P- ISSN 1991-8941 E-ISSN 2706-6703 2023,(17), (2):269– 272
Journal of University of Anbar for Pure Science (JUAPS) Open Access

Modified Unbiased Optimal Estimator for Linear Regression Model

Hussein Ali A. Jassim*, Mustafa I. Alheety

Department of Mathematics, College of education for pure science University of Anbar. Anbar. Iraq;

ARTICLE INFO

Received: 05 / 06 /2023
Accepted: 16 /07/ 2023
Available online: 18 / 12 / 2023

DOI: 10.37652/juaps.2023.140832.1078

Keywords:
Almost Unbiased Ridge Estimator , Modified Almost Unbiased Two-Parameter , Generalized Unbiased Estimator , Mean Squared Error .

ABSTRACT

In this paper, we propose a novel form of Generalized Unbiased Optimal Estimator where the explanatory variables are multicollinear. The proposed estimator's bias, variance, and mean square error matrix (MSE) are calculated. The MSE criterion is used to compare the performance of this estimator against that of other estimators. Finally, a numerical example is examined to understand more about the new estimator's performance.

1.Introdeuction:

Consider the following multiple linear regression model

$Y = X\beta + \epsilon$ (1)

Where X is an $n \times p$ known matrix of independent variable, Y is an $n \times 1$ vector of remarks the dependent variable, β is an $p \times 1$ vector of unknown regression coefficients, and ϵ is an $n \times 1$ vector of errors term disturbance, such that $E(\epsilon) = 0$ and $V(\epsilon) = \sigma^2 I$. The ordinary least square (OLS) estimator of β in model (1) is an given by,

$\hat{\beta} = S^{-1}X'Y$ (2)

Where $S = X'X$. For a long time, it was assumed that OLS was the best estimator; however, the results have shown that OLS is no longer a viable estimator when the variance is large. And this has been shown to be inefficient when there is a multicollinearity relationship. There are numerous biased capabilities available to address this issue by ordinary ridge regression (ORR) estimator, Hoerl and Kennard (1970).

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

The Liu (1993) tried to solve the problem estimator with a ridge estimator $\hat{\beta}_{LU}(d)$, the unbiased ridge regression estimator (URR) was developed by Crouse et al. (1995), almost unbiased ridge estimator (AURE) was in idea indicated Singh and Chaubey (1986), in a linear regression model, the modified almost unbiased Liu (AULE) estimator Arumairajan, S. et al. (2017), as the modified almost unbiased two-parameter (AUTP) estimator Lukman, A. F., et al. (2019). The two-parameter estimator (TPE) was proposed by Ozkale and Kaciranlar (2007). All these are some of the biased estimators proposed to handle the multicollinearity problem using only sample data. The estimators are as follows:

$\hat{\beta}_{URR} = S_k^{-1}(X'y + kJ)$ (4)

where $S_k = S + kI$

$\hat{\beta}_{AURE}(k) = [I - k^2(S + k)^{-2}]\hat{\beta}_{OLSE} = A_k\hat{\beta}_{OLSE}$, (5)

where $A_k = [I - k^2(S + k)^{-2}]$

$\hat{\beta}_{Lu}(d) = (S + I)^{-1}(S + dI)\hat{\beta}_{ORR} = F_d\hat{\beta}_{OLSE}$, (6)

where $F_d = (S + I)^{-1}(S + dI)$,

$\hat{\beta}_{AULE} = [I - (1 - d)^2(S + kI)^{-2}]\,\hat{\beta}_{OLSE} = T_d\,\hat{\beta}_{OLSE}$ (7)

*Corresponding author at: Department of Mathematics, College of education for pure science University of Anbar. Anbar. Iraq;
ORCID:https://orcid.org/0000-0000-0000-0000; Tel: +964
E-mail address: hus21u2001@uoanbar.edu.iq

where $T_d = I-(1-d)^2(S+kI)^{-2}$
$\hat{\beta}_{AUTP} = [I-k^2(1-d)^2(S+kI)^{-2}]\,\hat{\beta}_{OLSE} = P\,\hat{\beta}_{OLSE}$
(8)

where $P = I-k^2(1-d)^2(S+kI)^{-2}$
$\hat{\beta}_{TPE} = (S+kI)^{-1}(S+kdI)\hat{\beta}_{OLSE} = T_{kd}\hat{\beta}_{OLSE},$
(9)
$T_{kd} = (S+kI)^{-1}(S+kdI)$

