9$	
k		$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$
0.1	$\tilde{\beta}$	0.2801295	0.4486118	0.2801295	0.4486118	0.2801295	0.4486118
	$\tilde{\beta}(k)$	0.2640343	0.4350360	0.2640342	0.4350360	0.2640342	0.4350360
	$\tilde{\beta}(k,d)$	0.2656187	0.4363773	0.2720124	0.4417786	0.2784949	0.4472379
	$\tilde{\beta}_r$	0.0003777	0.1855473	0.0003777	0.1855473	0.0003777	0.1855473
	$\tilde{\beta}_r(k)$	0.0003774	0.1851850	0.0003774	0.1851850	0.0003774	0.1851850
	$\tilde{\beta}_r(k,d)$	0.0003774	0.1852212	0.0003775	0.1853661	0.0003776	0.1855110
0.5	$\tilde{\beta}$	0.2801295	0.4486118	0.2801295	0.4486118	0.2801295	0.4486118
	$\tilde{\beta}(k)$	0.2131531	0.3899955	0.2131531	0.3899955	0.2131531	0.3899955
	$\tilde{\beta}(k,d)$	0.2193777	0.3955276	0.2453274	0.4183884	0.2729588	0.4424207
	$\tilde{\beta}_r$	0.0003777	0.1855473	0.0003777	0.1855473	0.0003777	0.1855473
	$\tilde{\beta}_r(k)$	0.0003768	0.1837417	0.0003768	0.1837417	0.0003768	0.1837417
	$\tilde{\beta}_r(k,d)$	0.0003768	0.1839219	0.0003770	0.1846434	0.0003775	0.1853663
0.9	$\tilde{\beta}$	0.2801295	0.4486118	0.2801295	0.4486118	0.2801295	0.4486118
	$\tilde{\beta}(k)$	0.1769369	0.3558756	0.1769369	0.3558756	0.1769369	0.3558756
	$\tilde{\beta}(k,d)$	0.1860480	0.3642675	0.2251774	0.3997944	0.2686021	0.4384565
	$\tilde{\beta}_r$	0.0003777	0.1855473	0.0003777	0.1855473	0.0003777	0.1855473
	$\tilde{\beta}_r(k)$	0.0003773	0.1823080	0.0003773	0.1823080	0.0003773	0.1823080
	$\tilde{\beta}_r(k,d)$	0.0003771	0.1826306	0.0003768	0.1839241	0.0003774	0.1852221
1.5	$\tilde{\beta}$	0.2801295	0.4486118	0.2801295	0.4486118	0.2801295	0.4486118
	$\tilde{\beta}(k)$	0.1388449	0.3179755	0.1388449	0.3179755	0.1388449	0.3179755
	$\tilde{\beta}(k,d)$	0.1504966	0.3291388	0.2026073	0.3780150	0.2635243	0.4336479
	$\tilde{\beta}_r$	0.0003777	0.1855473	0.0003777	0.1855473	0.0003777	0.1855473
	$\tilde{\beta}_r(k)$	0.0003799	0.1801752	0.0003799	0.1801752	0.0003799	0.1801752
	$\tilde{\beta}_r(k,d)$	0.0003790	0.1807089	0.0003770	0.1828514	0.0003772	0.1850065

Table 1. Estimated MSE of the estimators for true and misspecified model when
$\rho = 0.95$ and $\sigma = 0.5$ (cont.)

		$d = 0.1$		$d = 0.5$		$d = 0.9$	
k		$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$
5	$\tilde{\beta}$	0.2801295	0.4486118	0.2801295	0.4486118	0.2801295	0.4486118
	$\tilde{\beta}(k)$	0.0526366	0.2207759	0.0526366	0.2207759	0.0526366	0.2207759
	$\tilde{\beta}(k,d)$	0.0666946	0.2360594	0.1422406	0.3138604	0.2486889	0.4183282
	$\tilde{\beta}_r$	0.0003777	0.1855473	0.0003777	0.1855473	0.0003777	0.1855473
	$\tilde{\beta}_r(k)$	0.0004400	0.1681494	0.0004400	0.1681494	0.0004400	0.1681494
	$\tilde{\beta}_r(k,d)$	0.0004267	0.1698506	0.0003892	0.1767411	0.0003768	0.1837689

