The Research on Biased Estimators Based on Mean Square Error Matrix Criteria

Nilgün YILDIZ

Department of Mathematics, Faculty of Art-Sciences, Marmara University, 34722, Istanbul, TURKEY

Abstract

Regression models with multicollinearity can be tackled by using various estimators such as $r-(k, d)$ class estimators, principal components regression, Liu-type estimators. In this study, we defined conditions where the $r-(k, d)$ class estimator is superior over the biased estimators in terms of mean square error matrix (MSEM) criterion. Finally, we showed theoretical results by means of a numerical example and a simulation study.

Keywords: $r-(k, d)$ class estimator, Liu-type estimators, $(r-d)$ class estimator, Ordinary Ridge Regression Principal Component Regression, Mean square error matrix, Multicollinearity

I. INTRODUCTION

The ordinary least squares (OLS) estimation is one of the most commonly used methods in the literature. But this estimator is unsuitable in the presence of multicollinearity. Through the years a lot of research has been dedicated to overcome this hurdle. To overcome this problem, different remedial actions have been proposed and the most popular of those are Stein estimator [1], the principal components regression (PCR) estimator which became one of the most popular estimators proposed by Massy [2]. Another very popular estimator is ordinary Ridge Regression (ORR) estimator, which was proposed by Hoerl and Kennard [3].

While early studies have proposed new estimators to tackle the multicollinearity in later years researchers combined various estimators to obtain better results. In 1984 Baye and Parker [4] introduced $(r-k)$ class estimator, which combines the ORR and PCR. In addition, Baye and Parker also showed that $(r-k)$ class estimator is superior to PCR estimator based on the scalar mean square error (SMSE) criterion.

Liu [5] introduced a new estimator called Liu estimator (LE), which combined the Stein estimator with the ORR estimator and Akdeniz and Kaçıranlar [6] named this estimator after Liu as Liu estimator (LE). Kaçıranlar and Sakalloğlu [7] introduced the $(r-d)$ class estimator, which is a combination of the Liu and PCR estimators. Liu [8] presented the Liu-type
 estimator (LTE) by combining the ordinary ridge regression estimator and LE. Based on the MSEM criterion Özkale and Kaciulanlar [9] have compared the \((r-\theta)\) class estimator with PCR, \((r-\kappa)\) class estimator and LE.

One of the latest additions to such efforts is the \(r - (k, d)\) estimator, which is a combination of the Liu-type and PCR estimators. Inan [10], who proposed the \(r - (k, d)\) estimator. She studied the scalar mean square error (SMSE) properties of this estimator. As the MSEM is stronger criterion than the SMSE, we use the MSEM criteria. But calculation of MSEM is more difficult than SMSE. In this study, we aimed to compare the \(r - (k, d)\) estimator with OLS, PCR, ORR, LE, \((r - \kappa)\) and Liu-type estimators based on the MSEM criteria. Furthermore, necessary and sufficient conditions for the \(r - (k, d)\) class estimator to dominate the OLS, PCR and Liu-type estimators sense are derived.

A Monte Carlo simulation and and real data have been conducted to show the performance of estimators. The article is organized as follows: The second section of the article describes our model and defines \(r - (k, d)\) estimator. The MSEM criteria comparison of \(r - (k, d)\) estimator with OLS, PCR, and Liu-type estimators also takes place in the second section. The third section provides the numerical example and the outcome of the Monte Carlo simulation. Finally, some conclusions remarks are given in section four.

II. MATERIAL AND METHOD

We considered the linear regression model given as

\[ y = X\beta + \varepsilon \]  

(1)

Where \(y\) is \((n \times 1)\) observable random vector, \(X\) is a \((n \times p)\) matrix of non-stochastics variables of rank \(p\); \(\beta\) is a \((p \times 1)\) vector of unknown parameters associated with \(X\), and \(\varepsilon\) is a \((n \times 1)\) vector of error terms. Let \(T = [t_1, t_2, \ldots, t_p]\) be an orthogonal matrix that consist of the eigenvalues of \(X'X\).

