MATEMATIKA, MJIM, Volume 39, Number 3, 315–327

An Almost Unbiased Regression Estimator: Theoretical Comparison and Numerical Comparison in Portland Cement Data

Set Foong Ng
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA
Cawangan Johor Kampus Pasir Gudang, 81750 Masai, Johor, Malaysia
Corresponding author: ngsetfoong061@uitm.edu.my

Article history
Received: 30 August 2023
Received in revised form: 24 November 2023
Accepted: 8 December 2023
Published on line: 31 December 2023

Abstract Multicollinearity is the problem when there is linear dependency among the independent variables. The Ordinary least squares estimator (OLSE) that is commonly adopted is not suitable for the linear regression model when the independent variables are correlated. This is due to the high variance in OLSE and hence the accuracy of OLSE reduces in the presence of multicollinearity. Hence, the estimator named k-almost unbiased regression estimator (KAURE) was proposed as an alternative to OLSE in this paper. KAURE was developed by using the definition of an almost unbiased estimator to further reduce the bias of Liu-type estimator-special case (LTESC). The properties of KAURE including bias, variance-covariance and mean squared error (MSE) were derived. Theoretical comparison and real-life data comparison were carried out to evaluate the performance of the KAURE based on the MSE criterion. The application of the real-life data supported the theoretical comparison that showed the superiority of KAURE over OLSE and LTESC. The results revealed that KAURE could be considered as an alternative estimator for the linear regression model to combat the problem of multicollinearity.

Keywords Multicollinearity; Almost unbiased estimator; Mean squared error; Linear regression model; Ordinary least squares estimator.

Mathematics Subject Classification 62F10

1 Introduction

Multicollinearity is a common issue that can cause significant problems in various fields, particularly when using linear regression for analysis. Belsley [1] stated that multicollinearity is a natural flaw in data due to the uncontrollable processes of the data-generating mechanism. Multicollinearity occurs when two or more independent variables in a regression model are highly correlated with each other. In other words, they are not providing distinct or independent information to the model. When a near exact relationship between two or more independent variables is present in a data set, multicollinearity exists.

The ordinary least squares estimator (OLSE) is the most popular known estimation technique in regression analysis. OLSE is an unbiased estimator in the regression model. However, it is not a

good estimator when multicollinearity exists in the data [2] due to its large variance. A consequence of having large variance is that the width of the confidence interval for the parameter will be inflated and hence affects the accuracy of the estimator. In other words, multicollinearity can lead to higher standard errors of the coefficients, which makes it harder to detect significant effects. The presence of multicollinearity will also mislead with the significance test telling us that some important variables are not needed in the model. Hence, multicollinearity can lead to coefficients being statistically insignificant even if the variables are theoretically important. This can mask real relationships in the data [3]. In addition, multicollinearity causes a reduction of statistical power in the ability of statistical tests. Therefore, the impact of multicollinearity is serious if the primary interest of a study is in estimating the parameters and identifying the important variables in the process.

As an alternative to the OLSE, many researchers have developed a number of biased estimators. The biased estimators include but are not limited to Shrunken Estimator, Iteration Estimator, Ridge Regression Estimator, Almost Unbiased Ridge Regression Estimator, Restricted Ridge Regression Estimator, Liu Estimator, New Ridge-type Estimator, Modified Two-parameter Regression Estimator [4-11].

In this paper, an almost unbiased estimator named k-almost unbiased regression estimator (KAURE) was developed for the linear regression model in the presence of multicollinearity. Two stages were involved in the development process. We derived a biased estimator named Liu-type estimator-special case (LTESC) in the first stage. In the second stage, motivated by the concept of reducing the distance between the biased estimator and the true value of the parameter, the estimator, KAURE, was developed by reducing the bias of LTESC.

This research paper is organised as follows. The development process of KAURE has been discussed in Section 2. Section 3 contains the theorems for the comparison between KAURE and other estimators in terms of the mean squared error (MSE) criterion. A numerical comparison using a real-life Portland cement data has been performed in Section 4. The conclusion is presented in Section 5.

2 Development Process of KAURE

Linear regression model with p independent variables and a dependent variable can be written in the matrix form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{X}$ is the matrix of independent variables, $\mathbf{Y}$ is the vector of dependent variable and $\boldsymbol{\varepsilon}$ is the vector of error.

Linear regression model in Equation (1) can be written in canonical form

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{Z}$=$\mathbf{XQ}$, $\boldsymbol{\alpha} = \mathbf{Q}'\boldsymbol{\beta}$. Here, $\mathbf{Q}$ and $\boldsymbol{\Lambda}$ satisfy $\mathbf{Q}'\,\mathbf{X}'\,\mathbf{XQ} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ where λ_j is the j^{th} eigenvalue of $\mathbf{X}'\,\mathbf{X}$ and $\mathbf{Q} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_p)$ consists of p eigenvectors of $\mathbf{X}'\,\mathbf{X}$. It is noted that $\mathbf{Q}'\,\mathbf{X}'\,\mathbf{XQ} = \mathbf{Z}'\,\mathbf{Z} = \boldsymbol{\Lambda}$ and $\lambda_j > 0$.

