Modern Management based on Big Data V
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA240239

On the Random Constrained Dual Parameter Ridge Estimator in Linear Regression

Lipei Liu[a,1]
[a] *School of Mathematics and Big Data, Chongqing University of Arts and Sciences, Chongqing, China, 402160*
ORCiD ID: https://orcid.org/0009-0004-7686-8786

Abstract. This article proposes a random constrained dual parameter ridge estimator with random constraints for the problem of multicollinearity in linear regression model. Under the mean square error matrix, the superiority of the new estimator over the dual parameter ridge estimator, common mixed estimator and random constrained ridge estimator is discussed. And the theoretical results are demonstrated through a numerical example and a simulation study.

Keywords. Common mixed estimator; Ridge estimator; Random constrained dual parameter estimator

1. Introduction

In a linear regression, the ordinary least square estimator (OLSE) is unbiased and has the smallest variance among all linear unbiased estimators. It has long been considered the best estimator. However, when there is infamous multicollinearity, OLSE may be highly variable, despite its minimum variance natures in linear unbiased estimation classes. Many statisticians think how to enhance OLSE. One approach is to use biased estimators such as Hoerl and Kennard [1], Liu [2], Yang and Chang [3], Wu and Yang [4], Akdeniz and Erol [5], and others.

Another method for dealing with multicollinearity problems is to consider parameter estimators with sample information. If the sample information is linearly constrained, Zhong and Yang [6] presented a constrained ridge estimator, Ozkale and Kaciranlar [7] proposed a constrained dual parameter estimator, and Wu [8] proposed a constrained r-d class estimator. When the sample information is subject to random constraints, Theil and Goldberger [9] proposed a common mixed estimator. Li and Yang [10] propose a stochastic constrained ridge estimator, Li and Yang [11] proposed a stochastic constrained dual parameter estimator, and Alheety [12] proposed a stochastic constrained ridge estimator.

For this article, when the random linear constraint is established, we combine the dual parameter ridge estimator and the mixed estimator to propose a dual parameter ridge

[1]Corresponding Author: Lipei Liu, Chongqing University of Arts and Sciences, Chongqing, China; Email: cqwu_huawei@outlook.com. The authors are highly obliged to the editor and the reviewers for the comments and suggestions which improved the paper in its present form. This work is supported by the Scientific Technological Research Program of Chongqing Municipal Education Commission [Grant No. KJQN202001321].

estimator with random constraints. We discussed the performance of the new estimator relative to other competing estimators regarding the mean square error matrix criterion. The new estimator has many advantages. Firstly, the new estimator is a general estimator which contains many existing estimators. Secondly, under the mean square error matrix criterion the new estimator is better than other existing estimators.

The remaining parts of this article are arranged as follows: In Section 2, we described the model and proposed new estimator. Section 3 compares the new estimator with the competing estimator based on the mean square error matrix criterion. In Section 4, the advantages of new estimator is demonstrated through an actual data example an a simulation study. Section 5 introduces some discussions.

2. The new estimator

Consider the following linear model

$$y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2) \tag{1}$$

where y is an $m \times 1$ vector of dependent variable, X is an $m \times q$ matrix of the control variables, β is a $q \times 1$ vector of unknown regression coefficients.

The ordinary least square estimator is often used to estimate regression coefficients

$$\hat{\beta} = (X'X)^{-1}X'y \tag{2}$$

To overcome multicollinearity, Hoerl and Kennard [1] proposed the ridge estimator (RE), which is defined as follows:

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'y \tag{3}$$

where $k > 0$, I shows an identity matrix.

The dual parameter ridge estimator (DPRE) [7] is defined as

$$\hat{\beta}(k, d) = T_{kd}\hat{\beta} \tag{4}$$

where $T_{kd} = SS_{kd}^{-1}, S = X'X, S_{kd} = X'X + kdI, k > 0, 0 < d < 1$.

Let's provide some prior information in the form of a set of n independent random linear constraints, as follows:

$$r = R\beta + e \tag{5}$$

where R is a $n \times q$ known matrix with $rank(R) = n$, e is a $q \times 1$ vector of disturbances with expectation 0 and covariance matrix $\sigma^2 W$, $W > 0$.