\(T'X'XT = \Lambda\), where \(\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)\) ranked according to the magnitude. Further let \(T_r [t_1, t_2, \ldots, t_r]\) be remaining columns of \(T\) having deleted \(r\) columns where \(r \leq p\). Obviously, \(T_r'X'XT_r = \Lambda_r = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)\) and \(T_{p-r}'X'XT_{p-r} = \Lambda_{p-r} = \text{diag}(\lambda_{r+1}, \lambda_{r+2}, \ldots, \lambda_p)\) where \(T_{p-r} = [t_{r+1}, t_{r+2}, \ldots, t_p]\) The \(r - (k, d)\) class estimator for \(\beta\) as proposed by Inan [10] is

\[ \hat{\beta}_r(k, d) = T_r(k + dI)(T_r'X'T_r + kI)^{-1}(T_r'X'T_r - dI)T_r'T_r'\hat{\beta}_{PCR} k \geq 0, -\infty < d < +\infty \]  

(2)

where \(\hat{\beta}_{PCR} = T_r(k + dI)(T_r'X'T_r)^{-1}T_r'X'y\) is the PCR estimator.

Inan[10] proposed \(r - (k, d)\) class estimator and saw that MSE was decreased. Thus, \(r - (k, d)\) class estimator is an improved method for undertaking multicollinearity.

2.1: Main Results

In this section, we have compare the \(r - (k, d)\) class estimator with the OLS estimator, PCR and the Liu-type estimator.

**Theorem 1.** The \(\hat{\beta}_r(k, d)\) is superior to the OLS estimator \(\hat{\beta}\) if and only if

\[ \beta' T_r(k + d)^2[2(k + d)I_r + (k^2 - d^2)\Lambda_r^{-1}]^{-1}T_r' \beta + \beta' T_{p-r} \Lambda_{p-r}^{-1} T_{p-r}' \beta \leq \sigma^2 \]

**Proof.** We first compared the \(\hat{\beta}_r(k, d)\) with the OLS estimator of \(\beta\) by the MSEM .

\[ \text{MSEM}(\hat{\beta}) = \sigma^2 S^{-1} \]  

where \(S = X'X\). If we write \(\Lambda = \begin{pmatrix} \Lambda_r & 0 \\ 0 & \Lambda_{p-r} \end{pmatrix}\) and \(T_r \Lambda_r^{-1} T_r' + T_{p-r} \Lambda_{p-r}^{-1} T_{p-r}'\) then

\[ \text{MSEM}(\hat{\beta}) = \sigma^2 \begin{pmatrix} T_r \Lambda_r^{-1} T_r' + T_{p-r} \Lambda_{p-r}^{-1} T_{p-r}' \end{pmatrix} \]  

(3)
The MSEM of \( \hat{\beta}_r(k, d) \) is given as

\[
MSEM(\hat{\beta}_r(k, d)) = \sigma^2 T_r S_r(k)^{-1} S_r(-d) A_r^{-1} S_r(-d) S_r(k)^{-1} T_r' + \left[(-k - d) T_r S_r(k)^{-1} T_r' - T_{p-r} T_{p-r}' \right] \beta' \left[(-k - d) T_r S_r(k)^{-1} T_r' - T_{p-r} T_{p-r}' \right] \beta
\]

by using (3) and (4), \( MSEM(\hat{\beta}) - MSEM(\hat{\beta}_r(k, d)) \) can be expressed as

\[
MSEM(\hat{\beta}) - MSEM(\hat{\beta}_r(k, d)) = \sigma^2 T_r S_r(k)^{-1} [2(k + d) I_r + (k^2 - d^2) A_r^{-1}] S_r(k)^{-1} T_r' + T_{p-r} [\sigma^2 A_r^{-1} - T_{p-r}' \beta' T_{p-r}] T_{p-r}'
\]