Ordinary least squares estimator (OLSE), $\hat{\boldsymbol{\alpha}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$, is an unbiased estimator for the parameter $\boldsymbol{\alpha}$. OLSE has no bias. However, its variance is unacceptably high in the presence of multicollinearity. Multicollinearity is the problem when there is linear dependency among the

independent variables. Hence, the accuracy of the parameter estimates by using OLSE is reduced due to its high variance.

Instead of using OLSE, biased estimators became the alternative for parameter estimates in linear regression model in the presence of multicollinearity. It is possible for a biased estimator to have an amount of bias but a much smaller variance. Hence, the mean squared error (MSE) of the biased estimator would be less than the MSE of the unbiased estimator OLSE. The mean squared error is a measure of goodness of an estimator. The MSE of an estimator is the sum of its variance and the square of its bias.

Ordinary Ridge regression estimator (ORRE), $\hat{\boldsymbol{\alpha}}_k = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}\mathbf{Z}'\mathbf{Y}$, was proposed by Hoerl and Kennard [6, 12]. The estimation procedure by ORRE was based on adding small positive number to the diagonal of $\mathbf{Z}'\mathbf{Z}$. ORRE was developed from augmenting $\mathbf{0} = \sqrt{k}\boldsymbol{\alpha}+\boldsymbol{\varepsilon}^{\circ}$ to the equation $\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha}+\boldsymbol{\varepsilon}$.

In the first stage to develop KAURE, we derived a biased estimator by augmenting $\frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} = \sqrt{k}\boldsymbol{\alpha}+\boldsymbol{\varepsilon}^{\circ}$ to the equation $\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$. Thus, we get

$$\begin{pmatrix} \mathbf{Y} \\ \frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \\ \sqrt{k}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}^{\circ} \end{pmatrix}.$$

The errors can be expressed as

$$\begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}^{\circ} \end{pmatrix} = \begin{pmatrix} \mathbf{Y} - \mathbf{Z}\boldsymbol{\alpha} \\ \frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} - \sqrt{k}\boldsymbol{\alpha} \end{pmatrix}.$$

Hence, the sum of squares of the errors is given by

$$\begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}^{\circ} \end{pmatrix}' \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}^{\circ} \end{pmatrix} = (\mathbf{Y} - \mathbf{Z}\boldsymbol{\alpha})' (\mathbf{Y} - \mathbf{Z}\boldsymbol{\alpha}) + \left(\frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} - \sqrt{k}\boldsymbol{\alpha}\right)' \left(\frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} - \sqrt{k}\boldsymbol{\alpha}\right).$$

A biased estimator that minimizes the sum of squares of the errors is obtained by solving

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}^{\circ} \end{pmatrix}' \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}^{\circ} \end{pmatrix} = 0$$

as below.

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \left[(\mathbf{Y} - \mathbf{Z}\boldsymbol{\alpha})' (\mathbf{Y} - \mathbf{Z}\boldsymbol{\alpha}) + \left(\frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} - \sqrt{k}\boldsymbol{\alpha}\right)' \left(\frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} - \sqrt{k}\boldsymbol{\alpha}\right) \right] = 0$$

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \left[(\mathbf{Y} - \mathbf{Z}\boldsymbol{\alpha})' (\mathbf{Y} - \mathbf{Z}\boldsymbol{\alpha}) + \left(\frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} - \sqrt{k}\boldsymbol{\alpha}\right)' \left(\frac{1}{\sqrt{k}}\hat{\boldsymbol{\alpha}} - \sqrt{k}\boldsymbol{\alpha}\right) \right] = 0$$

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \left[(\mathbf{Y})' (\mathbf{Y}) - 2 (\boldsymbol{\alpha})' (\mathbf{Z})' \mathbf{Y} + (\boldsymbol{\alpha})' (\mathbf{Z})' \mathbf{Z}\boldsymbol{\alpha} + \frac{1}{k} (\hat{\boldsymbol{\alpha}})' \hat{\boldsymbol{\alpha}} - 2 (\boldsymbol{\alpha})' \hat{\boldsymbol{\alpha}} + k (\boldsymbol{\alpha})' \boldsymbol{\alpha} \right] = 0$$

$$-2 (\mathbf{Z})' \mathbf{Y} + 2 (\mathbf{Z})' \mathbf{Z}\boldsymbol{\alpha} - 2\hat{\boldsymbol{\alpha}} + 2k\boldsymbol{\alpha} = 0$$

$$\begin{aligned} \boldsymbol{\alpha} &= (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}(\mathbf{Z}'\mathbf{Y} + \hat{\boldsymbol{\alpha}}) \\ &= (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1} \left[(\mathbf{Z}'\mathbf{Z})\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\alpha}}\right] \\ &= (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}(\mathbf{Z}'\mathbf{Z} + \mathbf{I})\hat{\boldsymbol{\alpha}} \end{aligned} \tag{3}$$

Here, we named biased estimator in Equation (3) as Liu-type estimator-special case (LTESC). LTESC can be expressed as Equation (4) or Equation (5).

