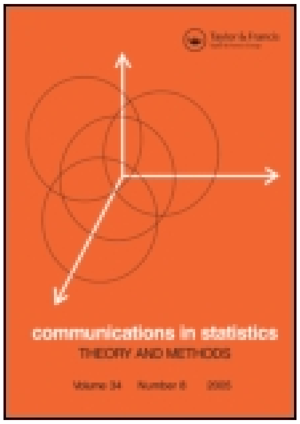

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

Modified Liu-Type Estimator Based on (r - k) Class Estimator

Mustafa Ismaeel Alheety [a] & B. M. Golam Kibria [b]

[a] Department of Mathematics, University of Al-Anbar, Ramadi, Iraq

[b] Department of Mathematics & Statistics, Florida International University, Miami, Florida, USA

Published online: 07 Dec 2012.

Communications in Statistics—Theory and Methods, 42: 304–319, 2013

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610926.2011.577552

Modified Liu-Type Estimator Based on $(r - k)$ Class Estimator

MUSTAFA ISMAEEL ALHEETY[1]
AND B. M. GOLAM KIBRIA[2]

[1]Department of Mathematics, University of Al-Anbar, Ramadi, Iraq
[2]Department of Mathematics & Statistics, Florida International University, Miami, Florida, USA

In this article, we introduced a new Liu-type estimator which includes the ordinary least squares estimator (OLS), ordinary ridge regression estimator (ORR), Liu estimator (LE), $(k - d)$ class estimator, principal components regression (PCR) estimator, $(r - d)$ class estimator, and $(r - k)$ class estimator. Under some conditions, the performance of the proposed estimator is superior to the other estimators by using the scalar mean squares error criterion. A simulation study has been conducted to compare the performance of the estimators. Finally, a numerical example has been analyzed to illustrate the theoretical results of the article.

Keywords $(r - d)$ class estimator; $(r - k)$ class estimator; $(k - d)$ class estimator; Liu estimator; Mean squares error; Multicollinearity; Ordinary least squares estimator; Ordinary ridge regression estimator.

Mathematics Subject Classification 62J12; 62J05; 62J07.

1. Introduction

In multiple linear regression model, we usually assume that the explanatory variables are independent. However, in practice, there may be moderate to strong linear relationships that exist among the explanatory variables. In that case, the independence assumptions are no longer valid, which causes the problem of multicollinearity. In the presence of multicollinearity, it is very difficult to estimate the unique effects of individual variables in the regression equation. Moreover, the regression coefficient will may experience with unexpected large sampling variance which affects both inference and prediction. Therefore, multicollinearity becomes one of the serious problems in the linear regression analysis. There are various methods available to solve this problem in the literature and different remedial actions have been proposed. Hoerl and Kennard (1970a,b) proposed a popular numerical technique to deal with multicollinearity which is known as ridge

Received September 25, 2010; Accepted March 29, 2011
Address correspondence to B. M. Golam Kibria, Department of Mathematics & Statistics, Florida International University, Miami, FL 33199, USA; E-mail: kibriag@fiu.edu

regression. The ordinary ridge regression (ORR) estimator is biased, but under certain conditions it gives smaller mean squares error compared to ordinary least squares (OLS) estimator. Also, Liu (1993) combined the Stein (1956) estimator with the ORR estimator to obtain Liu estimator (LE). Baye and Parker (1984) proposed a generalized estimator, known as $(r-k)$ class estimator, which includes as special cases the principal component regression (PCR), ORR and, of course, ordinary least squares (OLS) estimators. Kaçiranlar and Sakallioğlu (2001) combined both LE estimator and PCR estimator and called this the $r-d$ class estimator which includes as special cases the PCR, LE, and OLS estimators. Liu (2003) introduced a new estimator by combining the ORR estimator with any estimator of β. He called this new estimator as the Liu-type estimator. Using both theoretical results and simulation study, he showed that the new estimator has two advantages over ridge regression. Sakallioglu and Kaciranlar (2008) introduced a new biased estimator which includes as a special cases the OLS estimator, ridge regression estimator, and Liu estimator that provides an alternative method of dealing with multicollinearity. They called this new estimator as the $(k-d)$ class estimator.