In cruder to impure the perforuarace of the estimator Hussein and Alheety in (2023) prove three biased estimator based on URR. These estimators as follows: MUORE, MUDE, and MUAURE, based on URR. Since ORR, Liu, AURA, Hussein and Alheety presented a generalized form for these three estimators, the Generalized Unbiased Estimator (GMUE) which is given as

$\hat{\beta}_{GMUE} = A_i\hat{\beta}_{URR}$ (10)

where (A_i) is a positive definite matrix , $i = 1,2,\dots,6$ and ($A_1 = W$, $A_2 = F_d$, $A_3 = A_k$) The bias vector, dispersion matrix and MSE matrix of $\hat{\beta}_{GMUE}$ are given as:
The properties of the estimator GMUE are calculated as follows

$E(\hat{\beta}_{GMUE}) = A_i\beta$, (11)
$Bais(\hat{\beta}_{GMUE}) = (A_i - I)\beta,$
$Cov(\hat{\beta}_{GMUE}) = \sigma^2 A_i S_k^{-1} A_i'$, (12)
and
$MES(\hat{\beta}_{GMUE}) = \sigma^2 A_i S_k^{-1} A_i' + (A_i - I)\beta\beta'(A_i - I)'$
(13)

respectively.

Rather than modifying the matrix A_i to introduce a new biased estimator, we get the best choice of A_i in this study by minimizing the mean square error matrix (MES) of GMUR with respect to A_i.

2 The propertied Estimator

By using the trace operator as in (13), the following equation can be produced as

$tr[MES(\hat{\beta}_{GMUE})] = \sigma^2 tr\, A_i S_k^{-1} A_i' + (A_i - I)\beta\beta'(A_i - I)'$ (14)

where tr is the trace a matrix and it is the sum of the diagonal elements' of square matrix
Now we will use the follows theorems(Rao, C.R. 1995)

A) Theorem : Let b be an n vector and X be an $(n\times n)$ symmetric matrix, then
$$\frac{\partial}{\partial b} b'Xb = 2Xb .$$
B) Theorem: If a be an n vector, Y be a vector, and M be an $(n\times m)$ matrix then
$$\frac{\partial}{\partial M} a'MY = aY'$$
C) Theorem: Suppose that b is m-vector, A is symmetric matrix $(T\times T)$, and C be $(T\times m)$ matrix then
$$\frac{\partial b'C'}{\partial C} a'C'bC = 2ACbb'.$$
By minimizing (14) that respect to A_i, the optimum A_i can be acquired.
$$\frac{\partial\{tr[MES(\hat{\beta}_{GMUE})]\}}{\partial A_i} = \frac{\partial[\sigma^2 tr\,(A_i S_k^{-1} A_i')]}{\partial A_i} + \frac{\partial(A_i - I)\beta\beta'(A_i - I)'}{\partial A_i}$$
(15)

The following solution is obtained by taking the previous equation's derivative with respect to (A_i) and putting the result equal to zero:

$= 2A_i\sigma^2 S_k^{-1} + 2(A_i - I)\beta\beta'$ (16)
$[\,2A_i\sigma^2 S_k^{-1} + 2(A_i - I)\beta\beta' = 0] \div 2$
$A_i\sigma^2 S_k^{-1} + (A_i - I)\beta\beta' = 0$
$A_i\sigma^2 S_k^{-1} + A_i\beta\beta' - \beta\beta' = 0$
$A_i(\sigma^2 S_k^{-1} + \beta\beta') = \beta\beta'$
$\tilde{A}_i = \beta\beta'(\sigma^2 S_k^{-1} + \beta\beta')^{-1}$ (17)

Note that $(\sigma^2 S_k^{-1} + \beta\beta')^{-1}$ exists since $1 + \sigma^{-2}\beta S_k\beta' \neq 0$, it exists Rao, C.R. (1995)

New we are propose a modified unbiased optimal estimator (MUOE) which is given as

$\tilde{\beta}_{MUOE} = \tilde{A}_i\hat{\beta}_{URR}$
(18)

The properties of the estimator MUOE are calculated as follows

$E(\tilde{\beta}_{MUOE}) = \tilde{A}_i\beta$, (19)
$Bais(\tilde{\beta}_{MUOE}) = (\tilde{A}_i - I)\beta,$
$Cov(\tilde{\beta}_{MUOE}) = \sigma^2\tilde{A}_i S_k^{-1}\tilde{A}_i'$, (20)
and
$MES(\tilde{\beta}_{MUOE}) = \sigma^2\tilde{A}_i S_k^{-1}\tilde{A}_i' + (\tilde{A}_i - I)\beta\beta'(\tilde{A}_i - I)'$
(21)
respectively.