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_{LTESC} &= (\mathbf{Z}'\mathbf{Z}+k\mathbf{I})^{-1}(\mathbf{Z}'\mathbf{Z}+\mathbf{I})\hat{\boldsymbol{\alpha}} \qquad (4)\\ &= \left[(\mathbf{Z}'\mathbf{Z}+k\mathbf{I})^{-1}(\mathbf{Z}'\mathbf{Z}+k\mathbf{I}) - (\mathbf{Z}'\mathbf{Z}+k\mathbf{I})^{-1}(k-1)\right]\hat{\boldsymbol{\alpha}}\\ &= \left[\mathbf{I} - (\mathbf{Z}'\mathbf{Z}+k\mathbf{I})^{-1}(k-1)\right]\hat{\boldsymbol{\alpha}} \qquad (5)\end{aligned}$$

LTESC can also be written as

$$\hat{\boldsymbol{\alpha}}_{LTESC} = \mathbf{H}_{LTESC}\hat{\boldsymbol{\alpha}}, \qquad (6)$$

where $\mathbf{H}_{LTESC} = (\boldsymbol{\Lambda}+k\mathbf{I})^{-1}(\boldsymbol{\Lambda}+\mathbf{I}) = \mathbf{I}-(\boldsymbol{\Lambda}+k\mathbf{I})^{-1}(k-1)$ and $\mathbf{Z}'\mathbf{Z} = \boldsymbol{\Lambda}$. It is noted that $\hat{\boldsymbol{\alpha}} = \boldsymbol{\Lambda}^{-1}\mathbf{Z}'\mathbf{Y}$ is the OLSE and its variance is $\mathbf{var}(\hat{\boldsymbol{\alpha}}) = \boldsymbol{\Lambda}^{-1}\sigma^2$.

The bias, variance-covariance and MSE for LTESC are given as

$$\mathbf{bias}(\hat{\boldsymbol{\alpha}}_{LTESC}) = (\mathbf{H}_{LTESC}-\mathbf{I})\,\boldsymbol{\alpha} \qquad (7)$$

$$\mathbf{cov}(\hat{\boldsymbol{\alpha}}_{LTESC}) = \mathbf{H}_{LTESC}\boldsymbol{\Lambda}^{-1}\mathbf{H}'_{LTESC}\sigma^2 \qquad (8)$$

$$\begin{aligned}\mathbf{MSE}(\hat{\boldsymbol{\alpha}}_{LTESC}) &= \mathbf{H}_{LTESC}\boldsymbol{\Lambda}^{-1}\mathbf{H}'_{LTESC}\sigma^2 + \boldsymbol{\alpha}'(\mathbf{H}_{LTESC}-\mathbf{I})'(\mathbf{H}_{LTESC}-\mathbf{I})\boldsymbol{\alpha}\\ &= \sigma^2\sum_{j=1}^{p}\frac{1}{\lambda_j}\left(\frac{\lambda_j+1}{\lambda_j+k}\right)^2 + \sum_{j=1}^{p}\alpha_j^2\left(\frac{k-1}{\lambda_j+k}\right)^2 \qquad (9)\end{aligned}$$

Liu [13] proposed Liu-type estimator (LTE) as given in Equation (10). It is noted that LTE is equal to LTESC when $d = -1$.

$$\hat{\boldsymbol{\alpha}}_{LTE} = (\mathbf{Z}'\mathbf{Z}+k\mathbf{I})^{-1}(\mathbf{Z}'\mathbf{Z}-d\mathbf{I})\hat{\boldsymbol{\alpha}} \qquad (10)$$

The first stage development process involved the derivation of the biased estimator named LTESC. Hence, the objective of the first stage was achieved.

In the second stage to develop KAURE, we further improve the LTESC by deducting its bias and obtained the estimator, KAURE, by using the Definition 1. The definition of almost unbiased estimator is given in Definition 1.

Definition 1 *Assume* $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ *is the biased estimator of* $\boldsymbol{\theta}$ *where* $\mathbf{bias}\left(\hat{\boldsymbol{\theta}}_{\mathbf{0}}\right) = E\left(\hat{\boldsymbol{\theta}}_{\mathbf{0}}\right) - \boldsymbol{\theta} = \mathbf{W}\boldsymbol{\theta}$. *Then, the estimator* $\hat{\boldsymbol{\theta}}_A = \hat{\boldsymbol{\theta}}_{\mathbf{0}} - \mathbf{W}\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ *is called almost unbiased estimator based on the biased estimator* $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ *(see* [14]).

The rationale behind is to reduce the distance between the biased estimator and the true value of parameter. This is in line with the direction of obtaining an almost unbiased estimator that could be a better alternative for linear regression when multicollinearity presents in the data [7, 15-16]. Some studies [14, 17-20] that are related to almost unbiased estimator in regression analysis are found in literature.

To derive KAURE, we express the bias of $\hat{\boldsymbol{\alpha}}_{LTESC}$ as Equation (11).

$$\begin{aligned}\mathbf{bias}(\hat{\boldsymbol{\alpha}}_{LTESC}) &= (\mathbf{H}_{LTESC}-\mathbf{I})\,\boldsymbol{\alpha}\\ &= \left[\mathbf{I} - (\mathbf{Z}'\mathbf{Z}+k\mathbf{I})^{-1}(k-1) - \mathbf{I}\right]\boldsymbol{\alpha}\\ &= -(\mathbf{Z}'\mathbf{Z}+k\mathbf{I})^{-1}(k-1)\,\boldsymbol{\alpha} \qquad (11)\end{aligned}$$

Hence, $\mathbf{bias}\left(\hat{\boldsymbol{\alpha}}_{LTESC}\right) = \mathbf{W}\boldsymbol{\alpha}$, where $\mathbf{W} = -(\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}(k-1)$. By using Definition 1, an almost unbiased estimator based on the biased estimator $\hat{\boldsymbol{\alpha}}_{LTESC}$ is derived and we named the almost unbiased estimator as KAURE. The derivation of KAURE is shown as below.

