Research Article

Stochastic Restricted Biased Estimators in Misspecified Regression Model with Incomplete Prior Information

Manickavasagar Kayanan [iD] [1,2] and Pushpakanthie Wijekoon[3]

[1]*Department of Physical Science, Vavuniya Campus, University of Jaffna, Vavuniya, Sri Lanka*
[2]*Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka*
[3]*Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka*

Correspondence should be addressed to Manickavasagar Kayanan; mgayanan@vau.jfn.ac.lk

The analysis of misspecification was extended to the recently introduced stochastic restricted biased estimators when multicollinearity exists among the explanatory variables. The Stochastic Restricted Ridge Estimator (SRRE), Stochastic Restricted Almost Unbiased Ridge Estimator (SRAURE), Stochastic Restricted Liu Estimator (SRLE), Stochastic Restricted Almost Unbiased Liu Estimator (SRAULE), Stochastic Restricted Principal Component Regression Estimator (SRPCRE), Stochastic Restricted r-k (SRrk) class estimator, and Stochastic Restricted r-d (SRrd) class estimator were examined in the misspecified regression model due to missing relevant explanatory variables when incomplete prior information of the regression coefficients is available. Further, the superiority conditions between estimators and their respective predictors were obtained in the mean square error matrix (MSEM) sense. Finally, a numerical example and a Monte Carlo simulation study were used to illustrate the theoretical findings.

1. Introduction

Misspecification due to left out relevant explanatory variables is very often when considering the linear regression model, which causes these variables to become a part of the error term. Consequently, the expected value of error term of the model will not be zero. Also, the omitted variables may be correlated with the variables in the model. Therefore, one or more assumptions of the linear regression model will be violated when the model is misspecified, and hence the estimators become biased and inconsistent. Further, it is well-known that the ordinary least squares estimator (OLSE) may not be very reliable if multicollinearity exists in the linear regression model. As a remedial measure to solve multicollinearity problem, biased estimators based on the sample model $y = X\beta + \varepsilon$ with prior information which can be exact or stochastic restrictions have received much attention in the statistical literature. The intention of this work is to examine the performance of the recently introduced stochastic restricted biased estimators in the misspecified regression model with incomplete prior knowledge about

regression coefficients when there exists multicollinearity among explanatory variables.

When we consider the biased estimation in misspecified regression model without any restrictions on regression parameters, Sarkar [1] discussed the consequences of exclusion of some important explanatory variables from a linear regression model when multicollinearity exists. Şiray [2] and Wu [3] examined the efficiency of the r-d class estimator and r-k class estimator over some existing estimators, respectively, in the misspecified regression model. Chandra and Tyagi [4] studied the effect of misspecification due to the omission of relevant variables on the dominance of the r-(k, d) class estimator. Recently, Kayanan and Wijekoon [5] examined the performance of existing biased estimators and the respective predictors based on the sample information in a misspecified linear regression model without considering any prior information about regression coefficients.

It is recognized that the mixed regression estimator (MRE) introduced by Theil and Goldberger [6] outperforms ordinary least squares estimator (OLSE) when the regression

model is correctly specified. The biased estimation with stochastic linear restrictions in the misspecified regression model due to inclusion of an irrelevant variable with the incorrectly specified prior information was discussed by Teräsvirta [7]. Later Mittelhammer [8], Ohtani and Honda [9], Kadiyala [10], and Trenkler and Wijekoon [11] discussed the efficiency of MRE under misspecified regression model due to exclusion of a relevant variable with correctly specified prior information. Further, the superiority of MRE over the OLSE under the misspecified regression model with incorrectly specified sample and prior information was discussed by Wijekoon and Trenkler [12]. Hubert and Wijekoon [13] have considered the improvement of Liu estimator (LE) under a misspecified regression model with stochastic restrictions and introduced the Stochastic Restricted Liu Estimator (SRLE).

In this paper, the performance of the recently introduced stochastic restricted estimators, namely, the Stochastic Restricted Ridge Estimator (SRRE) proposed by Li and Yang [14], Stochastic Restricted Almost Unbiased Ridge Estimator (SRAURE), and Stochastic Restricted Almost Unbiased Liu Estimator (SRAULE) proposed by Wu and Yang [15], Stochastic Restricted Principal Component Regression Estimator (SRPCRE) proposed by He and Wu [16], Stochastic Restricted r-k (SRrk) class estimator, and Stochastic Restricted r-d (SRrd) class estimator proposed by Wu [17], was examined in the misspecified regression model when multicollinearity exists among explanatory variables. Further, a generalized form to represent these estimators is also proposed.

The rest of this article is organized as follows. The model specification and the estimators are written in Section 2. In Section 3, the mean square error matrix (MSEM) comparison between two estimators and respective predictors is considered. In Section 4, a numerical example and a Monte Carlo simulation study are given to illustrate the theoretical results in Scalar Mean Square Error (SMSE) criterion. Finally, some concluding remarks are mentioned in Section 5. The references and appendixes are given at the end of the paper.

2. Model Specification and the Estimators

Assume that the true regression model is given by

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon = X_1\beta_1 + \delta + \varepsilon, \tag{1}$$

where y is the $n \times 1$ vector of observations on the dependent variable, X_1 and X_2 are the $n \times l$ and $n \times p$ matrices of observations on the $m = l + p$ regressors, β_1 and β_2 are the $l \times 1$ and $p \times 1$ vectors of unknown coefficients, and ε is the $n \times 1$ vector of disturbances such that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \Omega = \sigma^2 I$.