Table 2. Estimated MSE of the estimators for true and misspecified model when
$\rho = 0.95$ and $\sigma = 1$

		$d = 0.1$		$d = 0.5$		$d = 0.9$	
k		$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$
0.1	$\tilde{\beta}$	1.1205180	1.1783938	1.1205180	1.1783938	1.1205180	1.1783938
	$\tilde{\beta}(k)$	1.0561408	1.1288003	1.0561408	1.1288003	1.0561408	1.1288003
	$\tilde{\beta}(k,d)$	1.0624786	1.1336995	1.0880518	1.1534299	1.1139803	1.1733743
	$\tilde{\beta}_r$	0.0015107	0.1888630	0.0015107	0.1888630	0.0015107	0.1888630
	$\tilde{\beta}_r(k)$	0.0015098	0.1884980	0.0015098	0.1884980	0.0015098	0.1884980
	$\tilde{\beta}_r(k,d)$	0.0015099	0.1885345	0.0015103	0.1886804	0.0015106	0.1888264
0.5	$\tilde{\beta}$	1.1205180	1.1783938	1.1205180	1.1783938	1.1205180	1.1783938
	$\tilde{\beta}(k)$	0.8526281	0.9646305	0.8526281	0.9646305	0.8526281	0.9646305
	$\tilde{\beta}(k,d)$	0.8775255	0.9847905	0.9813187	1.0681334	1.0918374	1.1558011
	$\tilde{\beta}_r$	0.0015107	0.1888630	0.0015107	0.1888630	0.0015107	0.1888630
	$\tilde{\beta}_r(k)$	0.0015066	0.1870440	0.0015066	0.1870440	0.0015066	0.1870440
	$\tilde{\beta}_r(k,d)$	0.0015069	0.1872255	0.0015085	0.1879524	0.0015103	0.1886807
0.9	$\tilde{\beta}$	1.1205180	1.1783938	1.1205180	1.1783938	1.1205180	1.1783938
	$\tilde{\beta}(k)$	0.7077687	0.8408833	0.7077687	0.8408833	0.7077687	0.8408833
	$\tilde{\beta}(k,d)$	0.7442128	0.8713751	0.9007244	1.0005850	1.0744122	1.1413835
	$\tilde{\beta}_r$	0.0015107	0.1888630	0.0015107	0.1888630	0.0015107	0.1888630
	$\tilde{\beta}_r(k)$	0.0015044	0.1855997	0.0015044	0.1855997	0.0015044	0.1855997
	$\tilde{\beta}_r(k,d)$	0.0015048	0.1859248	0.0015069	0.1872278	0.0015099	0.1885353
1.5	$\tilde{\beta}$	1.1205180	1.1783938	1.1205180	1.1783938	1.1205180	1.1783938
	$\tilde{\beta}(k)$	0.5554001	0.7045712	0.5554001	0.7045712	0.5554001	0.7045712
	$\tilde{\beta}(k,d)$	0.6020087	0.7449187	0.8104502	0.9219414	1.0541030	1.1239768
	$\tilde{\beta}_r$	0.0015107	0.1888630	0.0015107	0.1888630	0.0015107	0.1888630
	$\tilde{\beta}_r(k)$	0.0015030	0.1834511	0.0015030	0.1834511	0.0015030	0.1834511
	$\tilde{\beta}_r(k,d)$	0.0015032	0.1839887	0.0015051	0.1861471	0.0015093	0.1883182
5	$\tilde{\beta}$	1.1205180	1.1783938	1.120518	1.1783938	1.120518	1.1783938
	$\tilde{\beta}(k)$	0.2103956	0.3713567	0.2103956	0.3713567	0.2103956	0.3713567
	$\tilde{\beta}(k,d)$	0.2666748	0.4242071	0.5689763	0.6975050	0.9947728	1.0698368
	$\tilde{\beta}_r$	0.0015108	0.1888630	0.0015108	0.1888630	0.0015108	0.1888630
	$\tilde{\beta}_r(k)$	0.0015404	0.1713349	0.0015404	0.1713349	0.0015404	0.1713349
	$\tilde{\beta}_r(k,d)$	0.0015303	0.1730490	0.0015059	0.1799914	0.0015066	0.1870714