Let us define \( S^*(k, d) \) and \( A^*(k, d)^{-1} \) as follows

\[
S^*(k, d) = \begin{pmatrix} \frac{1}{k + d} S_r(k) & 0 \\ 0' & I_{p-r} \end{pmatrix}
\]

and

\[
A^*(k, d)^{-1} = \begin{pmatrix} \frac{1}{(k + d)^2} [2(k + d) I_r + (k^2 - d^2) A_r^{-1}] & 0 \\ 0' & A_r^{-1} \end{pmatrix}
\]

Then the expression in (5) equals

\[
MSEM(\hat{\beta}) - MSEM(\hat{\beta}_r(k, d)) = TS^*(k, d)^{-1} [\sigma^2 A^*(k, d)^{-1} - T' \beta' T] S^*(k, d)^{-1} T'
\]

So that the \( r = (k, d) \) class estimator is superior to the OLS estimator if and only if \( \beta' T A^*(k, d) T' \beta \leq \sigma^2 \) (Rao and Toutenburg, [11]). We conclude the proof.

There are many different theorems that compare two biased estimators in terms of MSEM such as Trenkler [12], Trenkler and Toutenburg [13] and etc. We used the theorem by Baksalary and Trenkler [14].

**Theorem 2.** The \( r = (k, d) \) class estimator is superior to the PCR estimator by the criterion of MSEM if and only if \( \beta \in N(W_1) \), where \( N(W_1) \) is the null space of \( W_1 = \left( \frac{k + d}{\sigma} \right) [2(k + d) I_r + (k^2 - d^2) A_r^{-1}]^{-\frac{1}{2}} T_r' \)

**Proof.**

\[
MSEM(\hat{\beta}_r) = \sigma^2 T_r A_r^{-1} T_r' + (T_r T_r' - I_p) \beta' (T_r T_r' - I_p)
\]

\[
MSEM(\hat{\beta}_r) - MSEM(\hat{\beta}_r(k, d)) \text{ can be expressed as}
\]
Since comparing \((\hat{\beta}_r, (k, d))\) with \((\hat{\beta}_r)\) in terms of MSEM by using the matrix would be challenging, we opted to used a theorem proposed by Baksalary and Trenkler [14].

We noted from (8) that in our case \(A = \sigma^2 T_r B T_r'\), where

\[
B = S_r(k)^{-1} [2(k + d)I_r + (k^2 - d^2)A_r^{-1}] S_r(k)^{-1}
\]

\[
A^* = \frac{1}{\sigma^2} T_r B^{-1} T_r'
\]

Where \(A^*\) is Moore-Penrose inverse of \(A\) and also \(AA^* = T_r T_r'\). Hence

\[
\alpha_1 = (T_r T_r' - I_p)\beta \quad \text{and} \quad \alpha_2 = \left[-(k + d)T_r S_r(k)^{-1} T_r' - T_{p-r} T_{p-r}' \right] \beta \quad \text{if} \quad \alpha_1 \in \mathcal{N}(A) \text{ if and only if } \alpha_1 = 0.
\]

From the part (b) of theorem we can obtain the definition of \(s\):

\[
s = \frac{[a_1(U_n - AA^*)'(U_n - AA^*)a_1]}{[a_2(U_n - AA^*)'(U_n - AA^*)a_2]} = 1.
\]

In addition, \(a_2 - a_1 = A\eta_1\), where

\[
\eta_1 = -\frac{(k + d)}{\sigma^2} T_r S_r(k)[2(k + d)I_r + (k^2 - d^2)A_r^{-1}]^{-1} T_r' \beta
\]

\[
(a_2 - a_1)' A^-(a_2 - a_1) = \eta_1' A\eta_1 \leq 0
\]

Then, \((\hat{\beta}_r, (k, d))\) is superior over \((\hat{\beta}_r)\) if and only if

\[
\eta_1' A\eta_1 = \left(\frac{(k + d)}{\sigma^2}\right)^2 \beta' T_r [2(k + d)I_r + (k^2 - d^2)A_r^{-1}]^{-1} T_r' \beta \leq 0.
\]