In this article, we proposed a new estimator, which is defined by combining the Liu estimator and the $(r-k)$ class estimator. We will call this new estimator as the $(r-(k-d))$ class estimator and we will discuss its properties in details. Under some conditions, the proposed estimator is superior by the scalar mean squares error criterion compare to other estimators. The organization of the article is as follows. The model and the proposed estimators are given in Sec. 2. The comparison of the estimators are provided in Sec. 3. A Monte-Carlo simulation has been conducted in Sec. 4. An example has been considered in Sec. 5. Finally, some concluding remarks are given in Sec. 6.

2. The Model and Some Estimators

We consider the following multiple linear regression model:

$$Y = X\beta + \epsilon, \tag{2.1}$$

where Y is an $n \times 1$ vector of observations on the response (or dependent) variable, X is an $n \times p$ model matrix of observations on p non stochastic explanatory (or independent) variables, β is an $p \times 1$ vector of unknown parameters associated with the p independent variables and ϵ is an $n \times 1$ vector of errors with expectation $E(\epsilon) = 0$ and dispersion matrix $\mathrm{Var}(\epsilon) = \sigma^2 I_n$. Throughout this article, we assume that the model matrix X has full column rank. Suppose there exists an orthogonal matrix $T = (t_1, \ldots, t_p)$ such that $T'ST = \Lambda$, where $S = X'X$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$ contains the eigen values of the matrix S. The orthogonal (canonical form) version of the model (2.1) is

$$Y = Z\alpha + \epsilon, \tag{2.2}$$

where $Z = XT_r$ and $\alpha = T_r'\beta$. Let $T_r = (t_1, \ldots, t_r)$, where $r \leq p$. So, $T_r'ST_r = \Lambda_r = diag(\lambda_1, \ldots, \lambda_r)$ and $T_{p-r}'ST_{p-r} = \Lambda_{p-r} = diag(\lambda_{p-r}, \ldots, \lambda_p)$. It is well known that the ordinary least squares (OLS) estimator, $\hat{\beta} = S^{-1}X'Y$, is unbiased and has minimum variance among all linear unbiased estimators when the fundamental assumptions of the linear model are satisfied. One of these assumptions is that the explanatory variables are independent. If a linear relationship exists among the explanatory variables, the situation is called multicollinearity. Hoerl and Kennard (1970a,b) stated that multicollinearity is a common problem in the field of engineering. To resolve this problem they suggested to use $S + kI_p$, $(k \geq 0)$ instead of S, for estimating β. The resulting estimator is given as

$$\hat{\beta}_k = (S + kI_p)^{-1}X'Y, \tag{2.3}$$

which is known as the ridge regression estimator (RRE). The constant, $k > 0$, is known as shrinkage or biasing or ridge parameter. As k increases from zero and continues up to infinity, the regression estimates tend toward zero. Although these estimators result in biased, for certain value of k, they yield minimum mean square error (MMSE) compared to the OLS estimator. For more with applications on ridge regression we refer to Hoerl and Kennard (1981), Saleh and Kibria (1993), Khalaf and Shukur (2005), Alkhamisi et al. (2006), Alkhamisi and Shukur (2008), and very recently, Alkhamisi (2010), among others. The combination of two different estimators might inherit the advantages of both estimators surely motivated Liu (1993) to study new biased estimator. Liu combined the Stein (1956) estimator with the ORR and defined the new estimator as

$$\hat{\beta}_d = (X'X + I)^{-1}(X'Y + d\hat{\beta}), \quad 0 < d < 1. \tag{2.4}$$

Liu (2003) introduced another estimator, which is defined as

$$\hat{\beta}_{k,d} = (X'X + kI)^{-1}(X'Y - d\beta^*), \quad k > 0, -\infty < d < \infty, \tag{2.5}$$

where β^* can be any estimator of β. This estimator is known as the Liu-type estimator. Sakallioglu and Kaciranlar (2008) introduced a new biased estimator based on the ridge regression estimator, which is defined as

$$\hat{\beta}(k, d) = (X'X + I)^{-1}(X'Y + d\hat{\beta}_k), \quad k > 0, -\infty < d < \infty, \tag{2.6}$$

and known as the $(k - d)$ class estimator.