Table 3. Estimated MSE of the estimators for true and misspecified model when $\rho = 0.99$ and $\sigma = 0.5$

k		$d = 0.1$		$d = 0.5$		$d = 0.9$	
		$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$
0.1	$\tilde{\beta}$	1.3733302	1.4155047	1.3733302	1.4155047	1.3733302	1.4155047
	$\tilde{\beta}(k)$	1.0501954	1.1614661	1.0501954	1.1614661	1.0501954	1.1614661
	$\tilde{\beta}(k,d)$	1.0802694	1.1854445	1.2055422	1.2845257	1.3387773	1.3886753
	$\tilde{\beta}_r$	0.0003490	0.2090398	0.0003490	0.2090398	0.0003490	0.2090398
	$\tilde{\beta}_r(k)$	0.0003487	0.2086612	0.0003487	0.2086612	0.0003487	0.2086612
	$\tilde{\beta}_r(k,d)$	0.0003487	0.2086991	0.0003488	0.2088504	0.0003489	0.2090019
0.5	$\tilde{\beta}$	1.3733302	1.4155047	1.3733302	1.4155047	1.3733302	1.4155047
	$\tilde{\beta}(k)$	0.4930231	0.6729832	0.4930231	0.6729832	0.4930231	0.6729832
	$\tilde{\beta}(k,d)$	0.5592052	0.7310081	0.8724862	0.9991682	1.2634509	1.3250253
	$\tilde{\beta}_r$	0.0003490	0.2090398	0.0003490	0.2090398	0.0003490	0.2090398
	$\tilde{\beta}_r(k)$	0.0003482	0.207153	0.0003482	0.207153	0.0003482	0.207153
	$\tilde{\beta}_r(k,d)$	0.0003482	0.2073413	0.0003484	0.2080953	0.0003488	0.2088507
0.9	$\tilde{\beta}$	1.3733302	1.4155047	1.3733302	1.4155047	1.3733302	1.4155047
	$\tilde{\beta}(k)$	0.2931221	0.4843209	0.2931221	0.4843209	0.2931221	0.4843209
	$\tilde{\beta}(k,d)$	0.3626887	0.5471788	0.7264088	0.865856	1.2268551	1.2921258
	$\tilde{\beta}_r$	0.0003490	0.2090398	0.0003490	0.2090398	0.0003490	0.2090398
	$\tilde{\beta}_r(k)$	0.0003486	0.2056542	0.0003486	0.2056542	0.0003486	0.2056542
	$\tilde{\beta}_r(k,d)$	0.0003484	0.2059915	0.0003482	0.2073435	0.0003487	0.2087000
1.5	$\tilde{\beta}$	1.3733302	1.4155047	1.3733302	1.4155047	1.3733302	1.4155047
	$\tilde{\beta}(k)$	0.1645441	0.3596925	0.1645441	0.3596925	0.1645441	0.3596925
	$\tilde{\beta}(k,d)$	0.2297062	0.4196822	0.6141691	0.7609554	1.1967351	1.2643319
	$\tilde{\beta}_r$	0.0003490	0.2090398	0.0003490	0.2090398	0.0003490	0.2090398
	$\tilde{\beta}_r(k)$	0.0003509	0.2034235	0.0003509	0.2034235	0.0003509	0.2034235
	$\tilde{\beta}_r(k,d)$	0.0003501	0.2039817	0.0003483	0.2062221	0.0003486	0.2084747
5	$\tilde{\beta}$	1.3733302	1.4155047	1.3733302	1.4155047	1.3733302	1.4155047
	$\tilde{\beta}(k)$	0.0280793	0.2169173	0.0280793	0.2169173	0.0280793	0.2169173
	$\tilde{\beta}(k,d)$	0.0697681	0.2569569	0.4428262	0.5944911	1.1459688	1.2158268
	$\tilde{\beta}_r$	0.0003490	0.2090398	0.0003490	0.2090398	0.0003490	0.2090398
	$\tilde{\beta}_r(k)$	0.0004035	0.1908219	0.0004035	0.1908219	0.0004035	0.1908219
	$\tilde{\beta}_r(k,d)$	0.0003918	0.1926062	0.000359	0.1998268	0.0003482	0.2071805