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_{KAURE} &= \hat{\boldsymbol{\alpha}}_{LTESC} - \mathbf{W}\hat{\boldsymbol{\alpha}}_{LTESC} \\
&= \hat{\boldsymbol{\alpha}}_{LTESC} + (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}(k-1)\,\hat{\boldsymbol{\alpha}}_{LTESC} \\
&= \left[\mathbf{I} + (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}(k-1)\right]\hat{\boldsymbol{\alpha}}_{LTESC} \\
&= \left[\mathbf{I} + (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}(k-1)\right]\left[\mathbf{I} - (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1}(k-1)\right]\hat{\boldsymbol{\alpha}} \\
&= \left[\mathbf{I} - (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-2}(k-1)^2\right]\hat{\boldsymbol{\alpha}} \qquad (12)
\end{aligned}$$

KAURE can also be written as

$$\hat{\boldsymbol{\alpha}}_{KAURE} = \mathbf{H}_{KAURE}\hat{\boldsymbol{\alpha}}, \qquad (13)$$

where $\mathbf{H}_{KAURE} = \mathbf{I} - (\boldsymbol{\Lambda} + k\mathbf{I})^{-2}(k-1)^2$.

The bias, variance-covariance and MSE for KAURE are given as

$$\mathbf{bias}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) = \left(\mathbf{H}_{KAURE} - \mathbf{I}\right)\boldsymbol{\alpha} \qquad (14)$$

$$\mathbf{cov}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) = \mathbf{H}_{KAURE}\boldsymbol{\Lambda}^{-1}\mathbf{H}'_{KAURE}\sigma^2 \qquad (15)$$

$$\begin{aligned}
\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) &= \mathbf{H}_{KAURE}\boldsymbol{\Lambda}^{-1}\mathbf{H}'_{KAURE}\sigma^2 + \boldsymbol{\alpha}'\left(\mathbf{H}_{KAURE} - \mathbf{I}\right)'\left(\mathbf{H}_{KAURE} - \mathbf{I}\right)\boldsymbol{\alpha} \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left[1 - \left(\frac{k-1}{\lambda_j + k}\right)^2\right]^2 + \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^4. \qquad (16)
\end{aligned}$$

3 Theoretical Comparison among Estimators using MSE

We made a theoretical comparison between the proposed estimator KAURE with OLSE and LTESC based on MSE in order to check the superiority of the KAURE. The MSE for these estimators are given as below.

$$\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left[1 - \left(\frac{k-1}{\lambda_j + k}\right)^2\right]^2 + \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^4$$

$$\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}$$

$$\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{LTESC}\right) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left(\frac{\lambda_j + 1}{\lambda_j + k}\right)^2 + \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^2$$

3.1 Comparison between KAURE and OLSE

The following comparison shows that KAURE is superior to OLSE when the values of k are according to the Theorem 1.

Theorem 1 *The estimator* $\hat{\boldsymbol{\alpha}}_{KAURE}$ *is superior to* $\hat{\boldsymbol{\alpha}}_{OLSE}$ *using the MSE criterion, that is* $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right) < 0$ *if and only if* $1 < k < \min\left(\omega_j\right)$, *where*

$$\omega_j = \frac{\lambda_j \sqrt{\dfrac{2\sigma^2}{\lambda_j \alpha_j^2 + \sigma^2}} + 1}{1 - \sqrt{\dfrac{2\sigma^2}{\lambda_j \alpha_j^2 + \sigma^2}}} \quad \textit{for } j = 1,\ 2,\ \ldots,\ p.$$

Proof. The difference between $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right)$ and $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right)$ is given by

$$\begin{aligned}
&\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right) \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \left[1 - \left(\frac{k-1}{\lambda_j + k}\right)^2\right]^2 + \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^4 - \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \left[1 - 2\left(\frac{k-1}{\lambda_j + k}\right)^2 + \left(\frac{k-1}{\lambda_j + k}\right)^4\right] - \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} + \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^4 \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \left[-2\left(\frac{k-1}{\lambda_j + k}\right)^2 + \left(\frac{k-1}{\lambda_j + k}\right)^4\right] + \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^4 \\
&= \sum_{j=1}^{p} \frac{1}{\lambda_j} \left(\frac{k-1}{\lambda_j + k}\right)^2 \left[-2\sigma^2 + \left(\frac{k-1}{\lambda_j + k}\right)^2 \sigma^2 + \left(\frac{k-1}{\lambda_j + k}\right)^2 \lambda_j \alpha_j^2\right]
\end{aligned}$$