With the constraint model (1) and (5), the common mixed estimator (CME) introduced by Theil and Goldberger [9] is defined as

$$\hat{\beta}_{CME} = (X'X + R'W^{-1}R)^{-1}(X'y + R'W^{-1}r) \tag{6}$$

Li and Yang [10] presented the random constraint ridge estimator (RCRE)

$$\hat{\beta}_{RCRE} = (X'X + R'W^{-1}R)^{-1}(T_k X'y + R'W^{-1}r) \tag{7}$$

where $T_k = SS_k^{-1}$, where $S_k = X'X + kI$.

Now, we are preparing to propose a new constraint estimator in a random constrained linear regression model. Then combining CME with DPRE, we can propose a random constrained dual parameter ridge estimator (RCDPRE) as follows:

$$\hat{\beta}_{RCDPRE}(k, d) = (X'X + R'W^{-1}R)^{-1}(T_{kd}X'y + R'W^{-1}r) \tag{8}$$

By the definitions of $\hat{\beta}_{RCDPRE}(k, d)$ that is a general estimator that include the CME, RCRE, DPRE as special cases.

3. The properties of RCDPRE

For this section, we will compare the new estimator with CME, RCRE and DPRE under the mean square error matrix. We provide the definition of the mean square error matrix (MSEM). The MSEM of one estimator is denoted as

$$MSEM(\beta^*) = E[(\beta^* - \beta)(\beta^* - \beta)'] \tag{9}$$

And (9) is equivalent to

$$MSEM(\beta^*) = Cov(\beta^*) + [Bias(\beta^*)][Bias(\beta^*)]' \tag{10}$$

Where $Bias(\beta^*) = E(\beta b^*) - \beta$ is the bias estimator. For two given estimators $\hat{\beta}_1, \hat{\beta}_2$, the estimator $\hat{\beta}_2$ is said to be superior to $\hat{\beta}_1$ in the MSEM criterion, if and only if

$$\Delta(\hat{\beta}_1, \hat{\beta}_2) = MSEM(\hat{\beta}_1) - MSEM(\hat{\beta}_2) \geq 0$$

Now we present some lemmas.

Lemma 1 (Farebrother [13]) Let $M > 0$ and α be some vector, the $M - \alpha\alpha' \geq 0 \Leftrightarrow$ if $\alpha' M^{-1}\alpha \leq 1$.

Lemma 2 (Rao et al. [14]) Let $M > 0, N > 0$,then $M > N \Leftrightarrow \lambda_1(NM^{-1}) \leq 1$, where $\lambda_1(NM^{-1})$ denotes the maximum eigenvalues of NM^{-1}.

Firstly, we can compute the MSEM of the estimators CME, RCRE, DPRE and RCDPRE as follows:

$$MSEM(\hat{\beta}_{DPRE}) = \sigma^2 T_{kd} S^{-1} T_{kd} + b_1 b_1' \tag{11}$$

$$MSEM(\hat{\beta}_{CME}) = \sigma^2 A \tag{12}$$

$$MSEM(\hat{\beta}_{RCRE}) = \sigma^2 A(T_k S T_k + R'W^{-1}R)A + b_2 b_2' \tag{13}$$

$$MSEM(\hat{\beta}_{RCDPRE}) = \sigma^2 A(T_{kd} S T_{kd} + R'W^{-1}R)A + b_3 b_3' \tag{14}$$

where $A = (X'X + R'W^{-1}R)^{-1}$, $b_1 = (T_{kd} - I)\beta$, $b_2 = A(T_k - I)\beta$, $b_3 = A(T_{kd} - I)\beta$.