Baye and Parker (1984) introduced a new estimator based on ridge estimation and the principal components regression (PCR), which is defined as

$$\hat{\beta}_r(k) = T_r(T_r'X'XT_r + kI_r)^{-1}T_r'X'Y, \quad k \geq 0 \tag{2.7}$$

and known as the $(r - k)$ class estimator.

Kaçiranlar and Sakallioğlu (2001) introduced a new estimator based on Liu estimation and the principal components regression (PCR), which is defined as

$$\hat{\beta}_r(d) = T_r(T_r'X'XT_r + I_r)^{-1}(T_r'X'Y + dT_r'\hat{\beta}_r), \quad 0 < d < 1, \tag{2.8}$$

where $\hat{\beta}_r = T_r(T_r'X'XT_r)^{-1}T_r'X'Y$ is *PCR*. The $\hat{\beta}_r(d)$ is known as the $(r - d)$ class estimator.

Now, we are ready to introduce a new alternative estimator for β as follows:

$$\hat{\beta}_r(k, d) = T_r(T_r'X'XT_r + I_r)^{-1}(T_r'X'Y + dT_r'\hat{\beta}_r(k)), \tag{2.9}$$

$-\infty < d < \infty$ and k > 0. We will call the estimator in (2.9) as the $(r-(k-d))$ class estimator for β. This new estimator $\hat{\beta}_r(k, d)$ is a general estimator, which includes the OLS, ORR, LE,$(k-d)$, PCR, $(r-k)$, and $(r-d)$ estimators:

$$\hat{\beta}_p(0, 1) = \hat{\beta},$$
$$\hat{\beta}_p(0, d) = \hat{\beta}(d),$$
$$\hat{\beta}_p(k, d) = \hat{\beta}(k, d),$$
$$\hat{\beta}_r(0, 1) = \hat{\beta}_r,$$
$$\hat{\beta}_r(0, d) = \hat{\beta}_r(d).$$

3. Comparison of the Estimators by MSE Criterion

The mean squares error matrix of an estimator b^* for β is defined as:

$$\mathrm{MSE}(b^*) = E(b^* - \beta)(b^* - \beta)' = \mathrm{Var}(b^*) + (\mathrm{Bias}(b^*))(\mathrm{Bias}(b^*))', \tag{3.1}$$

where $\mathrm{Bias}(b^*) = E(b^*) - \beta$ is the bias of b^* and $Var(b^*) = E[(b^* - E(b^*))(b^* - E(b^*))']$ is a variance of b^*. For a given value of β, b_1^* is preferred to an alternative estimator b_2^*, when $MSE(b_2^*) - MSE(b_1^*)$ is a non negative definite (nnd) matrix.

Another criterion measure of the goodness of an estimator is called the scalar mean squares error of b^* and it is given as: $mse(b^*) = tr(MSE(b^*)) = tr(Var(b^*)) + Bias(b^*))'(Bias(b^*))$, where tr is a trace. If $MSE(b_2^*) - MSE(b_1^*)$ is a non negative definite (nnd) matrix, then $mse(b_2^*) - mse(b_1^*) \geq 0$.

3.1. *Comparison between the* $(r-(k-d))$ *Class Estimator and* $(r-d)$ *Class Estimator*

The matrix (MSE) and scalar (mse) for the $\hat{\beta}_r(k, d)$ and $\hat{\beta}_r(d)$ are given as:

$$\begin{aligned}\mathrm{MMSE}(\hat{\beta}_r(k, d)) = {} & \sigma^2 T_r S_r^{-1}(1)(I_r + dS_r^{-1}(k))T_r'ST_r(I_r + dS_r^{-1}(k))S_r^{-1}(1)T_r' \\ & + (T_r S_r^{-1}(1)(I_r - dS_r^{-1}(k)T_r'ST_r)T_r' + T_{p-r}T_{p-r}')\beta\beta' \\ & (T_r S_r^{-1}(1)(I_r - dT_r'ST_rS_r^{-1}(k))T_r' + T_{p-r}T_{p-r}').\end{aligned} \tag{3.2}$$