Table 4. Estimated MSE of the estimators for true and misspecified model when $\rho = 0.99$ and $\sigma = 1$

k		$d = 0.1$		$d = 0.5$		$d = 0.9$	
		$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$	$EMSE_T$	$EMSE_M$
0.1	$\widetilde{\beta}$	5.4933208	4.9729356	5.4933208	4.9729356	5.4933208	4.9729356
	$\widetilde{\beta}(k)$	4.2007888	3.9736359	4.2007888	3.9736359	4.2007888	3.9736359
	$\widetilde{\beta}(k,d)$	4.3210844	4.0679655	4.8221727	4.4577291	5.3551101	4.8674052
	$\widetilde{\beta}_r$	0.0013960	0.2122746	0.0013960	0.2122746	0.0013960	0.2122746
	$\widetilde{\beta}_r(k)$	0.0013951	0.2118936	0.0013951	0.2118936	0.0013951	0.2118936
	$\widetilde{\beta}_r(k,d)$	0.0013952	0.2119316	0.0013956	0.2120840	0.0013959	0.2122365
0.5	$\widetilde{\beta}$	5.4933208	4.9729356	5.4933208	4.9729356	5.4933208	4.9729356
	$\widetilde{\beta}(k)$	1.9721133	2.0517057	1.9721133	2.0517057	1.9721133	2.0517057
	$\widetilde{\beta}(k,d)$	2.2368417	2.2799880	3.4899605	3.3349853	5.0538077	4.6169719
	$\widetilde{\beta}_r$	0.0013960	0.2122746	0.0013960	0.2122746	0.0013960	0.2122746
	$\widetilde{\beta}_r(k)$	0.0013923	0.2103753	0.0013923	0.2103753	0.0013923	0.2103753
	$\widetilde{\beta}_r(k,d)$	0.0013926	0.2105648	0.0013940	0.2113239	0.0013956	0.2120843
0.9	$\widetilde{\beta}$	5.4933208	4.9729356	5.4933208	4.9729356	5.4933208	4.9729356
	$\widetilde{\beta}(k)$	1.1725128	1.3109990	1.1725128	1.3109990	1.1725128	1.3109990
	$\widetilde{\beta}(k,d)$	1.4507813	1.5580588	2.9056605	2.8110399	4.9074278	4.4876081
	$\widetilde{\beta}_r$	0.0013960	0.2122746	0.0013960	0.2122746	0.0013960	0.2122746
	$\widetilde{\beta}_r(k)$	0.0013904	0.2088665	0.0013904	0.2088665	0.0013904	0.2088665
	$\widetilde{\beta}_r(k,d)$	0.0013907	0.2092061	0.0013926	0.2105671	0.0013952	0.2119325
1.5	$\widetilde{\beta}$	5.4933208	4.9729356	5.4933208	4.9729356	5.4933208	4.9729356
	$\widetilde{\beta}(k)$	0.6581957	0.8249431	0.6581957	0.8249431	0.6581957	0.8249431
	$\widetilde{\beta}(k,d)$	0.9188519	1.0601478	2.4567130	2.4000657	4.7869520	4.3785418
	$\widetilde{\beta}_r$	0.0013960	0.2122746	0.0013960	0.2122746	0.0013960	0.2122746
	$\widetilde{\beta}_r(k)$	0.0013891	0.2066210	0.0013891	0.2066210	0.0013891	0.2066210
	$\widetilde{\beta}_r(k,d)$	0.0013893	0.2071829	0.0013910	0.2094382	0.0013947	0.2117058
5	$\widetilde{\beta}$	5.4933208	4.9729356	5.4933208	4.9729356	5.4933208	4.9729356
	$\widetilde{\beta}(k)$	0.1121679	0.2972680	0.1121679	0.2972680	0.1121679	0.2972680
	$\widetilde{\beta}(k,d)$	0.2789891	0.4501910	1.7713719	1.7610914	4.5839115	4.1907251
	$\widetilde{\beta}_r$	0.0013960	0.2122746	0.0013960	0.2122746	0.0013960	0.2122746
	$\widetilde{\beta}_r(k)$	0.0014216	0.1939342	0.0014216	0.1939342	0.0014216	0.1939342
	$\widetilde{\beta}_r(k,d)$	0.0014128	0.1957307	0.0013915	0.2030001	0.0013923	0.2104030

Since ρ affects the structure of the design matrix, the estimated MSEs of $\widetilde{\beta}$ and $\widetilde{\beta}_r$ are the same for all values of k and d for a fixed σ in true and misspecified models. As expected, for higher value of σ, the estimated MSEs inflate for all the estimators in true and misspecified model as well. Similarly, when the collinearity among the regressors increases, the estimated MSEs of the estimators inflate in both the models.