Under the assumption \((k - d + 2)(k + d) \geq 0\),

if we let \(Q_1 = W_1 \beta Q_1 = W_1 \beta\) where

\[
W_1 = \left(\frac{k + d}{\sigma}\right) \left[2(k + d)I_r + (k^2 - d^2)A_r^{-1}\right]^{-\frac{1}{2}} T_r'
\]

\[
\text{MSEM}(\hat{\beta}_r) - \text{MSEM}(\hat{\beta}_r, (k, d)) \text{ is nd if and only if } W_1 \beta = 0 W_1 \beta = 0.
\]

**Theorem 3.** The \(r - (k, d)\) class estimator is superior to the Liu-type estimator by the criterion of MSEM if and only if \(\beta \in N(W_2)\), where \(N(W_2)\) is the null space of \(W_2 = \left(\frac{1}{\sigma}\right) A_{p-r}^{-\frac{3}{2}} T_{p-r}'\).

**Proof.**

\[
\text{MSEM}(\hat{\beta}_r, (k, d)) = \sigma^2 TS(k)^{-1} S(-d)A^{-1} S(-d)S(k)^{-1} T_r' + (k + d)^2 TS(k)^{-1} T_r' \beta \beta' TS(k)^{-1} T_r'
\]

(9)
where $S(k) = (A + kI_p)$. In consequence of (4) and (9), it can be seen that

$$
\begin{align*}
MSEM(\hat{\beta}_r(k)) - MSEM((\hat{\beta}_r(k, d)) &= \sigma^2 T_{p-r} S_{p-r}(k)^{-1} S_{p-r}(-d) A_{p-r}^{-1} S_{p-r}(-d) S_{p-r}(k)^{-1} T_{p-r}' \\
&+ (k + d)^2 T S(k)^{-1} T' \beta \beta' T S(k)^{-1} T' \\
&- [(k - d) T_r S_r(k)^{-1} T_r' - T_{p-r} \tau_{p-r}'] \eta_2 \\
&\times \beta' [(k - d) T_r S_r(k)^{-1} T_r' - T_{p-r} \tau_{p-r}']
\end{align*}
$$

Similar to the previous case, theorem by Baksalarly and Trenkler [14] was used below.

Letting $A = \sigma^2 T_{p-r} B T_{p-r}'$, where $B = S_{p-r}(k)^{-1} S_{p-r}(-d) A_{p-r}^{-1} S_{p-r}(-d) S_{p-r}(k)^{-1}$

$$
\begin{align*}
a_1 &= (k + d) T S(k)^{-1} T' \beta \quad \text{and} \quad a_2 = [(k - d) T_r S_r(k)^{-1} T_r' - T_{p-r} \tau_{p-r}] \beta.
\end{align*}
$$

$$
A^+ = \frac{1}{\sigma^2} T_{p-r} B^{-1} T_{p-r}' \quad \text{and} \quad A^+ = T_{p-r} T_{p-r}'.
\quad \alpha_1 \in \mathbb{R}(A) \quad \text{and} \quad \alpha_2 \in \mathbb{R}(A: \alpha_1) \quad \text{because} \quad s = -1,
\quad \text{where}
$$

$$
\eta_2 = -\frac{1}{\sigma^2} T_{p-r} S_{p-r}(k) S_{p-r}(-d)^{-1} A_{p-r} T_{p-r}' \beta.
$$

then the $(\hat{\beta}_r(k, d))$ dominates the $\hat{\beta}_r(k)$ if and only if

$$
\eta_2'A\eta_2 = \frac{1}{\sigma^2} \beta' T_{p-r} A_{p-r} T_{p-r}' \beta \leq 0
$$

However, it is obvious that $\eta_2'A\eta_2$ is always $\geq 0$ so the condition turns out to be $\eta_2'A\eta_2 = 0$. If we let $Q_2 = W_2 \beta$ where

$$
W_2 = \left(\frac{1}{\sigma^2}\right) (A_{p-r})^{\frac{1}{2}} T_{p-r}',
$$

we obtain $\eta_2'A\eta_2 = Q_2'Q_2$

Then $MSEM(\hat{\beta}_r(k)) - MSEM((\hat{\beta}_r(k, d))$ is nnd if and only if $Q_2 = 0; Q_2 = W_2 \beta = 0$

We can say that the necessary and sufficient conditions for the r-(k, d) class estimator to dominate the OLS, PCR and Liu-type estimators have been obtained in Theorems 1–3.