$$\mathrm{mse}(\hat{\beta}_r(k, d)) = \sum_{i=1}^{r} \frac{\sigma^2\lambda_i(\lambda_i + k + d)^2 + (\lambda_i + k - d\lambda_i)^2\alpha_i^2}{(\lambda_i + k)^2(\lambda_i + 1)^2} + \sum_{i=p-r}^{p} \alpha_i^2, \tag{3.3}$$

where $S_r^{-1}(1) = (\Lambda_r + I_r)$.

By minimize mse $(\hat{\beta}_r(k, d))$ with respect to d we get

$$d_{opt} = \frac{\sum_{i=1}^{r} \lambda_i(\alpha_i^2 - \sigma^2)/(\lambda_i + k)(\lambda_i + 1)^2}{\sum_{i=1}^{r} \lambda_i(\lambda_i\alpha_i^2 + \sigma^2)/(\lambda_i + k)^2(\lambda_i + 1)^2}. \tag{3.4}$$

When $k = 0$ in (3.3), we get the mse of the $(r-d)$ class estimator

$$\mathrm{mse}(\hat{\beta}_r(d)) = \mathrm{mse}(\hat{\beta}_r(0, d)) = \sum_{i=1}^{r} \frac{\sigma^2(\lambda_i + d)^2 + \lambda_i(1-d)^2\alpha_i^2}{\lambda_i(\lambda_i + 1)^2} + \sum_{i=p-r}^{p} \alpha_i^2. \tag{3.5}$$

Now let k be fixed. So,

$$\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) = d^2 \sum_{i=1}^{r} \frac{(\lambda_i \alpha_i^2 + \sigma^2)(\lambda_i^2 - (\lambda_i + k)^2)}{\lambda_i(\lambda_i + k)^2)(\lambda_i + 1)^2} + 2dk \sum_{i=1}^{r} \frac{(\alpha_i^2 - \sigma^2)}{(\lambda_i + k)(\lambda_i + 1)^2}.$$

Since k, α_i^2, λ_i, and σ^2 are positive numbers, then

$$\sum_{i=1}^{r} \frac{(\lambda_i \alpha_i^2 + \sigma^2)(\lambda_i^2 - (\lambda_i + k)^2)}{\lambda_i(\lambda_i + k)^2)(\lambda_i + 1)^2} < 0 \to \sum_{i=1}^{r} \frac{(\lambda_i \alpha_i^2 + \sigma^2)((\lambda_i + k)^2 - \lambda_i^2)}{\lambda_i(\lambda_i + k)^2)(\lambda_i + 1)^2)} > 0.$$

Let $C_1 = \sum_{i=1}^{r} \frac{(\lambda_i \alpha_i^2 + \sigma^2)(\lambda_i^2 - (\lambda_i + k)^2)}{\lambda_i(\lambda_i + k)^2)(\lambda_i + 1)^2}$ and $C_2 = k \sum_{i=1}^{r} \frac{(\alpha_i^2 - \sigma^2)}{(\lambda_i + k)(\lambda_i + 1)^2}$; then,

$$\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) = d^2 C_1 + 2dC_2$$

Now, when $C_2 > 0$, we want to find the conditions that make $\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) > 0$.

$$\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) = d^2 C_1 + 2dC_2 = d(dC_1 + 2C_2).$$

Therefore, $\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d))$ will be positive when $d < 0$ and $dC_1 + 2C_2 < 0$. In this case,

$$dC_1 + 2C_2 < 0 \Leftrightarrow dC_1 < -2C_2 \Leftrightarrow d(-C_1) > 2C_2 \Leftrightarrow d > \frac{2C_2}{-C_1} = d^* > 0.$$

This result means that the values of d are not all negative, therefore, this result will be canceled.

Also, $\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d))$ will be positive when $d > 0$ and $dC_1 + 2C_2 > 0$. In this case,

$$dC_1 + 2C_2 > 0 \Leftrightarrow dC_1 > -2C_2 \Leftrightarrow d(-C_1) < 2C_2 \Leftrightarrow d < d^* > 0.$$

Therefore, $\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) > 0$ for $0 < d < d^*$.