As the theoretical results suggest, the MSE of the estimator may increase due to the omission of relevant regressors depending on the values of unknown

parameters. When we compare the performances of the estimators in true and misspecified model for all choices of the parameters involved almost all the estimators have larger estimated MSE in the misspecified model than in the case where there is no misspecification.

While examining the variations in the estimated MSE of the estimators with respect to the variations in k and d from Tables 1-4, we observe that as the value of k increases, the values of the estimated MSEs decrease for all the estimators considered here where k is involved. However, $\tilde{\beta}_r(k, d)$ in true model exhibits a pattern of concave up function of k, that is the estimated MSE of $\tilde{\beta}_r(k, d)$ first decreases and then increases after attaining a minimum value of the MSE with the increase in the value of k. In our simulation, the minimum value of the MSE of the $r - (k, d)$ class estimator when $d = 0.1, 0.5$ is attained for some value of k in between 0.9 to 1.5 and 1.5 to 5 for $\sigma = 0.5$ and $\sigma = 1$ respectively.

However, with the increase in the value of d, the estimated MSEs of $\tilde{\beta}(k, d)$ and $\tilde{\beta}_r(k, d)$ increase for the selected values of ρ and σ for both the models. The values of the estimated MSEs show that the $r - (k, d)$ class estimator performs better than the OLS, ORR, PCR, two-parameter class estimator for all chosen values of k, d, σ and ρ, although the dominance of the $r - (k, d)$ class estimator over the $r - k$ class estimator depends on the choices of k and d. In fact, the difference in the estimated MSE values of the $r - (k, d)$ class estimator and $r - k$ class estimator do not show much difference if seen up to the third or forth decimal places for small σ, however, if observed up to the sixth or seventh decimal places, the MSE of the $r - k$ class estimator is found to be less than that of the $r - (k, d)$ class estimator. For $\sigma = 1$, the $r - k$ class estimator shows dominance over the $r - (k, d)$ class estimator in the misspecified model, see Table 2 and 4, the reason being the condition of dominance of the $r - (k, d)$ class estimator over the $r - k$ class estimator (see Theorem 3.4) is not satisfied in this simulation.

5. Numerical example

In order to illustrate our theoretical results, in this section we now consider the data set on Total National Research and Development Expenditures as a Per cent of Gross National Product originally due to Gruber (1998), also analysed by Zhong and Yang (2007). It represents the relationship between the dependent variable Y, the percent spent by the U.S., and the four other independent variables X_1, X_2, X_3 and X_4. The variables X_1, X_2, X_3 and X_4, respectively represents the percent spent by France, West Germany, Japan and the former Soviet Union. The variables are standardized and the OLS estimator of $\beta = (\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4)'$ is obtained as $\beta = (0.6455, 0.0896, 0.1436, 0.1526)'$. We obtain the eigenvalues of $X'X$ as $\lambda_1 = 302.9626$, $\lambda_2 = 0.7283$, $\lambda_3 = 0.0446$, and $\lambda_4 = 0.0345$, and the condition number is approximately 8,776.382. Hence, the design matrix is quite ill-conditioned.