Let us say that the researcher misspecifies the regression model by excluding p regressors as

$$y = X_1\beta_1 + u. \tag{2}$$

Let us also assume that there exists prior information on β_1 in the form of

$$r = R\beta_1 + g + v, \tag{3}$$

where r is the $q \times 1$ vector, R is the given $q \times l$ matrix with rank q, g is the $q \times 1$ unknown fixed vector, v is the $q \times 1$ vector of disturbances such that $E(v) = 0$, $D(v) = E(vv') = \Psi = \sigma^2 W$, where W is positive definite, and $E(vu') = 0$

By combining sample model (2) and prior information (3), Theil and Goldberger [6] proposed the mixed regression estimator (MRE) as

$$\begin{aligned} \widehat{\beta}_{MRE} &= \left(X_1'\Omega^{-1}X_1 + R'\Psi^{-1}R\right)^{-1}\left(X_1'\Omega^{-1}y + R'\Psi^{-1}r\right) \\ &= \left(X_1'X_1 + R'W^{-1}R\right)^{-1}\left(X_1'y + R'W^{-1}r\right). \end{aligned} \tag{4}$$

To combat multicollinearity, several researchers introduce different types of stochastic restricted estimators in place of MRE. Seven such estimators are SRRE, SRAURE, SRLE, SRALUE, SRPCRE, SRrk class estimator, and SRrd class estimator defined below, respectively:

$$\widehat{\beta}_{SRRE} = \left(X_1'X_1 + kI\right)^{-1}X_1'X_1\widehat{\beta}_{MRE}$$

$$\widehat{\beta}_{SRAURE} = \left(I - k^2\left(X_1'X_1 + kI\right)^{-2}\right)\widehat{\beta}_{MRE}$$

$$\widehat{\beta}_{SRLE} = \left(X_1'X_1 + I\right)^{-1}\left(X_1'X_1 + dI\right)\widehat{\beta}_{MRE}$$

$$\widehat{\beta}_{SRAULE} = \left(I - (1-d)^2\left(X_1'X_1 + I\right)^{-2}\right)\widehat{\beta}_{MRE} \tag{5}$$

$$\widehat{\beta}_{SRPCRE} = T_hT_h'\widehat{\beta}_{MRE}$$

$$\widehat{\beta}_{SRrk} = T_hT_h'\left(X_1'X_1 + kI\right)^{-1}X_1'X_1\widehat{\beta}_{MRE}$$

$$\widehat{\beta}_{SRrd} = T_hT_h'\left(X_1'X_1 + I\right)^{-1}\left(X_1'X_1 + dI\right)\widehat{\beta}_{MRE},$$

where $k > 0$, $0 < d < 1$, and $T_h = (t_1, t_2, \ldots, t_h)$ are the first h columns of $T = (t_1, t_2, \ldots, t_h, \ldots, t_l)$ which is an orthogonal matrix of the standardized eigenvectors of $X_1'X_1$.

According to Kadiyala [10], now we apply the simultaneous decomposition to the two symmetric matrices $X_1'X_1$ and $R'\Psi^{-1}R$, as

$$B'X_1'X_1B = I,$$

$$B'R'\Psi^{-1}RB = \Lambda, \tag{6}$$

where $X_1'X_1$ is a positive definite matrix and $R'\Psi^{-1}R$ is a positive semidefinite matrix, B is a $l \times l$ nonsingular matrix, and Λ is a $l \times l$ diagonal matrix with eigenvalues $\lambda_i > 0$ for $i = 1, 2, \ldots, q$ and $\lambda_i = 0$ for $i = q + 1, \ldots, l$.

Let $X_* = X_1B$, $R_* = RB$, $\gamma = B^{-1}\beta_1$, $X_*'X_* = I$, and $R_*'\Psi^{-1}R_* = \Lambda$; then the models (1), (2), and (3) can be written as

$$y = X_*\gamma + \delta + \varepsilon, \tag{7}$$

$$y = X_*\gamma + u, \tag{8}$$

$$r = R_*\gamma + g + v. \tag{9}$$

According to Wijekoon and Trenkler [12], the corresponding MRE is given by

$$
\begin{aligned}
\widehat{\gamma}_{\mathrm{MRE}} &= \left(X'_* X_* + R'_* \Psi^{-1} R_*\right)^{-1} \left(X'_* y + R'_* W^{-1} r\right) \\
&= \left(I + \sigma^2 \Lambda\right)^{-1} \left(X'_* y + R'_* W^{-1} r\right).
\end{aligned}
\tag{10}
$$

Hence, the respective expectation vector, bias vector, and dispersion matrix are given by

$$
E\left(\widehat{\gamma}_{\mathrm{MRE}}\right) = \gamma + \left(I + \sigma^2 \Lambda\right)^{-1} \left(X'_* \delta + R'_* W^{-1} g\right),
$$

$$
\mathrm{Bias}\left(\widehat{\gamma}_{\mathrm{MRE}}\right) = \left(I + \sigma^2 \Lambda\right)^{-1} \left(X'_* \delta + R'_* W^{-1} g\right),
\tag{11}
$$

$$
D\left(\widehat{\gamma}_{\mathrm{MRE}}\right) = \sigma^2 \left(I + \sigma^2 \Lambda\right)^{-1}.
$$

In the case of misspecification, now the SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd for model (7) can be written as