Now we study the following differences:

$$\begin{aligned}\Delta_1 &= MSEM(\hat{\beta}_{DPRE}) - MSEM(\hat{\beta}_{RCDPRE}) \\ &= \sigma^2 T_{kd} S^{-1} T_{kd} + b_1 b_1' - \sigma^2 A(T_{kd} S T_{kd} + R'W^{-1}R)A - b_3 b_3' \\ &= \sigma^2 D_1 + b_1 b_1' - b_3 b_3' \end{aligned} \tag{15}$$

$$\begin{aligned}\Delta_2 &= MSEM(\hat{\beta}_{CME}) - MSEM(\hat{\beta}_{RCDPRE}) \\ &= \sigma^2 A - \sigma^2 A(T_{kd} S T_{kd} + R'W^{-1}R)A - b_3 b_3' \\ &= \sigma^2 D_2 - b_3 b_3' \end{aligned} \tag{16}$$

$$\begin{aligned}\Delta_3 &= MSEM(\hat{\beta}_{RCRE}) - MSEM(\hat{\beta}_{RCDPRE}) \\ &= \sigma^2 A(T_k S T_k + R'W^{-1}R)A + b_2 b_2' - \sigma^2 A(T_{kd} S T_{kd} + R'W^{-1}R)A - b_3 b_3' \\ &= \sigma^2 D_3 + b_2 b_2' - b_3 b_3' \end{aligned} \tag{17}$$

where $D_1 = T_{kd} S^{-1} T_{kd} - A(T_{kd} S T_{kd} + R'W^{-1}R)A$, $D_2 = A - A(T_{kd} S T_{kd} + R'W^{-1}R)A$, $D_3 = A(T_k S T_k + R'W^{-1}R)A$-$A(T_{kd} S T_{kd} + R'W^{-1}R)A$.

Below we will provide their comparison.

Theorem 1 When $\lambda_1(A(T_{kd} S T_{kd} + R'W^{-1}R)A(T_{kd} S^{-1} T_{kd})^{-1}) < 1$, the RCDPRE is better than the DPRE in the MSEM sense, if $b_3'(\sigma^2 D_1 + b_1 b_1')b_3 \leq 1$.

Proof Since $T_{kd} S^{-1} T_{kd} > 0$, $A(T_{kd} S T_{kd} + R'W^{-1}R)A > 0$. When $\lambda_1(A(T_{kd} S T_{kd} + R'W^{-1}R)A(T_{kd} S^{-1} T_{kd})^{-1}) < 1$, by Lemma 2, we can see that $D_1 > 0$, so from (15) and Lemma 1, we get that $\Delta_1 \geq 0$ if and only if $b_2'(\sigma^2 D_1 + b_1 b_1')b_2 \leq 1$.

Theorem 2 If $b_3' D_2^{-1} b_3 \leq \sigma^2$,the RCDPRE is better than the CME in the MSEM sense.

Proof We can that $D_2 = A - A(T_{kd} S T_{kd} + R'W^{-1}R)A = A(S - T_{kd} S T_{kd})A$. As $S > 0$,we can get that $S = HTH'$,where $T = diag(\eta_{1,\dots,}\eta_q)$ and $\eta_i > 0$. Thus, we have

$$S - T_{kd}ST_{kd} = Hdiag(\frac{kd(kd+2\eta_i)\eta_i}{(kd+\eta_i)^2})H'$$

So we have $D_2 > 0$, that is the RCDPRE is superior to the CME if and only if $b_3'D_2^{-1}b_3 \leq \sigma^2$.

Theorem 3 If $b_3'(\sigma^2 D_3 + b_2 b_2')b_3 \leq 1$, the RCDPRE is better than the RCRE in the MSEM sense.

Proof We can that $D_3 = A(T_kST_k + R'W^{-1}R)A - A(T_{kd}ST_{kd} + R'W^{-1}R)A = A(T_kST_k - AT_{kd}ST_{kd})A$. By theorem 2, we have

$$T_kST_k - T_{kd}ST_{kd} = Hdiag(\frac{kd(kd+2\eta_i+k)\eta_i}{(kd+\eta_i)^2(k+\eta_i)^2})H'$$

So we have $D_3 > 0$, that is the RCDPRE is superior to the RCRE if and only if $b_3'(\sigma^2 D_3 + b_2 b_2')b_3 \leq 1$.