**III. RESULTS**

To motivate the problem of estimation in the linear regression model, we consider the data about Total National Research and Development Expenditures as a percent of Gross National Product by country from 1972–1986, which was discussed in Gruber [15] and Akdeniz and Erol [16].

We firstly obtain the eigenvalue of $X'X$ as

$$
\lambda_1 = 312.930, \lambda_2 = 0.7536, \lambda_3 = 0.0453, \lambda_4 = 0.0372, \lambda_5 = 0.0019
$$

The condition number indicates a serious multicollinearity among the regression vector. For the OLS, PCR, ORR, LE, LTE, $r - k$ and $r - (k, d)$ were obtained by replacing the corresponding theoretical MSE expressions in all unknown model parameters with their OLS are summarized in Table 1 and 2.

We can use the following formula to choose $k$ as given by Liu [8]

$$
\hat{k} = \frac{\lambda_1 - 100 * \lambda_5}{99}
$$
Table 1. Estimated MSE with $d=1$ and various values of $k$

<table>
<thead>
<tr>
<th></th>
<th>$k = 0.10$</th>
<th>$k = 0.30$</th>
<th>$k = 0.50$</th>
<th>$k = 0.70$</th>
<th>$k = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
</tr>
<tr>
<td><strong>PCR</strong></td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
</tr>
<tr>
<td><strong>ORR</strong></td>
<td>0.6068</td>
<td>0.6816</td>
<td>0.7058</td>
<td>0.7186</td>
<td>0.7300</td>
</tr>
<tr>
<td><strong>LE</strong></td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
</tr>
<tr>
<td><strong>LTE</strong></td>
<td>153.2568</td>
<td>22.1533</td>
<td>9.7610</td>
<td>5.9868</td>
<td>3.7802</td>
</tr>
<tr>
<td><strong>r-k</strong></td>
<td>0.7417</td>
<td>0.7435</td>
<td>0.7459</td>
<td>0.7482</td>
<td>0.7513</td>
</tr>
<tr>
<td>$r-(k,d)$</td>
<td>0.6102</td>
<td>0.6054</td>
<td>0.6023</td>
<td>0.6001</td>
<td>0.5978</td>
</tr>
</tbody>
</table>

Table 2. Estimated MSE with $\hat{\beta}_d$ and various values of $d$

<table>
<thead>
<tr>
<th></th>
<th>$d = -2$</th>
<th>$d = -1$</th>
<th>$d = 0$</th>
<th>$d = 0.30$</th>
<th>$d = 0.50$</th>
<th>$d = 0.70$</th>
<th>$d = 1$</th>
<th>$d = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
<td>0.8814</td>
</tr>
<tr>
<td><strong>PCR</strong></td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
<td>0.7417</td>
</tr>
<tr>
<td><strong>ORR</strong></td>
<td>0.7550</td>
<td>0.7550</td>
<td>0.7550</td>
<td>0.7550</td>
<td>0.7550</td>
<td>0.7550</td>
<td>0.7550</td>
<td>0.7550</td>
</tr>
<tr>
<td><strong>LE</strong></td>
<td>31.9632</td>
<td>13.5274</td>
<td>3.1667</td>
<td>1.6332</td>
<td>1.0146</td>
<td>0.7191</td>
<td>0.8814</td>
<td>6.6712</td>
</tr>
<tr>
<td><strong>LTE</strong></td>
<td>0.4579</td>
<td>0.4419</td>
<td>0.7550</td>
<td>0.9139</td>
<td>1.0363</td>
<td>1.1720</td>
<td>1.4004</td>
<td>2.377</td>
</tr>
<tr>
<td><strong>r-k</strong></td>
<td>0.7622</td>
<td>0.7622</td>
<td>0.7622</td>
<td>0.7622</td>
<td>0.7622</td>
<td>0.7622</td>
<td>0.7622</td>
<td>0.7622</td>
</tr>
<tr>
<td>$r-(k,d)$</td>
<td>0.5571</td>
<td>0.5640</td>
<td>0.5756</td>
<td>0.5800</td>
<td>0.5832</td>
<td>0.5865</td>
<td>0.5919</td>
<td>0.6130</td>
</tr>
</tbody>
</table>