Since $\frac{1}{\lambda_j}\left(\frac{k-1}{\lambda_j+k}\right)^2 > 0$, then $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right) < 0$ if and only if

$$\begin{aligned}
-2\sigma^2 + \left(\frac{k-1}{\lambda_j + k}\right)^2 \sigma^2 + \left(\frac{k-1}{\lambda_j + k}\right)^2 \lambda_j \alpha_j^2 &< 0 \\
\left(\frac{k-1}{\lambda_j + k}\right)^2 &< \frac{2\sigma^2}{\lambda_j \alpha_j^2 + \sigma^2} \\
0 < \frac{k-1}{\lambda_j + k} &< \sqrt{\frac{2\sigma^2}{\lambda_j \alpha_j^2 + \sigma^2}} \\
1 < k &< \frac{\lambda_j \sqrt{\dfrac{2\sigma^2}{\lambda_j \alpha_j^2 + \sigma^2}} + 1}{1 - \sqrt{\dfrac{2\sigma^2}{\lambda_j \alpha_j^2 + \sigma^2}}}
\end{aligned}$$

Hence, $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right) < 0$ if and only if $1 < k < \min\left(\omega_j\right)$ where

$$\omega_j = \frac{\lambda_j \sqrt{\dfrac{2\sigma^2}{\lambda_j \alpha_j^2 + \sigma^2}} + 1}{1 - \sqrt{\dfrac{2\sigma^2}{\lambda_j \alpha_j^2 + \sigma^2}}} \quad \text{for } j = 1,\ 2,\ \ldots,\ p.$$

3.2 Comparison between KAURE and LTESC

The following comparison shows that KAURE is superior to LTESC when the values of k are according to the Theorem 2.

Theorem 2 *The estimator $\hat{\boldsymbol{\alpha}}_{KAURE}$ is superior to $\hat{\boldsymbol{\alpha}}_{LTESC}$ using the MSE criterion, that is* $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{LTESC}\right) < 0$ *if and only if* $k > \max\left(\tau_j\right)$, *where*

$$\tau_j = \frac{2 - \lambda_j + \lambda_j \sqrt{\dfrac{9\sigma^2 + \lambda_j \alpha_j^2}{\sigma^2 + \lambda_j \alpha_j^2}}}{3 - \sqrt{\dfrac{9\sigma^2 + \lambda_j \alpha_j^2}{\sigma^2 + \lambda_j \alpha_j^2}}} \quad \text{for } j = 1,\ 2,\ \ldots,\ p.$$

Proof. The difference between $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right)$ and $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{LTESC}\right)$ is given by

$$\begin{aligned}
&\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{LTESC}\right) \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left[1 - \left(\frac{k-1}{\lambda_j + k}\right)^2\right]^2 + \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^4 - \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left(\frac{\lambda_j + 1}{\lambda_j + k}\right)^2 - \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^2 \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left[\left(1 + \frac{k-1}{\lambda_j + k}\right)^2 \left(1 - \frac{k-1}{\lambda_j + k}\right)^2 - \left(\frac{\lambda_j + 1}{\lambda_j + k}\right)^2\right] - \sum_{j=1}^{p} \alpha_j^2 \left[\left(\frac{k-1}{\lambda_j + k}\right)^2 - \left(\frac{k-1}{\lambda_j + k}\right)^4\right] \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left(1 - \frac{k-1}{\lambda_j + k}\right)^2 \left[\left(1 + \frac{k-1}{\lambda_j + k}\right)^2 - 1\right] - \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^2 \left[1 - \left(\frac{k-1}{\lambda_j + k}\right)^2\right] \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left(1 - \frac{k-1}{\lambda_j + k}\right)^2 \left[\left(\frac{k-1}{\lambda_j + k}\right)^2 + 2\left(\frac{k-1}{\lambda_j + k}\right)\right] - \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^2 \left(1 - \frac{k-1}{\lambda_j + k}\right)\left(1 + \frac{k-1}{\lambda_j + k}\right) \\
&= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j}\left(1 - \frac{k-1}{\lambda_j + k}\right)^2 \left(\frac{k-1}{\lambda_j + k}\right)\left(\frac{k-1}{\lambda_j + k} + 2\right) - \sum_{j=1}^{p} \alpha_j^2 \left(\frac{k-1}{\lambda_j + k}\right)^2 \left(1 - \frac{k-1}{\lambda_j + k}\right)\left(1 + \frac{k-1}{\lambda_j + k}\right) \\
&= \sum_{j=1}^{p} \frac{1}{\lambda_j}\left(\frac{k-1}{\lambda_j + k}\right)\left(1 - \frac{k-1}{\lambda_j + k}\right)\left[\left(1 - \frac{k-1}{\lambda_j + k}\right)\left(\frac{k-1}{\lambda_j + k} + 2\right)\sigma^2 - \left(\frac{k-1}{\lambda_j + k}\right)\left(1 + \frac{k-1}{\lambda_j + k}\right)\lambda_j \alpha_j^2\right] \\
&= \sum_{j=1}^{p} \frac{1}{\lambda_j}\left(\frac{k-1}{\lambda_j + k}\right)\left(1 - \frac{k-1}{\lambda_j + k}\right)\left[-\left(\sigma^2 + \lambda_j \alpha_j^2\right)\left(\frac{k-1}{\lambda_j + k}\right)^2 - \left(\sigma^2 + \lambda_j \alpha_j^2\right)\left(\frac{k-1}{\lambda_j + k}\right) + 2\sigma^2\right].
\end{aligned}$$