$$
\begin{aligned}
\widehat{\gamma}_{\mathrm{SRRE}} &= \left(X'_* X_* + kI\right)^{-1} X'_* X_* \widehat{\gamma}_{\mathrm{MRE}} \\
&= (1+k)^{-1} \widehat{\gamma}_{\mathrm{MRE}} = C_k \widehat{\gamma}_{\mathrm{MRE}} \\
\widehat{\gamma}_{\mathrm{SRAURE}} &= \left(I - k^2 \left(X'_* X_* + kI\right)^{-2}\right) \widehat{\gamma}_{\mathrm{MRE}} \\
&= (1+k)^{-2}(1+2k) \widehat{\gamma}_{\mathrm{MRE}} \\
&= (1+2k)\left(C_k\right)^2 \widehat{\gamma}_{\mathrm{MRE}} = C_k^* \widehat{\gamma}_{\mathrm{MRE}} \\
\widehat{\gamma}_{\mathrm{SRLE}} &= \left(X'_* X_* + I\right)^{-1} \left(X'_* X_* + dI\right) \widehat{\gamma}_{\mathrm{MRE}} \\
&= 2^{-1}(1+d) \widehat{\gamma}_{\mathrm{MRE}} = C_d \widehat{\gamma}_{\mathrm{MRE}} \\
\widehat{\gamma}_{\mathrm{SRAULE}} &= \left(I - (1-d)^2 \left(X'_* X_* + I\right)^{-2}\right) \widehat{\gamma}_{\mathrm{MRE}} \\
&= 2^{-2}(1+d)(3-d) \widehat{\gamma}_{\mathrm{MRE}} \\
&= 2^{-1}(3-d) C_d \widehat{\gamma}_{\mathrm{MRE}} = C_d^* \widehat{\gamma}_{\mathrm{MRE}} \\
\widehat{\gamma}_{\mathrm{SRPCRE}} &= T_h T'_h \widehat{\gamma}_{\mathrm{MRE}} = C_h \widehat{\gamma}_{\mathrm{MRE}} \\
\widehat{\gamma}_{\mathrm{SRrk}} &= (1+k)^{-1} T_h T'_h \widehat{\gamma}_{\mathrm{MRE}} = C_k C_h \widehat{\gamma}_{\mathrm{MRE}} \\
&= C_{hk} \widehat{\gamma}_{\mathrm{MRE}} \\
\widehat{\gamma}_{\mathrm{SRrd}} &= 2^{-1}(1+d) T_h T'_h \widehat{\gamma}_{\mathrm{MRE}} = C_d C_h \widehat{\gamma}_{\mathrm{MRE}} \\
&= C_{hd} \widehat{\gamma}_{\mathrm{MRE}},
\end{aligned}
\tag{12}
$$

respectively, where $C_k = (1+k)^{-1}$, $C_k^* = (1+2k)(C_k)^2$, $C_d = 2^{-1}(1+d)$, $C_d^* = 2^{-1}(3-d)C_d$, $C_h = T_r T'_r$, $C_{hk} = C_k C_h$, and $C_{hd} = C_d C_h$.

It is clear that C_k, C_k^*, C_d, and C_d^* are positive definite and C_h, C_{hk}, and C_{hd} are nonnegative definite.

Since all these estimators can be written by incorporating $\widehat{\gamma}_{\mathrm{MRE}}$, now we write a generalized form to represent SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd as given below:

$$
\widehat{\gamma}_{(j)} = G_{(j)} \widehat{\gamma}_{\mathrm{MRE}},
\tag{13}
$$

where $G_{(j)}$ is positive definite matrix if it stands for C_k, C_k^*, C_d, and C_d^*, and it is nonnegative definite matrix if it stands for C_h, C_{hk}, and C_{hd}.

Now the expectation vector, bias vector, the dispersion matrix, and the mean square error matrix can be written as

$$
\begin{aligned}
E\left(\widehat{\gamma}_{(j)}\right) &= G_{(j)} E\left(\widehat{\gamma}_{\mathrm{MRE}}\right) \\
&= G_{(j)} \left(\gamma + \left(I + \sigma^2 \Lambda\right)^{-1} \left(X'_* \delta + R'_* W^{-1} g\right)\right) \\
&= G_{(j)} \left(\gamma + \tau A\right) \\
\mathrm{Bias}\left(\widehat{\gamma}_{(j)}\right) &= E\left(\widehat{\gamma}_{(j)} - \gamma\right) = G_{(j)} \left(\gamma + \tau A\right) - \gamma \\
&= \left(G_{(j)} - I\right) \gamma + G_{(j)} \tau A \\
D\left(\widehat{\gamma}_{(j)}\right) &= G_{(j)} D\left(\widehat{\gamma}_{\mathrm{MRE}}\right) G'_{(j)} = \sigma^2 G_{(j)} \left(I + \sigma^2 \Lambda\right)^{-1} \\
&\quad \cdot G'_{(j)} = \sigma^2 G_{(j)} \tau G'_{(j)} \\
\mathrm{MSEM}\left(\widehat{\gamma}_{(j)}\right) &= E\left(\widehat{\gamma}_{(j)} - \gamma\right)\left(\widehat{\gamma}_{(j)} - \gamma\right)' = D\left(\widehat{\gamma}_{(j)}\right) \\
&\quad + \mathrm{Bias}\left(\widehat{\gamma}_{(j)}\right) \mathrm{Bias}\left(\widehat{\gamma}_{(j)}\right)' = \sigma^2 G_{(j)} \tau G'_{(j)} \\
&\quad + \left(\left(G_{(j)} - I\right) \gamma + G_{(j)} \tau A\right)\left(\left(G_{(j)} - I\right) \gamma + G_{(j)} \tau A\right)',
\end{aligned}
\tag{14}
$$

where $\tau = (I + \sigma^2 \Lambda)^{-1}$ and $A = (X'_* \delta + R'_* W^{-1} g)$.

Based on (14), the respective bias vector, dispersion matrix, and MSEM of the MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd can easily be obtained and are given in Table B1 in Appendix B.

By using the approach of Kadiyala [10] and (3) and (4), the generalized prediction function can be defined as follows:

$$
\begin{aligned}
y_0 &= X_* \gamma + \delta \\
\widehat{y}_{(j)} &= X_* \widehat{\gamma}_{(j)},
\end{aligned}
\tag{15}
$$

where y_0 is the actual value and $\widehat{y}_{(j)}$ is the corresponding predictor.