Now, let us consider that the investigator has omitted $Z = [X_4]$ mistakenly, which results in the misspecified model (2.2) with X matrix having 3 variables X_1, X_2, and X_3. The eigenvalues of the X matrix in the misspecified model are 161.38584077, 0.10961836 and 0.04454088, and the condition number is 3623.32, which indicates an ill-conditioned design matrix in the misspecified model. The OLS estimators of β, γ and σ^2 in the model (2.2) is obtained as $\beta =$ (0.80878236, 0.41402294, −0.09630492)', $\hat{\gamma} = \hat{\beta}_4 = 0.1526$, $\hat{\sigma}^2 = 0.002745$ respectively and we chose $r=2$. The values of k and d are chosen as: $k_0 = \hat{\sigma}^2/\alpha_{\max}^2 = 0.01178$, where $\alpha_{\max}$ is the maximum element of $\alpha=T'\beta$, which was suggested by Hoerl and Kennard (1970) and $d=0.0557$ is the positive solution of $\frac{\partial \, \text{MSE}(\hat{\beta}_r(k_0, d))}{\partial d} = 0$. The MSEs of the estimators are estimated by replacing β with the PCR estimator, which is an unbiased estimator, and are presented in Table 5 along with the estimated values of regression coefficients for both true and misspecified model. Figure 2 represents the estimated MSEs of the estimators in the two models.

From Table 5, we can see the sign of β_3 has changed in the misspecified model from positive to negative, which gives an evidence of the well-established results that the omission affects the estimation of parameters. Further, the estimated MSEs increase in the misspecified model as compared to the true model. We observe that the $r-(k,d)$ class estimator outperforms the OLS, ORR, two-parameter class and PCR estimators in MSE sense. However, the MSEs of the r-k class estimator and the r-(k,d) class estimator are almost equal and the difference can be only noticed at sixth decimal place. The dominance of the estimators can be easily seen in Figure 2.

On the other hand, from the results stated in Table 5 for the misspecified model, we see that the $r-(k, d)$ class estimator is superior to the OLS, ORR, two-parameter class and PCR estimators, and does not perform better than the $r-k$ class estimator under the MSE criterion. Moreover, the theoretical findings obtained in this study support the numerical results given in Table 5. Now, in order to verify the conditions of the dominance under MSE criterion, let us take Theorem 3.1, where we get $\sigma^2 - \lambda_i \alpha_i^2 + \eta_i^2/\lambda_i = -57.0280, -0.0042, 0.0027$, clearly condition (ii) of the theorem will be applied and the lower limits of d are −0.1592445−1.9733796, thus the $r-(k, d)$ class estimator dominates the OLS estimator for all values of d, which is the result obtained in the numerical illustration. Next, let us take the condition of dominance of $r-(k, d)$ class over the $r-k$ class given in Theorem 3.4; $\lambda_i \alpha_i^2 - \alpha_i \eta_i = 46.2405, 0.0037$ for $i =1,2$ and the value of d is 0.0557, which satisfies the condition (i) in Theorem 3.4 as the values of $2(\lambda_i \alpha_i^2 - \alpha_i \eta_i)/(\sigma^2 + \lambda_i \alpha_i^2 + 2\alpha_i \eta_i - \frac{\eta_i^2}{\lambda_i})$ for $i=1,2$ are 1.0856, 0.1421. Further, the value of the lower bound of k in condition (i) of Theorem 3.4 comes out to be 63.872591. Evidently the condition is not satisfied, hence the $r-(k, d)$ class estimator does not dominate the $r-k$ class estimator in this numerical illustration. Similarly, other conditions can also be verified.

Table 5. Estimated values of regression coefficients and estimated MSEs for true and misspecified model.