Since $\dfrac{1}{\lambda_j}\left(\dfrac{k-1}{\lambda_j + k}\right)\left(1 - \dfrac{k-1}{\lambda_j + k}\right) > 0$, then $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{LTESC}\right) < 0$ if and only if

$$-\left(\sigma^2 + \lambda_j \alpha_j^2\right)\left(\frac{k-1}{\lambda_j + k}\right)^2 - \left(\sigma^2 + \lambda_j \alpha_j^2\right)\left(\frac{k-1}{\lambda_j + k}\right) + 2\sigma^2 < 0$$

$$\left(\sigma^2 + \lambda_j \alpha_j^2\right)\left(\frac{k-1}{\lambda_j + k}\right)^2 + \left(\sigma^2 + \lambda_j \alpha_j^2\right)\left(\frac{k-1}{\lambda_j + k}\right) - 2\sigma^2 > 0$$

$$\frac{k-1}{\lambda_j + k} > \frac{-\left(\sigma^2 + \lambda_j \alpha_j^2\right) + \sqrt{\left(\sigma^2 + \lambda_j \alpha_j^2\right)^2 + 8\sigma^2\left(\sigma^2 + \lambda_j \alpha_j^2\right)}}{2\left(\sigma^2 + \lambda_j \alpha_j^2\right)}$$

$$\frac{k-1}{\lambda_j+k} > -\frac{1}{2}+\frac{1}{2}\sqrt{\frac{9\sigma^2+\lambda_j\alpha_j^2}{\sigma^2+\lambda_j\alpha_j^2}}$$

$$k > \frac{2-\lambda_j+\lambda_j\sqrt{\frac{9\sigma^2+\lambda_j\alpha_j^2}{\sigma^2+\lambda_j\alpha_j^2}}}{3-\sqrt{\frac{9\sigma^2+\lambda_j\alpha_j^2}{\sigma^2+\lambda_j\alpha_j^2}}}$$

Hence, $\mathbf{MSE}\left(\hat{\alpha}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\alpha}_{LTESC}\right) < 0$ if and only if $k > \max\left(\tau_j\right)$, where

$$\tau_j = \frac{2-\lambda_j+\lambda_j\sqrt{\frac{9\sigma^2+\lambda_j\alpha_j^2}{\sigma^2+\lambda_j\alpha_j^2}}}{3-\sqrt{\frac{9\sigma^2+\lambda_j\alpha_j^2}{\sigma^2+\lambda_j\alpha_j^2}}} \quad \text{for } j = 1,\ 2,\ \ldots,\ p.$$

4 Application in Real-life Portland Cement Data

To evaluate the performance of the proposed KAURE on real-life data, real-life Portland cement data [2, 10, 21] is applied to perform the numerical comparison. The linear regression model was constructed based on four independent variables and a dependent variable. The dependent variable V represents the heat evolved after 180 days of curing measured in calories per gram of cement. Four independent variables W_1, W_2, W_3 and W_4 represent tricalcium aluminate, tricalcium silicate, tetracalcium aluminoferrite and dicalcium silicate, respectively. Portland cement data is shown in Table 1.

Table 1: Portland cement data

Tricalcium aluminate, W_1	Tricalcium silicate, W_2	Tetracalcium aluminoferrite, W_3	β dicalcium silicate, W_4	Heat evolved after 180 days of curing, V
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

The standardization is done on the dependent variable and independent variables [22]. Each element of the standardized dependent variable and standardized independent variable is obtained by equation (17) and (18), respectively.

$$y_i = \frac{v_i - \bar{v}}{\sqrt{\sum_{i=1}^{n} (v_i - \bar{v})^2}}, \tag{17}$$

$$x_{ij} = \frac{w_{ij} - \bar{w}_j}{\sqrt{\sum_{i=1}^{n} \left(w_{ij} - \bar{w}_j\right)^2}}, \tag{18}$$

where $\bar{v}$ and $\bar{w}_j$ are mean of V and W_j, respectively. It is noted that $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$ where $p = 4$ and $n = 13$.

Hence, the linear regression model is represented by $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The canonical form of the linear regression model is given by $\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$. It is noted that $\mathbf{Z}$=$\mathbf{XQ}$, $\boldsymbol{\alpha} = \mathbf{Q}'\boldsymbol{\beta}$and $\mathbf{Q}'\,\mathbf{X}'\,\mathbf{XQ} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$. The matrix $\boldsymbol{\Lambda}$ consists of eigenvalues of $\mathbf{X}'\,\mathbf{X}$ while the matrix $\mathbf{Q}$ consists of the corresponding eigenvectors of $\mathbf{X}'\,\mathbf{X}$. In this Portland cement data, the eigenvalues λ_j are 2.235704, 1.576066, 0.186606 and 0.001624. The estimated OLSE $\hat{\alpha}_j$ are -1.198979, -0.018413, -1.549323, 0.573396. The mean squared error of the regression model is $\hat{\sigma}^2 = 0.002$.

Multicollinearity diagnostics were performed on this Portland cement data. The correlation matrix is equivalent to the matrix $\mathbf{X}'\,\mathbf{X}$, where each element in the matrix represents the correlation coefficient between the independent variables. Table 2 displays the correlation coefficient between the independent variables.