The MSEM of the generalized predictor is given by

$$
\begin{aligned}
\mathrm{MSEM}\left(\widehat{y}_{(j)}\right) &= E\left(\widehat{y}_{(j)} - y_0\right)\left(\widehat{y}_{(j)} - y_0\right)' \\
&= X_* \left(\mathrm{MSEM}\left(\widehat{\gamma}_{(j)}\right)\right) X'_* \\
&\quad - X_* \left(\mathrm{Bias}\left(\widehat{\gamma}_{(j)}\right)\right) \delta' \\
&\quad - \delta \left(\mathrm{Bias}\left(\widehat{\gamma}_{(j)}\right)\right)' X'_* + \delta \delta'.
\end{aligned}
\tag{16}
$$

Note that the predictors based on the MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd are denoted by $\widehat{y}_{\mathrm{MRE}}$, $\widehat{y}_{\mathrm{SRRE}}$, $\widehat{y}_{\mathrm{SRAURE}}$, $\widehat{y}_{\mathrm{SRLE}}$, $\widehat{y}_{\mathrm{SRAULE}}$, $\widehat{y}_{\mathrm{SRPCRE}}$, $\widehat{y}_{\mathrm{SRrk}}$, and $\widehat{y}_{\mathrm{SRrd}}$, respectively.

3. Mean Square Error Matrix (MSEM) Comparisons

If two generalized biased estimators $\widehat{\gamma}_{(i)}$ and $\widehat{\gamma}_{(j)}$ are given, the estimator $\widehat{\gamma}_{(j)}$ is said to be superior to $\widehat{\gamma}_{(i)}$ with respect to

MSEM sense if and only if $\mathrm{MSEM}(\widehat{\gamma}_{(i)}) - \mathrm{MSEM}(\widehat{\gamma}_{(j)}) \geq 0$. Also, if two generalized predictors $\widehat{y}_{(i)}$ and $\widehat{y}_{(j)}$ are given, the predictor $\widehat{y}_{(j)}$ is said to be superior to $\widehat{y}_{(i)}$ with respect to MSEM sense if and only if $\mathrm{MSEM}(\widehat{y}_{(i)}) - \mathrm{MSEM}(\widehat{y}_{(j)}) \geq 0$.

Now let $D_{(i,j)} = D(\widehat{\gamma}_{(i)}) - D(\widehat{\gamma}_{(j)})$, $b_{(i)} = \mathrm{Bias}(\widehat{\gamma}_{(i)})$, $b_{(j)} = \mathrm{Bias}(\widehat{\gamma}_{(j)})$, and $\Delta_{(i,j)} = \mathrm{MSEM}(\widehat{\gamma}_{(i)}) - \mathrm{MSEM}(\widehat{\gamma}_{(j)}) = D_{(i,j)} + b_{(i)}b_{(i)}' - b_{(j)}b_{(j)}'$.

By applying Lemma A1 (see Appendix A), the following theorem can be stated for the superiority of $\widehat{\gamma}_{(j)}$ over $\widehat{\gamma}_{(i)}$ with respect to the MSEM criterion.

Theorem 1. *If $D_{(i,j)}$ is positive definite, then $\widehat{\gamma}_{(j)}$ is superior to $\widehat{\gamma}_{(i)}$ in MSEM sense when the regression model is misspecified due to excluding relevant variables if and only if*

$$b_{(j)}' \left(D_{(i,j)} + b_{(i)}b_{(i)}'\right)^{-1} b_{(j)} \leq 1. \tag{17}$$

Proof. Let $D_{(i,j)}$ be a positive definite matrix. According to Lemma A1 (see Appendix A), $\Delta_{(i,j)}$ is nonnegative definite matrix if $b_{(j)}'(D_{(i,j)} + b_{(i)}b_{(i)}')^{-1}b_{(j)} \leq 1$. This completes the proof. □

The following theorem can be stated for the superiority of $\widehat{y}_{(j)}$ over $\widehat{y}_{(i)}$ with respect to the MSEM criterion.

Theorem 2. *If $A \geq 0$, $\widehat{y}_{(j)}$ is superior to $\widehat{y}_{(i)}$ in MSEM sense when the regression model is misspecified due to excluding relevant variables if and only if $\theta \in \Re(A)$ and $\theta' A^{-1}\theta \leq 1$, where $A = X_*\Delta_{(i,j)}X_*' + X_*(b_{(i)} - b_{(j)})(b_{(i)} - b_{(j)})'X_*' + \delta\delta'$, $\theta = \delta + X_*(b_{(i)} - b_{(j)})$, and $\Re(A)$ stands for column space of A and A^{-1} is an independent choice of g-inverse of A.*

Proof. According to (16), we can write $\mathrm{MSEM}(\widehat{y}_{(i)}) - \mathrm{MSEM}(\widehat{y}_{(j)})$ as

$$\begin{aligned} &\mathrm{MSEM}\left(\widehat{y}_{(i)}\right) - \mathrm{MSEM}\left(\widehat{y}_{(j)}\right) \\ &\quad = X_*\left(\mathrm{MSEM}\left(\widehat{\gamma}_{(i)}\right) - \mathrm{MSEM}\left(\widehat{\gamma}_{(j)}\right)\right)X_*' \\ &\qquad - X_*\left(\mathrm{Bias}\left(\widehat{\gamma}_{(i)}\right) - \mathrm{Bias}\left(\widehat{\gamma}_{(j)}\right)\right)\delta' \\ &\qquad - \delta\left(\mathrm{Bias}\left(\widehat{\gamma}_{(i)}\right) - \mathrm{Bias}\left(\widehat{\gamma}_{(j)}\right)\right)'X_*' \\ &\quad = X_*\Delta_{(i,j)}X_*' - X_*\left(b_{(i)} - b_{(j)}\right)\delta' \\ &\qquad - \delta\left(b_{(i)} - b_{(j)}\right)'X_*'. \end{aligned} \tag{18}$$

After some straight forward calculation, it can be written as

$$\mathrm{MSEM}\left(\widehat{y}_{(i)}\right) - \mathrm{MSEM}\left(\widehat{y}_{(j)}\right) = A - \theta\theta', \tag{19}$$

where $A = X_*(\Delta_{(i,j)} + (b_{(i)} - b_{(j)})(b_{(i)} - b_{(j)})')X_*' + \delta\delta'$ and $\theta = \delta + X_*(b_{(i)} - b_{(j)})$.