Note that since $P = \beta\beta'$ is symmetric and $P^2 = \beta\beta'\beta'\beta = \|\beta\|^2\beta\beta' = cP$ where $c = \|\beta\|^2 =$

$\sum_{i=1}^{P} \beta_i^2$ it can be defined $P = c\left(\frac{1}{c}\beta\beta'\right) = cR$ .Then $R' = R$ and $R^2 = R$. Therefore R is symmetric and idempotent matrix. Now it can create the optimal matrix A_i as $\widetilde{A_\iota} = cR(\sigma^2 S_k^{-1} + cR)^{-1}$

The properties of the estimator MUOE are can be rewritten as calculated as follows

$E(\tilde{\beta}_{\mathrm{MUOE}}) = \widetilde{A_\iota}\beta$, (19)

$\mathrm{Bais}(\tilde{\beta}_{\mathrm{MUOE}}) = -\sigma^2(\sigma^2 I + cRS_k)^{-1}$

$\mathrm{Cov}(\tilde{\beta}_{\mathrm{MUOE}}) = c^2\sigma^2 R(\sigma^2 S_k^{-1} + cR)^{-1} S_k^{-1}(\sigma^2 S_k^{-1} + cR)^{-1} R$ (20)

and

$\mathrm{MES}(\tilde{\beta}_{\mathrm{MUOE}}) = c^2\sigma^2 R(\sigma^2 S_k^{-1} + cR)^{-1} S_k^{-1}(\sigma^2 S_k^{-1} + cR)^{-1} R + c\sigma^2 (\sigma^2 I + cRS_k)^{-1} R(\sigma^2 I + cRS_k)^{-1}$
(21)

respectively.

In practice, we must replace the unknown parameters β and σ^2. For an estimated value for β it is ORR, Liu, AURA, use estimate possible. For the estimated value for $\hat{\sigma}^2$ and $\hat{\beta}$ in unbiased ridge regression estimator (URR). In this section we will discuss the superiority of the estimators when replacing each of these estimators using a numerical example for further illustration.

3.Numerical Example

Using a data set originally created by Woods et al (1932),and used by many researchers as Alheety (2020), we undertake an experiment to confirm the theoretical predictions .Following the identical Ohtani, K, (1986), we substitute the unknown parameters Hoerl .et al (1970), with their unbiased estimators in this experiment. Matlab R2012b was used to calculate the results. The estimator of the ordinary least squares as:

$\hat{\alpha}_{\mathrm{OLSE}} = S^{-1}X'Y = (62.4054,\ 1.55102,\ 0.5102,\ 0.1019,\ -0.1441)'$, (22)

with $\sigma^2 = 5.9830$

TABLE1 : : The mse values of the for MUORE, MUOLE, MUAUER and MUOE with d =0.01

k	MUORE	MUOLE	MUAURE	MUOE
0.01	3097.6	3803.6	3321.1	0.19255
0.1	3797.2	3803.6	5200.4	0.192
0.5	3871.4	3803.6	5517.2	0.18957
0.9	3879.7	3803.6	5639.5	0.18719
1	3880.8	3803.6	5667.2	0.18661
1.5	3883.9	3803.6	5801	0.18374

TABLE2 : : The mse values of the for MUORE, MUOLE, MUAUER and MUOE with d =0.5

k	MUORE	MUOLE	MUAURE	MUOE
0.01	3097.6	1103.9	3321.1	0.19255
0.1	3797.2	985.06	5200.4	0.192
0.5	3871.4	973.24	5517.2	0.18957
0.9	3879.7	971.91	5639.5	0.18719
1	3880.8	971.74	5667.2	0.18661
1.5	3883.9	971.24	5801	0.18374

TABLE3 : : The mse values of the for MUORE, MUOLE, MUAUER and MUOE with d =0.7

k	MUORE	MUOLE	MUAURE	MUOE
0.01	3097.6	610.94	3321.1	0.19255
0.1	3797.2	378.32	5200.4	0.192
0.5	3871.4	355.18	5517.2	0.18957
0.9	3879.7	352.59	5639.5	0.18719
1	3880.8	352.26	5667.2	0.18661
1.5	3883.9	351.28	5801	0.18374

TABLE1 : : The mse values of the for MUORE, MUOLE, MUAUER and MUOE with d =0.9

k	MUORE	MUOLE	MUAURE	MUOE
0.01	3097.6	470.99	3321.1	0.19255
0.1	3797.2	86.764	5200.4	0.192
0.5	3871.4	48.544	5517.2	0.18957
0.9	3879.7	44.251	5639.5	0.18719
1	3880.8	43.713	5667.2	0.18661
1.5	3883.9	42.101	5801	0.18374

Through Table 1, when the value of d = 0.01 and for all values of k the proposed MUOE estimator performed well compared to the estimators within the limits of this paper. It can also be seen that the proposed estimator began to improve a lot when increasing the value of k, where the value of the proposed estimator is the best possible when k = 1.5 can be seen in Tables 1-4. The MUAURE and MUORE methods are not significantly affected by increasing d values.