We observed that under some conditions on $d$ and $k$, $r-(k,d)$ class estimator performed well compared to others. The estimated MSE values of the $r-(k,d)$ are indeed than those of the OLS, PCR, ORR, LE, r-k and LTE, which agrees with our theoretical findings described in the theorem.

3.1 Monte Carlo Simulation

In this study, the properties of $r-(k,d)$ estimator is examined by Monte Carlo simulation. $r-(k,d)$ estimator was compared with the PCR, $r-k$ estimator in terms of MSE.

According to Liu [8] and Kibria [17] the explanatory variables and response variable are generated by using the following equations:

\[ x_{ij} = (1 - \gamma^2)^{1/2}z_{ij} + \gamma z_{ip} y_i = (1 - \gamma^2)^{1/2}z_{ij} + \gamma z_{ip} y_i = (1 - \gamma^2)^{1/2}z_{ij} + \gamma z_{ip} \quad i = 1,2,\ldots,n; j = 1,2,\ldots,p \]

Where $z_{ij}$ are independent standard normal pseudo-random numbers and $\gamma$ is specified so that correlation between any two explanatory variables is given by $\gamma^2$. In this simulation, three different sets of correlations namely, $\gamma = 0.90$, $0.95$ and $0.99$ were considered to show collinearity between the explanatory variables. By applying the variance inflation factors and condition indices it can easily be shown that the explanatory variables are weak, strong and severely collinear when $\gamma = 0.90$, $0.95$ and $0.99$, respectively. In this experiment, we selected $p = 4$ for $n = 30, 50, 100$. Then, the experiment was replicated 1000 times by generating new error terms.

Let us consider the PCR, $r-k$ and $r-(k,d)$ and compute their respective estimated MSE values with the different levels of multicollinearity. Based on the simulation results shown in Tables 3, we can see that with the increase of the levels of multicollinearity, the estimated MSE values of the PCR, $r-k$ and $r-(k,d)$ increase in general. For fixed $k$ and $d$, the estimated MSE of estimators increase with the increasing level of multicollinearity. We can see that $r-(k,d)$ is much better than the competing estimators when the explanatory variables are severely collinear. The simulation results are given in Tables 3.
We have repeated the simulation studies and our results showed that for three different values of $n$ and multicollinearity there was a deviation in the MSE values. The least deviation was seen with $r - (k, \tilde{a})$. Generally speaking, the smallest MSE value calculated was seen with $r - (k, \tilde{a})$, while for $\gamma = 0.90$ and $n = 100$ the $r - (k, \tilde{a})$ showed a better performance. So the simulation results support the findings in this article.

### IV. DISCUSSION AND CONCLUSION

In this study, we have used the MSEM criterion to compare the $r - (k, \tilde{a})$ class estimator with the OLS, PCR, ORR, LE $r-k$ and Liu-type estimators. We proved that the $r - (k, \tilde{a})$ class estimator is superior overall seven aforementioned estimators in terms of mean squared error matrix under certain conditions in a numerical example. We also determined that those conditions depend on some unknown parameters. Finally, we illustrated our findings with a numerical example and a Monte Carlo simulation. Both numerical example and simulation results which agrees with our theoretical findings.
REFERENCES