Due to Lemma A3 (see Appendix A), $\mathrm{MSEM}(\widehat{y}_{(i)}) - \mathrm{MSEM}(\widehat{y}_{(j)})$ is nonnegative definite matrix if and only if $A \geq 0$, $\theta \in \Re(A)$ and $\theta' A^{-1}\theta \leq 1$, where $\Re(A)$ stands for column space of A and A^{-1} is an independent choice of g-inverse of A. This completes the proof. □

Based on Theorems 1 and 2, we can define Corollaries C1–C28, written in Appendix C, for the superiority conditions between two selected estimators and for the respective predictors by substituting the relevant expressions for $\mathrm{Bias}(\widehat{\gamma}_{(i)})$, $\mathrm{Bias}(\widehat{\gamma}_{(j)})$, $D(\widehat{\gamma}_{(i)})$, and $D(\widehat{\gamma}_{(j)})$ given in Table B1 in Appendix B.

4. Illustration of Theoretical Results

4.1. Numerical Example. To illustrate the theoretical results, we considered the dataset which gives the total National Research and Development Expenditures as a Percent of Gross National Product by Country from 1972 to 1986. The dependent variable Y of this dataset is the percentage spent by the United States, and the four other independent variables are X_1, X_2, X_3, and X_4. The variable X_1 represents the percent spent by the former Soviet Union, X_2 that spent by France, X_3 that spent by West Germany, and X_4 that spent by the Japan. The data has been analysed by Gruber [18], Akdeniz and Erol [19], and Li and Yang [14], among others. Now we assemble the data as follows:

$$X = \begin{pmatrix} 1.9 & 2.2 & 1.9 & 3.7 \\ 1.8 & 2.2 & 2.0 & 3.8 \\ 1.8 & 2.4 & 2.1 & 3.6 \\ 1.8 & 2.4 & 2.2 & 3.8 \\ 2.0 & 2.5 & 2.3 & 3.8 \\ 2.1 & 2.6 & 2.4 & 3.7 \\ 2.1 & 2.6 & 2.6 & 3.8 \\ 2.2 & 2.6 & 2.6 & 4.0 \\ 2.3 & 2.8 & 2.8 & 3.7 \\ 2.3 & 2.7 & 2.8 & 3.8 \end{pmatrix}$$

$$y = \begin{pmatrix} 2.3 \\ 2.2 \\ 2.2 \\ 2.3 \\ 2.4 \\ 2.5 \\ 2.6 \\ 2.6 \\ 2.7 \\ 2.7 \end{pmatrix}. \tag{20}$$

Note that the eigenvalues of $X'X$ are 302.96, 0.728, 0.044, and 0.035, the condition number is 93, and the Variance Inflation Factor (VIF) values are 6.91, 21.58, 29.75, and 1.79. This implies the existence of serious multicollinearity in the dataset.

The corresponding OLS estimator of β is $\widehat{\beta} = (X'X)^{-1}X'y = (0.645, 0.089, 0.143, 0.152)$ and the estimate of σ^2 is $\widehat{\sigma}^2 = 0.00153$. In this example, we consider

$R = (1, -2, -2, -2)$ and $g = c(1, -1, 2, 0)$. The SMSE values of the estimators are summarized in Tables B2-B3 in Appendix B.

Table B2 shows the estimated SMSE values of MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd for the regression model when $(l, p) = (4, 0)$, $(l, p) = (3, 1)$, and $(l, p) = (2, 2)$ with respect to shrinkage parameters (k/d), where l denotes the number of variables in the model and p denotes the number of misspecified variables. Table B3 shows the estimated SMSE values of the predictor of MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd for the regression model when $(l, p) = (4, 0)$, $(l, p) = (3, 1)$, and $(l, p) = (2, 2)$ for some selected shrinkage parameters (k/d).

Note that when $(l, p) = (4, 0)$ the model is correctly specified, when $(l, p) = (3, 1)$ one variable is omitted from the model, and when $(l, p) = (2, 2)$ two variables are omitted from the model. For simplicity, we choose shrinkage parameter values k and d in the range $(0, 1)$.

From Table B2, we can observe that the MRE is superior to the other estimators when $(l, p) = (4, 0)$ and SRAULE, SRRE, SRLE, and SRAURE outperform the other estimators for $(k/d) < 0.2$, $0.2 \leq (k/d) < 0.5$, $0.5 \leq (k/d) < 0.7$, and $(k/d) \geq 0.7$, respectively, when $(l, p) = (3, 1)$. Similarly, SRLE and SRRE are superior to the other estimators for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (2, 2)$.

From Table B3, we further observe that predictors based on SRLE and SRRE outperform the other predictors for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (4, 0)$ and $(l, p) = (3, 1)$, and predictors based on SRrd and SRrk are superior to the other predictors for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (2, 2)$.

4.2. Simulation. For further clarification, a Monte Carlo simulation study is done at different levels of misspecification using R 3.2.5. Following McDonald and Galarneau [20], we can generate the explanatory variables as follows:

$$x_{ij} = \left(1 - \rho^2\right)^{1/2} z_{ij} + \rho z_{i,m};$$
$$i = 1, 2, \ldots, n. \ j = 1, 2, \ldots, m, \tag{21}$$

where z_{ij} is an independent standard normal pseudorandom number and ρ is specified so that the theoretical correlation between any two explanatory variables is given by ρ^2. A dependent variable is generated by using the following equation:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i;$$
$$i = 1, 2, \ldots, n, \tag{22}$$

where ε_i is a normal pseudorandom number with mean zero and variance one. Also, we select $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ as the normalized eigenvector corresponding to the largest eigenvalue of $X'X$ for which $\beta'\beta = 1$. Further we choose $R = (1, 1, 1, 1, 1)$ and $g = (1, -2, 0, 3, 1)$.