4. Numerical example and Simulation study

4.1 A numerical example

To explain our theoretical results, we have now studied in this section the dataset originally provided by Gruber [15] and later discussed by Akdeniz and Erol [5]. In this article, we studied the same data in an attempt to demonstrate that RCDPRE outperforms CME, RCRE and DPRE.

With this data, we know that $\hat{\beta} = (0.6455,0.0896,0.1436,0.1536)'$, with $MSEM(\hat{\beta}) = 0.0808, \sigma_{OLSE}^2 = 0.0015$.

Consider the following stochastic restrictions: $r = R\beta + e, R = (1,-2,-2,-2), e \sim N(0,\sigma^2)$. Then we can get the MSE of CME, RCRE, DPRE and RCDPRE.

Table 1. MSE values of CME, RCRE and DPRE and RCDPRE for d=0.1

k	0	0.001	0.002	0.003	0.005	0.01	0.02	0.05
DPRE	0.0808	0.0769	0.0736	0.0708	0.0662	0.0601	0.0601	0.0857
CME	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454
RCRE	0.0454	0.0439	0.0440	0.0452	0.0520	0.0672	0.0969	0.1160
RCDPRE	0.0454	0.0452	0.0450	0.0448	0.0444	0.0439	0.0440	0.0500

Table 2. MSE values of CME, RCRE and DPRE and RCDPRE for d=0.2

k	0	0.001	0.002	0.003	0.005	0.01	0.02	0.05
DPRE	0.0808	0.0799	0.0792	0.0784	0.0769	0.0736	0.0683	0.0601
CME	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454
RCRE	0.0454	0.0439	0.0440	0.0452	0.0520	0.0672	0.0969	0.1160
RCDPRE	0.0454	0.0450	0.0446	0.0443	0.0439	0.0440	0.0473	0.0672

Table 3. MSE values of CME, RCRE and DPRE and RCDPRE for d=0.4

k	0	0.001	0.002	0.003	0.005	0.01	0.02	0.05
DPRE	0.0808	0.0792	0.0776	0.0762	0.0736	0.0683	0.0618	0.0601
CME	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454	0.0454
RCRE	0.0454	0.0439	0.0440	0.0452	0.0520	0.0672	0.0969	0.1160
RCDPRE	0.0454	0.0445	0.0436	0.0428	0.0414	0.0388	0.0368	0.0437

With Tables 1-3, we get that in most cases the new estimator is better than the CME, RCRE and DPRE. And we also can see that with small d, the new estimator performs well. So, we can see our new estimator is meaningful in practice.

4.2 A simulation study

In this subsection, we give a simulation study to show the performance of the new estimator. Following Akdeniz and Erol [5], we use the following formula to generate correlated variables:

$$x_{ij} = (1-\theta^2)^{0.5} z_{ij} + \theta z_{ip}$$

In this paper, we can consider $\theta = 0.99, n = 50,100,150, q = 4, \sigma^2=0.05$. We list the simulation results in Tables 4-6.

By the simulation, we find that the new estimator is always better than DPRE and RCRE. In most cases, the new estimator is better than CME. With large sample size n, the new estimator performs well.

Table 4. MSE values of CME, RCRE and DPRE and RCDPRE for n *=50*

k	0	0.001	0.002	0.003	0.005	0.01	0.02	0.05
DPRE	0.152	0.152	0.152	0.153	0.154	0.163	0.201	0.457
CME	0.0898	0.0898	0.0898	0.0898	0.0898	0.0898	0.0898	0.0898
RCRE	0.0898	0.0897	0.0898	0.0900	0.0910	0.0961	0.1174	0.2626
RCDPRE	0.0898	0.0897	0.0897	0.0897	0.0897	0.0898	0.0904	0.0961

Table 5. MSE values of CME, RCRE and DPRE and RCDPRE for *n=100*

k	0	0.001	0.002	0.003	0.005	0.01	0.02	0.05
DPRE	0.0757	0.0756	0.0756	0.0756	0.0757	0.0766	0.0811	0.1139
CME	0.0517	0.0517	0.0517	0.0517	0.0517	0.0517	0.0517	0.0517
RCRE	0.0517	0.0516	0.0516	0.0515	0.0515	0.0515	0.0520	0.0578
RCDPRE	0.0517	0.0516	0.0516	0.0515	0.0515	0.0515	0.0515	0.0515