	$\widehat{\boldsymbol{\beta}}$	$\widehat{\boldsymbol{\beta}}(\boldsymbol{k})$	$\widehat{\boldsymbol{\beta}}(\boldsymbol{k},\boldsymbol{d})$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}(\boldsymbol{k})$	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}(\boldsymbol{k},\boldsymbol{d})$
True Model						
$\widehat{\boldsymbol{\beta}}_1$	0.645458	0.551069	0.556329	0.209956	0.209236	0.209276
$\widehat{\boldsymbol{\beta}}_2$	0.089588	0.115598	0.114148	0.240076	0.23948	0.239514
$\widehat{\boldsymbol{\beta}}_3$	0.143557	0.180012	0.17798	0.304667	0.302885	0.302984
$\widehat{\boldsymbol{\beta}}_4$	0.152618	0.163265	0.162671	0.186063	0.187951	0.187845
$\widehat{\mathrm{MSE}}$	**0.086702**	**0.061178**	**0.062331**	**0.025063**	**0.022632**	**0.022639**
Misspecified Model						
$\widehat{\boldsymbol{\beta}}_1$	0.808782	0.71553	0.720727	0.409013	0.399393	0.399929
$\widehat{\boldsymbol{\beta}}_2$	0.414023	0.438716	0.437339	0.662707	0.635374	0.636897
$\widehat{\boldsymbol{\beta}}_3$	-0.0963	-0.04272	-0.0457	-0.01613	0.020682	0.01863
$\widehat{\mathrm{MSE}}$	**0.276168**	**0.184595**	**0.189148**	**0.214464**	**0.146008**	**0.149414**
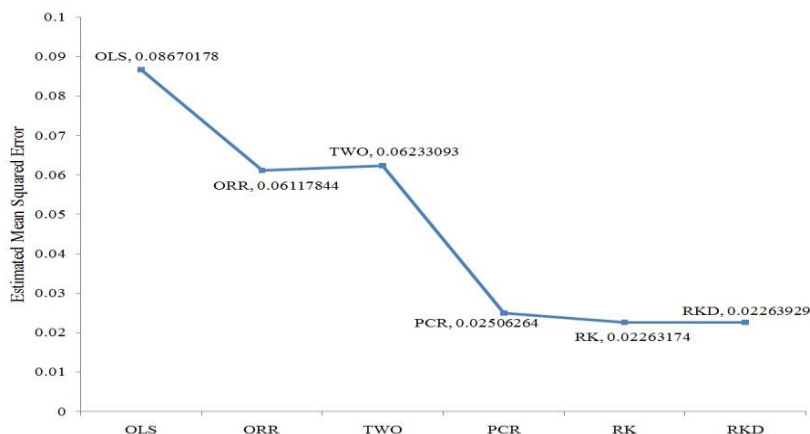
(a) In the case of no misspecification

Figure 2. Estimated MSE of the estimators
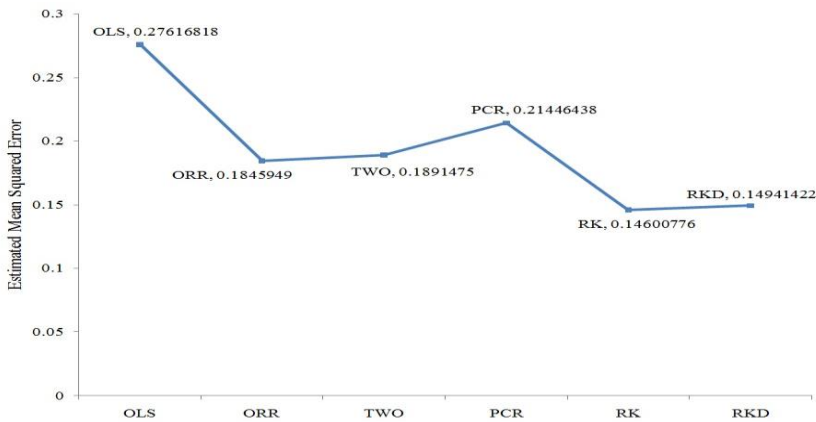
b)   When there is misspecification

Figure 2. Estimated MSE of the estimators (cont.)

6. Conclusion

In this paper the effect of misspecification due to omission of relevant regressors in a linear regression model when the problem of multicollinearity exists, on the dominance of the *r*-(*k*,*d*) class estimator over the other competing estimators have been studied. The dominance conditions of the *r*-(*k*,*d*) class estimator over the OLS, ORR, PCR, *r*-*k* class and the two-parameter class estimators have been derived under scalar mean squared error criterion. It has been observed that the MSE of the estimators may increase or decrease due to misspecification depending on the values of the unknown parameters. Similarly, the ranges of dominance of the *r*-(*k*,*d*) class estimator over the others may shrink or widen in the misspecified model. To understand the effect of misspecification on dominance of the *r*-(*k*,*d*) class estimator over the others a Monte Carlo simulation and a numerical example have been given and it is observed that the MSE of the estimators increases in the misspecified model as compared to the model assumed to be true. The *r*-(*k*,*d*) class estimator performs better than the OLS, ORR, two-parameter class estimator and the PCR estimator in the misspecified model as well for all chosen values of the parameters. However, the *r*-(*k*,*d*) class estimator and the *r*-*k* class estimator do equally well when observed up to few decimal places in simulation, whereas in the numerical example the *r*-*k* class estimator is found to be the most suited as an alternative to the OLS estimator in the misspecified model with multicollinearity. Hence, the study stuggests that the r-k class estimator or the r-(k,d) class estimators are a better choice over the other estimators considered in this study in the case of the misspecified model with multicollinearity.