Then the following setup is considered to investigate the effects of different degrees of multicollinearity on the estimators:

(i) $\rho = 0.9$, condition number $= 9.49$, and VIF $= (5.99, 5.88, 5.94, 5.96, 20.47)$.

(ii) $\rho = 0.99$, condition number $= 34.77$, and VIF $= (57.66, 56.50, 57.26, 57.31, 225.06)$.

(iii) $\rho = 0.999$, condition number $= 115.66$, and VIF $= (574.3, 562.8, 570.7, 570.8, 2271.4)$.

Three different sets of observations are considered by selecting $(l, p) = (5, 0)$, $(l, p) = (4, 1)$, and $(l, p) = (3, 2)$ when $n = 50$, where l denotes the number of variables in the model and p denotes the number of misspecified variables. Note that when $(l, p) = (5, 0)$ the model is correctly specified, when $(l, p) = (4, 1)$ one variable is omitted from the model, and when $(l, p) = (3, 2)$ two variables are omitted from the model. For simplicity, we select values k and d in the range $(0, 1)$.

The simulation is repeated 2000 times by generating new pseudorandom numbers and the simulated SMSE values of the estimators and predictors are obtained using the following equations:

$$\text{SMSE}\left(\widehat{\gamma}_{(j)}\right) = \frac{1}{2000} \sum_{r=1}^{2000} \text{tr}\left(\text{MSEM}\left(\widehat{\gamma}_{(j)r}\right)\right),$$

$$\text{SMSE}\left(\widehat{y}_{(j)}\right) \tag{23}$$

$$= \frac{1}{2000} \sum_{r=1}^{2000} \text{tr}\left(\text{MSEM}\left(\widehat{y}_{(j)r}\right)\right) \text{ respectively.}$$

The simulation results are summarized in Tables B4–B9 in Appendix B.

Tables B4, B5, and B6 show the estimated SMSE values of the estimators for the regression model when $(l, p) = (5, 0)$, $(l, p) = (4, 1)$, and $(l, p) = (3, 2)$ and $\rho = 0.9$, $\rho = 0.99$, and $\rho = 0.999$ for the selected values of shrinkage parameters (k/d), respectively. Tables B7, B8, and B9 show the corresponding estimated SMSE values of the predictors for the above regression models, respectively.

From Table B4, we can observe that MRE and SRAULE outperform the other estimators for $(k/d) < 0.8$ and $(k/d) \geq 0.8$, respectively, when $(l, p) = (5, 0)$ and $(l, p) = (4, 1)$. Further, SRLE and SRRE are superior to the other estimators for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (3, 2)$ under $\rho = 0.9$.

From Table B5, we can observe that SRAULE, MRE, and SRAURE outperform the other estimators for $(k/d) < 0.3$, $0.3 \leq (k/d) < 0.7$, and $(k/d) \geq 0.7$, respectively, when $(l, p) = (5, 0)$. Similarly, SRAULE, SRRE, SRLE, and SRAURE are superior to the other estimators when $(k/d) < 0.2$, $0.2 \leq (k/d) < 0.5$, $0.5 \leq (k/d) < 0.7$, and $(k/d) \geq 0.7$, respectively, when $(l, p) = (4, 1)$, and both SRLE and SRRE outperform the other estimators for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (3, 2)$ and $\rho = 0.99$.

The results in Table B6 indicate that MRE is superior to the other estimators when $(l, p) = (5, 0)$, and SRAULE, SRRE, SRLE, and SRAURE outperform the other estimators for $(k/d) < 0.2$, $0.2 \leq (k/d) < 0.5$, $0.5 \leq (k/d) < 0.7$, and $(k/d) \geq 0.7$, respectively, when $(l, p) = (4, 1)$. Further, SRLE and SRRE outperform the other estimators for $(k/d) < 0.5$
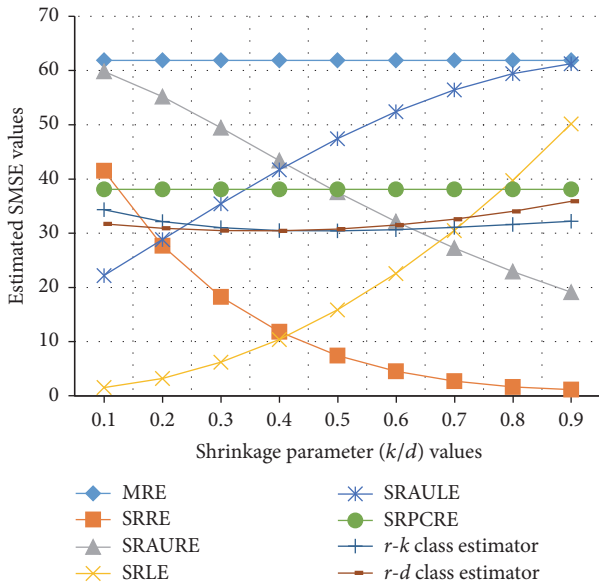
FIGURE 1: SMSE values of the estimators in the misspecified regression model $((l, p) = (3, 2))$ when $n = 50$ and $\rho = 0.9$.
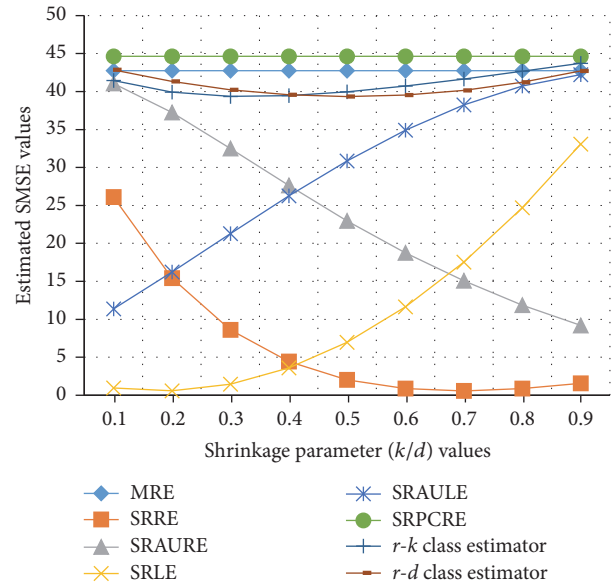