Table 2: Correlation between independent variables

	X_1	X_2	X_3	X_4
X_1	1	0.228579	-0.824134*	-0.245445
X_2	0.228579	1	-0.139242	-0.972955*
X_3	-0.824134*	-0.139242	1	0.029537
X_4	-0.245445	-0.972955*	0.029537	1

*Correlation between X_1 and X_3, correlation between X_2 and X_4 are significant at 0.01 level (2-tailed)

The high correlation coefficient between X_1 and X_3 as well as high correlation coefficient between X_2 and X_4 show that multicollinearity exists in the data. However, the values of correlation coefficient would not be enough since high values of correlation coefficients would only identify multicollinearity involving two independent variables but might miss those involving more than two independent variables.

In addition, variance inflation factors (VIF_j) were investigated. The values of VIF_j are the diagonal element of matrix $(\mathbf{X}'\,\mathbf{X})^{-1}$. The variance inflation factors VIF_j are 38.496211, 254.423162, 46.868386 and 282.512861. It shows that all VIF_j are higher than 10, indicating the existence of multicollinearity in the data [3]. The variance of the parameter estimate is directly proportional to VIF_j. Large value of VIF_j would result in having inflated variances of the parameter estimates.

Hence, inflated width of the confidence intervals of the parameters might perhaps cause one or more confidence intervals to be useless [22]. Furthermore, the condition indices

$$CI_j = \sqrt{\frac{\lambda_{\max}}{\lambda_j}}$$

are 1, 1.191022, 3.461339 and 37.106342. The largest condition index is greater than 30, indicating that moderate to strong dependencies among the independent variables.

4.1 Numerical Comparison between KAURE and OLSE

Using the Portland cement data, the MSE of OLSE is obtained, that is

$$\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right) = \hat{\sigma}^2 \sum_{j=1}^{4} \frac{1}{\lambda_j} = 1.244601.$$

Using Theorem 1, the estimated value of

$$\hat{\omega}_j = \frac{\lambda_j \sqrt{\dfrac{2\hat{\sigma}^2}{\lambda_j \hat{\alpha}_j^2 + \hat{\sigma}^2}} + 1}{1 - \sqrt{\dfrac{2\hat{\sigma}^2}{\lambda_j \hat{\alpha}_j^2 + \hat{\sigma}^2}}}$$

are obtained. Table 3 displays the values of $\hat{\omega}_j$.

Table 3: The estimated value of $\hat{\omega}_j$

j	$\hat{\sigma}^2$	λ_j	$\hat{\alpha}_j$	$\hat{\omega}_j$
1	0.002	2.235704	-1.198979	1.118
2	0.002	1.576066	-0.018413	-11.627
3	0.002	0.186606	-1.549323	1.124
4	0.002	0.001624	0.573396	-3.908

According to Theorem 1, the values of $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right)$, $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right)$ and $\mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{KAURE}\right) - \mathbf{MSE}\left(\hat{\boldsymbol{\alpha}}_{OLSE}\right)$ for k that satisfies $1 < k < \min\left(\omega_j\right)$ are presented in Table 4.

Table 4: Numerical comparison between KAURE and OLSE

k	$\mathbf{MSE}(\hat{\alpha}_{KAURE})$	$\mathbf{MSE}(\hat{\alpha}_{OLSE})$	$\mathbf{MSE}(\hat{\alpha}_{KAURE}) - \mathbf{MSE}(\hat{\alpha}_{OLSE})$
1.001	1.244599	1.244601	-0.000002
1.010	1.244359	1.244601	-0.000242
1.050	1.239011	1.244601	-0.005590
1.110	1.220595	1.244601	-0.024006
1.115	1.218623	1.244601	-0.025978
1.117	1.217818	1.244601	-0.026783

Table 4 shows that the values of $\mathbf{MSE}(\hat{\alpha}_{KAURE}) - \mathbf{MSE}(\hat{\alpha}_{OLSE})$ are less than zero for k that satisfies $1 < k < \min\left(\omega_j\right)$. Hence, the numerical comparison between KAURE and OLSE is aligned with Theorem 1.

4.2 Numerical Comparison between KAURE and LTESC

Using Theorem 2, the estimated value of

$$\hat{\tau}_j = \frac{2 - \lambda_j + \lambda_j \sqrt{\dfrac{9\hat{\sigma}^2 + \lambda_j \hat{\alpha}_j^2}{\hat{\sigma}^2 + \lambda_j \hat{\alpha}_j^2}}}{3 - \sqrt{\dfrac{9\hat{\sigma}^2 + \lambda_j \hat{\alpha}_j^2}{\hat{\sigma}^2 + \lambda_j \hat{\alpha}_j^2}}}$$

are obtained for the Portland cement data. Table 5 displays the values of $\hat{\tau}_j$.

Table 5: The estimated value of $\hat{\tau}_j$

j	$\hat{\sigma}^2$	λ_j	$\hat{\alpha}_j$	$\hat{\tau}_j$
1	0.002	2.235704	-1.198979	1.004
2	0.002	1.576066	-0.018413	15.847
3	0.002	0.186606	-1.549323	1.011
4	0.002	0.001624	0.573396	6.778

According to Theorem 2, the values of $\mathbf{MSE}(\hat{\alpha}_{KAURE})$, $\mathbf{MSE}(\hat{\alpha}_{LTESC})$ and $\mathbf{MSE}(\hat{\alpha}_{KAURE}) - \mathbf{MSE}(\hat{\alpha}_{LTESC})$ for k that satisfies $k > \max\left(\tau_j\right)$ are presented in Table 6.