Acknowledgement

REFERENCES

BAYE, M. R., PARKER, D. F., (1984). Combining ridge and principal component regression: A money demand illustration. Communications in Statistics- Theory and Methods, 13, pp. 197–205.

GIBBONS, D. G., (1981). A simulation study of some ridge estimators. Journal of the American Statistical Association, 76, pp. 131–139.

GRAHAM, M. H., (2003). Confronting multicollinearity in ecological multiple regression. Ecology, 84, pp. 2809–2815.

GRUBER, M., (1998). Improving efficiency by shrinkage: The James-Stein and ridge regression estimator. Marcel Dekker, Inc., New York.

HAMILTON, J. L., (1972). The demand for cigarettes: Advertising, the health scare, and the cigarette advertising ban. The Review of Economics and Statistics, 54, pp. 401–411.

HEIKKILA, E., (1988). Multicollinearity in regression models with multiple distance measures. Journal of Regional Science, 28, pp. 345–362.

HOERL, A. E., KENNARD, R. W., (1970). Ridge regression: Biased estimation for non-orthogonal problems. Technometrics, 12, pp. 55–67.

JOHNSON, R. A., WICHERN, D. W., (2007). Applied multivariate statistical analysis. Pearson-Prentice Hall, New Jersey.

KAÇIRANLAR, S., SAKALLIOĞLU, S., (2001). Combining the Liu estimator and the principal component regression estimator. Communications in Statistics – Theory and Methods, 30, pp. 2699–2705.

KADIYALA, K., (1986). Mixed regression estimator under misspecification. Economic Letters, 21, pp. 27–30.

KIBRIA, B., (2003). Performance of some new ridge regression estimators. Communications in Statistics – Theory and Methods, 32, pp. 419–435.

MAHAJAN, V., JAIN, A. K., BERGIER, M., (1977). Parameter estimation in marketing models in the presence of multicollinearity: An application of ridge regression. Journal of Marketing Research, 14, pp. 586–591.

MASSY, W. F., (1965). Principal components regression in exploratory statistical research. Journal of the American Statistical Association, 60, pp. 234–256.

MCDONALD, G. C., GALARNEAU, D. I., (1975). A Monte Carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association, 70, pp.407–416.

NOMURA, M., OHKUBO, T., (1985). A note on combining ridge and principal component regression. Communications in Statistics-Theory and Methods, 14, pp. 2489–2493.

ÖZKALE, M. R., KAÇIRANLAR, S., (2007). The restricted and unrestricted two parameter estimators. Communications in Statistics- Theory and Methods, 36, pp. 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, 82, pp. 653–688.

ÖZKALE, M., KAÇIRANLAR, S., (2008). Comparison of the $r - k$ class estimator to the ordinary least squares estimator under the Pitman's measure of closeness criterion. Statistical Papers, 49, pp. 503–512.

SARKAR, N., CHANDRA, S., (2015). Comparison of the r-(k,d) class estimator with some estimators for multicollinearity under the Mahalanobis loss function. Forthcoming paper in International Econometric Review, Vol. 7, issue 1, pp. 1–12.

SARKAR, N., (1989). Comparisons among some estimators in misspecified linear models with multicollinearity. Annals of Institute of Statistical Methods, 41, pp. 717–724.

SARKAR, N., (1996). Mean square error matrix comparison of some estimatorsin linear regressions with multicollinearity. Statistics & Probability Letters, 30, pp. 133–138.

TRENKLER, G., WIJEKOON, P., (1989). Mean squared error matrix superiority of the mixed regression estimator under misspecification. Statistica, 44, pp. 65–71.

WIJEKOON, P., TRENKLER, G., (1989). Mean squared error matrix superiority of estimators under linear restrictions and misspecification. Economics Letters, 30, pp. 141–149.

ZHONG, Z., YANG, H., (2007). Ridge estimation to the restricted linear model. Communications in Statistics- Theory and Methods, 36, pp. 2099–2115.