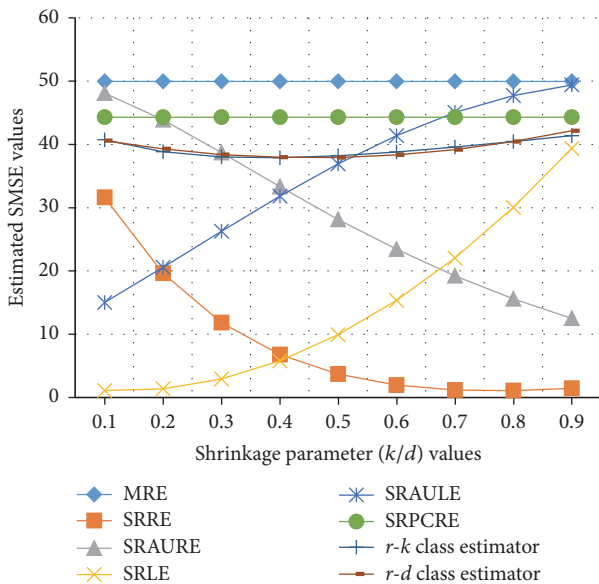FIGURE 2: SMSE values of the estimators in the misspecified regression model when $n = 50$ and $\rho = 0.99$.

FIGURE 3: SMSE values of the estimators in the misspecified regression model when $n = 50$ and $\rho = 0.999$.
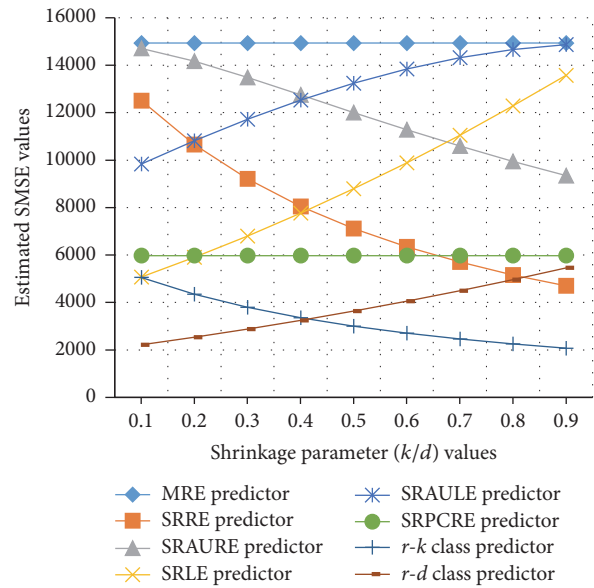

FIGURE 4: SMSE values of the predictors in the misspecified regression model when $n = 50$ and $\rho = 0.9$.

and $(k/d) \geq 0.5$, respectively, when $(l, p) = (3, 2)$ and $\rho = 0.999$.

From Tables B7–B9, we further observe that the predictors based on SRrd and SRrk always outperform the other predictors for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (5, 0)$, $(l, p) = (4, 1)$, and $(l, p) = (3, 2)$.

The SMSE values of the selected estimators are plotted with different ρ values to demonstrate the results graphically when $(l, p) = (3, 2)$. Figures 1–3 show the graphical illustration of the performance of estimators in the misspecified regression model $((l, p) = (3, 2))$ when $\rho = 0.9$, $\rho = 0.99$,

and $\rho = 0.999$, respectively. Similarly, Figures 4–6 present the graphical illustration of the performance of predictors in the misspecified regression model $((l, p) = (3, 2))$ when $\rho = 0.9$, $\rho = 0.99$, and $\rho = 0.999$, respectively.

5. Conclusion

Theorems 1 and 2 give the common form of superiority conditions to compare the estimators (MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd) and their respective predictors in MSEM criterion in the misspecified linear
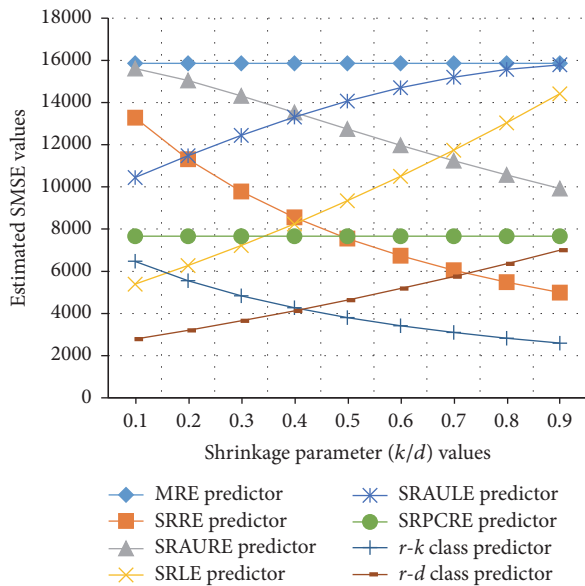
FIGURE 5: SMSE values of the predictors in the misspecified regression model when $n = 50$ and $\rho = 0.99$.
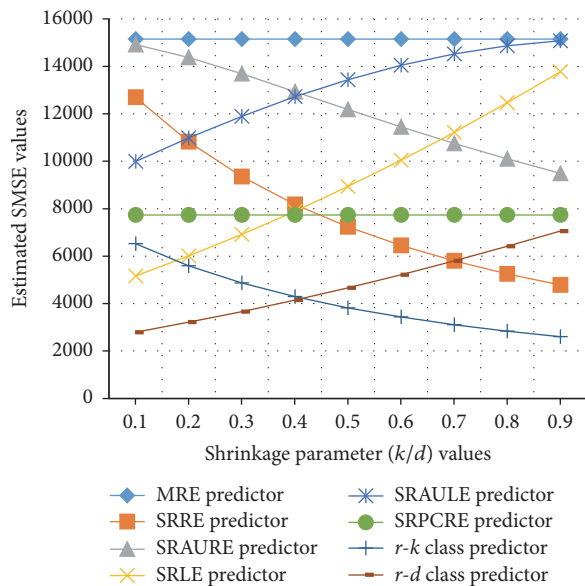