Conclusions:

In this paper, a new unbiased rate improvement estimator in multiple linear regression is proposed when there is a multiple linearity problem. These estimators outperform other current estimators that rely on sample information. Based on Tables 1–4, the proposed estimators have smaller mse values compared to MUORE, MUOLE, and MUAUER. Thus the that

MUOE is the best estimator compared to other proposed estimators.

REFERENCES

[1] C. Stein,(1956), Inadmissibility of the usual estimator for mean of multivariate normal distribution, Proceedings of the Third Berkley Symposium on Mathematical and Statistics Probability, J. Neyman, ed., Vol. 1, pp. 197–206.

[2] A.E. Hoerl and R.W. Kennard, (1970), Ridge regression: Biased estimation for non-orthogonal problems, Technometrics .

[3] K. Liu, (1993), A new class of biased estimate in linear regression, Commmun. Stat. Theory Methods.

[4] Liu, K. (1993) A New Class of Biased Estimate in Linear Regression. Communications in Statistics Theory and Methods, 22, 393-402.

[5] Singh. B., Chaubey. Y.P. and Dwivedi, T.D. (1986) An Almost Unbiased Ridge Estimator. Sankhya: The Indian Journal of Statistics B, 48, 342-346.

[6] Akdeniz, F. and Kaçiranlar, S. (1995) On the Almost Unbiased Generalized Liu Estimator and Unbiased Estimation of the Bias and MSE. Communications in Statistics Theory and Methods, 34, 1789-1797.

[7] Lukman, A. F., Adewuyi, E., Oladejo, N., & Olukayode, A. (2019, November). Modified Almost Unbiased Two-Parameter Estimator in linear regression model. In IOP Conference Series: Materials Science and Engineering .

[8] Özkale, M. R., & Kaçiranlar, S. (2007). The restricted and unrestricted two-parameter estimators. Communications in Statistics Theory and Methods.

[9] Wang, S. G., Shen, C. B., Long, K., Zhang, T., Wang, F. H., & Zhang, Z. D. (2006). The electrochemical corrosion of bulk nanocrystalline ingot iron in acidic sulfate solution. The Journal of Physical Chemistry B.

[10] Trenkler, G., & Toutenburg, H. (1990). Mean squared error matrix comparisons between biased estimators—An overview of recent results. Statistical papers, 31(1), 165-179.

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

[12] H. Woods, H. H. Steinour, and H. R. Starke, (1932) “Effect of composition of Portland cement on heat evolved during hardening,” Industrial and Engineering Chemistry.

[13] Alheety, Mustafa Ismaeel Naif. (2020) "New versions of liu-type estimator in weighted and non-weighted mixed regression model." Baghdad Science Journal 17.1 (Suppl): 0361-0361

[14] Ohtani, K., (1986). On small sample properties of the almost unbiased generalized ridge estimator. Communications in Statistics-Theory and Methods, 15, 1571-1578.

[15] Rao, C.R. and Touterburg, H. (1995) Linear Models, Least Squares and Alternatives. Springer Verlag.

المقدر الأمثل غير المتحيز المعدل لنموذج الانحدار الخطي

حسين علي ، مصطفى اسماعيل الهيتي

قسم الرياضيات، كلية التربية للعلوم الصرفة، جامعة الانبار، الانبار، العراق

Email **hus21u2001@uoanbar.edu.iq**

الخلاصة:

في هذا البحث، تم اقتراح نوعا جديدًا من المقدر الأمثل غير المتحيز المعمم عندما تعاني المتغيرات التوضيحية من مشكلة التعدد الخطي. تم حساب التحيز والتباين ومصفوفة متوسط الخطأ المربّع (MSE) للمقدر المقترح. كما تم استخدام معيار MSE لمقارنة أداء هذا المقدّر بأداء المقدرين الآخرين. أخيرًا، تم استخدام مثال من بيانات حقيقية كتطبيق لفهم المزيد عن أداء المقدر الجديد.

الكلمات المفتاحية: مقدر الحرف الغير متحيز على الاكثر، المقدر ذو المعلمتين الغير متحيز على الاكثر المعدل، المقدر الغير متحيز المعمم، متوسط مربعات الخطأ.