Table 6: Numerical comparison between KAURE and LTESC

k	$\mathbf{MSE}\,(\hat{\alpha}_{KAURE})$	$\mathbf{MSE}\,(\hat{\alpha}_{LTESC})$	$\mathbf{MSE}\,(\hat{\alpha}_{KAURE}) - \mathbf{MSE}\,(\hat{\alpha}_{LTESC})$
15.900	2.694088	3.323601	-0.629513
16.010	2.701571	3.328571	-0.627
17.005	2.766013	3.370997	-0.604984
19.500	2.905430	3.460625	-0.555195
20.100	2.934960	3.479248	-0.544288
22.025	3.021098	3.532883	-0.511785

The values of $\mathbf{MSE}\,(\hat{\alpha}_{KAURE}) - \mathbf{MSE}\,(\hat{\alpha}_{LTESC})$ in Table 6 are less than zero for k where $k > \max\left(\tau_j\right)$. Hence, the numerical comparison between KAURE and LTESC is aligned with Theorem 2.

5 Conclusion

In this paper, the k-almost unbiased regression estimator (KAURE) was developed as an alternative estimator for the linear regression model in the presence of multicollinearity. We derived KAURE by using the definition of almost unbiased estimator to further reducing the bias of Liu-type estimator-special case (LTESC). We compared the superiority of KAURE with OLSE and LTESC theoretically using mean squared error as criterion. Numerical comparison was also done on a real-life Portland cement data where multicollinearity was presence in the data. The numerical comparisons supported the theoretical findings where KAURE is superior to OLSE and LTESC based on MSE criterion according to the theoretical comparison.

References

[1] Belsley, D.A. *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. New York: John Wiley & Sons. 1991.

[2] Lukman, A. F., Ayinde, K., Binuomote, S. and Clement, O. A. Modified ridge-type estimator to combat multicollinearity: application to chemical data. Journal of Chemometrics. 2019. 33(5): 1-12.

[3] Rawlings, J.O., Pantula, S.G. and Dickey, D.A. *Applied Regression Analysis - A Research Tool*. New York: Springer-Verlag. 1998.

[4] Stein, C.M. Multiple regression. *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*. 1960. 424–443.

[5] Trenkler, G. An iteration estimator for the linear model. *Compstat*. 1978. 125–131.

[6] Hoerl, A. E. and Kennard, R.W. Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*. 1970. 12(1): 55–67.

[7] Akdeniz, F. and 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.

[8] Sarkar, N. A new estimator combining the ridge regression and the restricted least squares methods of estimation. *Communications in Statistics - Theory and Methods*. 1992. 21(7): 1987–2000.

[9] Liu, K. A new class of biased estimate in linear regression. *Communications in Statistics - Theory and Methods*. 1993. 22(2): 393–402.

[10] Kibria, B. M. and Lukman, A. F. A new ridge-type estimator for the linear regression model: Simulations and applications. *Scientifica*. 2020. 1–16.

[11] Dawoud, I. Modified two parameter regression estimator for solving the multicollinearity. *Thailand Statistician*. 2022. 20(4): 842–859.

[12] Hoerl, A. E. and Kennard, R.W. Ridge regression: Applications to nonorthogonal problems. *Technometrics*. 1970. 12(1): 69–82.

[13] Liu, K. Using Liu-type estimator to combat collinearity. *Communications in Statistics - Theory and Methods*. 2003. 32(5): 1009–1020.

[14] Xu, J. and Yang, H. More on the bias and variance comparisons of the restricted almost unbiased estimators. *Communications in Statistics-Theory and Methods*. 2011. 40(22): 4053–4064.

[15] Singh, B., Chaubey, Y.P. and Dwivedi, T.D. An almost unbiased ridge estimator. *Sankhya: The Indian Journal of Statistics*. 1986. 48: 342–346.

[16] Akdeniz, F. and Kaciranlar, S. On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE. *Communications in Statistics - Theory and Methods*. 1995. 24(7): 1789–1797.

[17] Omara, T. M. Almost unbiased modified ridge-type estimator: an application to tourism sector data in Egypt. *Heliyon*. 2022. 8(9): 1–7.

[18] Alheety, M. I., Qasim, M., Mansson, K. and Kibria, B. G. Modified almost unbiased two-parameter estimator for the Poisson regression model with an application to accident data. *SORT-Statistics and Operations Research Transactions*. 2021. 45(2): 121–142.

[19] Al-Taweel, Y. and Algamal, Z. Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model. *Periodicals of Engineering and Natural Sciences*. 2020. 8(1): 248–255.

[20] Alheety, M. I. and Kibria, B. G. On the Liu and almost unbiased Liu estimators in the presence of multicollinearity with heteroscedastic or correlated errors. *Surveys in Mathematics and its Applications*. 2009. 4: 155–167.

[21] Woods, H., Steinour, H. H. and Starke, H. R. Effect of composition of Portland cement on heat evolved during hardening. *Industrial & Engineering Chemistry*. 1932. 24(11): 1207–1214.

[22] Ryan, T. P. *Modern Regression Methods*. New York: John Wiley & Sons. 1997.