FIGURE 6: SMSE values of the predictors in the misspecified regression model when $n = 50$ and $\rho = 0.999$.

regression model when the prior information of the regression coefficients is incomplete, and the multicollinearity exists among the explanatory variables.

From the simulation study, the superior estimators and predictors over the others when the conditions are different can be identified. The results obtained in this research will produce significant improvements in the parameter estimation in misspecified regression models with incomplete prior information, and the results are applicable to real-world applications.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

Supplementary material contains three appendix sections named Appendices A, B, and C. Appendix A: Lemmas A1–A3, used to prove the theorems. Appendix B: Tables B1–B9, which show the stochastic properties of estimators and results of numerical example and simulation study. Appendix C: Corollaries C1–C28. *(Supplementary Materials)*

References

[1] N. Sarkar, "Comparisons among some estimators in misspecified linear models with multicollinearity," *Annals of the Institute of Statistical Mathematics*, vol. 41, no. 4, pp. 717–724, 1989.

[2] G. Şiray, "r-d class estimator under misspecification," *Communications in Statistics—Theory and Methods*, vol. 44, no. 22, pp. 4742–4756, 2015.

[3] J. Wu, "Superiority of the r-k class estimator over some estimators in a misspecified linear model," *Communications in Statistics—Theory and Methods*, vol. 45, no. 5, pp. 1453–1458, 2016.

[4] S. Chandra and G. Tyagi, "On the performance of some biased estimators in a misspecified model with correlated regressors," in *STATISTICS IN TRANSITION new series*, pp. 27–52, 2017.

[5] M. Kayanan and P. Wijekoon, "Performance of Existing Biased Estimators and the respective Predictors in a Misspecified Linear Regression Model," *Open Journal of Statistics*, pp. 876–900, 2017.

[6] H. Theil and A. S. Goldberger, "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, vol. 2, no. 1, pp. 65–78, 1961.

[7] T. Teräsvirta, *Linear restrtctions in misspecified linear models and polynomial distributed lag estimation*, vol. 5, Departntent of Statistics University of Helsinki, Helsinki, Finland, 1980.

[8] R. C. Mittelhammer, "On specification error in the general linear model and weak mean square error superiority of the mixed estimator," *Communications in Statistics—Theory and Methods*, vol. 10, no. 2, pp. 167–176, 1981.

[9] K. Ohtani and Y. Honda, "On small sample properties of the mixed regression predictor under misspecification," *Communications in Statistics - Theory and Methods*, pp. 2817–2825, 1984.

[10] K. Kadiyala, "Mixed regression estimator under misspecification," *Economics Letters*, vol. 21, no. 1, pp. 27–30, 1986.

[11] G. Trenkler and P. Wijekoon, "Mean square error matrix superiority of the mixed regression estimator under misspecification," *Statistica*, vol. 49, no. 1, pp. 65–71, 1989.

[12] P. Wijekoon and G. Trenkler, "Mean Square Error Matrix Superiority of Estimators under Linear Restrictions and Misspecification," *Economics Letters*, vol. 30, pp. 141–149, 1989.

[13] M. Hubert and P. Wijekoon, "Superiority of the stochastic restricted Liu estimator under misspecification," *Statistica*, vol. 64, no. 1, pp. 153–162, 2004.

[14] Y. Li and H. Yang, "A new stochastic mixed ridge estimator in linear regression model," *Statistical Papers*, pp. 315–323, 2010.

[15] J. Wu and H. Yang, "On the stochastic restricted almost unbiased estimators in linear regression model," *Communications in*

Statistics - Simulation and Computation, vol. 43, no. 2, pp. 428–440, 2014.

[16] D. He and Y. Wu, "A Stochastic Restricted Principal Components Regression Estimator in the Linear Model," *The Scientific World Journal*, vol. 2014, Article ID 231506, 6 pages, 2014.

[17] J. Wu, "On the Stochastic Restricted r-k Class Estimator and Stochastic Restricted r-d Class Estimator in Linear Regression Model," *Journal of Applied Mathematics*, vol. 2014, Article ID 173836, 6 pages, 2014.

[18] M. H. Gruber, *Improving Efficiency by Shrinkage*, vol. 156 of *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, NY, USA, 1998.

[19] F. Akdeniz and H. Erol, "Mean Squared Error Matrix Comparisons of Some Biased Estimators in Linear Regression," *Communications in Statistics—Theory and Methods*, pp. 2389–2413, 2003.

[20] G. C. McDonald and D. I. Galarneau, "A monte carlo evaluation of some ridge-type estimators," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 407–416, 1975.

Journal of
Applied Mathematics

The Scientific
World Journal

Journal of
Probability and Statistics

Advances in
Operations Research

Advances in
Decision Sciences

International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Optimization

International Journal of
Engineering
Mathematics

Hindawi

Submit your manuscripts at
www.hindawi.com

International Journal of
Analysis

Journal of
Complex Analysis

Advances in
Numerical Analysis

Mathematical Problems
in Engineering

International Journal of
Differential Equations

Discrete Dynamics in
Nature and Society

International Journal of
Stochastic Analysis

Journal of
Mathematics

Journal of
Function Spaces

Abstract and
Applied Analysis

Advances in
Mathematical Physics