Table 6. MSE values of CME, RCRE and DPRE and RCDPRE for *n=150*

k	0	0.001	0.002	0.003	0.005	0.01	0.02	0.05
DPRE	0.0600	0.0600	0.0600	0.0600	0.0601	0.0609	0.0644	0.0906
CME	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415	0.0415

RCRE	0.0415	0.0415	0.0414	0.0414	0.0413	0.0413	0.0414	0.0440
RCDPRE	0.0415	0.0414	0.0412	0.0412	0.0408	0.0407	0.0406	0.0400

5. Conclusions

In this article, when it is assumed that the additional random linear constraint on the parameter vector is true, the RCDPRE of the parameter vector in the linear regression model is given, and some properties of the estimator are studied. Specifically, we have demonstrated that under certain conditions, the proposed estimator outperforms CME, RCRE, and DPRE in terms of mean square error matrix. Finally, we use a numerical example and a simulation study to explain our findings.

References

[1] Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. Technometrics. 1970 Feb; 12(1):55–67, doi: 10.2307/1267351

[2] Liu KJ. A new class of biased estimate in linear regression. Communications in Statistics - Theory and Methods. 1993; 22(2): 393–402, doi: 10.1080/03610929308831027

[3] Yang H, Chang XF, Liu DQ. Improvement of the Liu estimator in weighted mixed regression. Communications in Statistics - Theory and Methods. 2009; 38(2):285-292, doi: 10.1080/03610920802192513

[4] Wu JB, Yang H. Efficiency of an almost unbiased two-parameter estimator in linear regression model. Statistics. 2013; 47(3):535-545. doi: 10.1080/02331888.2011.605891

[5] Akdeniz F, Erol H. Mean squared error matrix comparisons of some biased estimators in linear regression. Communications in Statistics - Theory and Methods. 2003; 32(12):2389–2413.doi:10.1081/STA-120025385

[6] Zhong Z, Yang H. Ridge estimation to the restricted linear model. Communications in Statistics - Theory and Methods. 2007; 36(11): 2099–2115.doi: 10.1080/03610920601144095

[7] Ozkale MR, Kaciranlar S. The restricted and unrestricted two-parameter estimators. Communications in Statistics - Theory and Methods. 2007; 36 (15):2707–2725, doi: 10.1080/03610920701386877

[8] Wu JB Efficiency of restricted r-d class estimator in linear regression. Applied Mathematics and Computation. 2014 June; 236:572-579, doi: 10.1016/j.amc.2014.03.019

[9] Theil H, Goldberger AS. On pure and mixed statistical estimation in economics. In: Raj B., Koerts J. editors. Advanced Studies in Theoretical and Applied Econometrics, vol 23. Springer, Dordrecht. 1992; 317-332, doi: 10.1007/978-94-011-2546-8_18

[10] Li YL, Yang H. A new stochastic mixed ridge estimator in linear regression model, Stat Papers, 2010; 51:315-323, doi: 10.1007/s00362-008-0169-5

[11] Li YL, Yang H. Efficiency of a stochastic restricted two-parameter estimator in linear regression. Applied Mathematics and Computation. 2014 Dec; 249: 371-381, doi: 10.1016/j.amc.2014.10.011

[12] Alheety MI, A new stochastic restricted biased estimator under heteroscedastic or correlated error. ESAIM: Probability and Statistics. 2011 February; 15: 30-40, doi: 10.1051/ps/2009009

[13] Farebrother RW. Further results on the mean square error of ridge regression. Journal of the Royal Statistical Society: Series B (Methodological). 1976; 38(3):248–250, doi: 10.1111/J.2517-6161.1976.TB01588.X

[14] C.R. Rao, H. Toutenburg, Shalabh, C. Heumann, Linear Models and Generalizations: Least Squares and Alternatives, Springer-Verlag, New York, 2008.

[15] M. Gruber, Improving Efficiency by Shrinkage: the James–Stein and Ridge Regression Estimators, Marcel Dekker, Inc., New York, 1998.