Now, we are searching for the condition that makes $\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) < 0$. This inequality will be held when $d < 0$ and $dC_1 + 2C_2 > 0 \Leftrightarrow d < d^* > 0$. Therefore, $\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) < 0$ for $d < 0$ or $d < d^*$.

By the same way, when $C_2 < 0$, $\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) > 0$ for $d^* < d < 0$. Also, $\mathrm{mse}(\hat{\beta}_r(k,d)) - \mathrm{mse}(\hat{\beta}_r(d)) < 0$ for $d > 0$ and $d > d^*$.

Thus, we may state Theorem 3.1.

Theorem 3.1.

(a) *When* $\sum_{i=1}^{r} \frac{(\alpha_i^2 - \sigma^2)}{(\lambda_i + k)(\lambda_i + 1)^2} > 0$ *then*:

(1) $mse(\hat{\beta}_r(k,d)) > mse(\hat{\beta}_r(d))$ *for* $0 < d < d^*$.
(2) $mse(\hat{\beta}_r(k,d)) < mse(\hat{\beta}_r(d))$ *for* $d < d^*$ *or* $d < 0$.

(b) *When* $\sum_{i=1}^{r} \frac{(\alpha_i^2-\sigma^2)}{(\lambda_i+k)(\lambda_i+1)^2} < 0$ *then*:

(1) $mse(\hat{\beta}_r(k,d)) > mse(\hat{\beta}_r(d))$ *for* $d^* < d < 0$.
(2) $mse(\hat{\beta}_r(k,d)) < mse(\hat{\beta}_r(d))$ *for* $d > 0$ *and* $d > d^*$.

where

$$d^* = \frac{2k\sum_{i=1}^{r} \frac{(\hat{\alpha}_i^2-\sigma^2)}{(\lambda_i+1)^2(\lambda_i+k)}}{\sum_{i=1}^{r} \frac{((\lambda_i+k)^2-\lambda_i^2)(\lambda_i\alpha_i^2+\sigma^2)}{\lambda_i(\lambda_i+1)^2(\lambda_i+k)^2}}.$$

3.2. Comparison between the $(r-(k-d))$ Class Estimator and $(r-k)$ Class Estimator

The MSE and mse of $(r-k)$ class estimator are, respectively,

$$\begin{aligned} \mathrm{MSE}(\hat{\beta}_r(k)) &= \sigma^2 T_r S_r^{-1}(k)\Lambda_r S_r^{-1}(k)T_r' \\ &\quad + [T_r S_r^{-1}(k)\Lambda_r T_r' - I_p]\beta\beta'[T_r S_r^{-1}(k)\Lambda_r T_r' - I_p]. \end{aligned} \tag{3.6}$$

$$\begin{aligned} \mathrm{mse}(\hat{\beta}_r(k)) &= \mathrm{mse}(\hat{\beta}_r(k, 1-k)) \\ &= \sum_{i=1}^{r} \frac{\sigma^2\lambda_i + k^2\alpha_i^2}{(\lambda_i+k)^2} + \sum_{i=p-r}^{p} \alpha_i^2. \end{aligned} \tag{3.7}$$

When $d = 1-k$ in (3.3), we get the mse of the $(r-k)$ class estimator. However, $\mathrm{mse}(\hat{\beta}_r(k,d))$ is minimized at d_{opt}, thus, we may state the following theorem.

Theorem 3.2. *When*

$$d = \frac{\sum_{i=1}^{r} \lambda_i(\alpha_i^2-\sigma^2)/(\lambda_i+k)(\lambda_i+1)^2}{\sum_{i=1}^{r} \lambda_i(\lambda_i\alpha_i^2+\sigma^2)/(\lambda_i+k)^2(\lambda_i+1)^2}$$

$mse(\hat{\beta}_r(k,d)) \le mse(\hat{\beta}_r(k))$.

3.3. Comparison between the $(r-(k-d))$ Class Estimator and PCR Estimator

The MSE and mse of PCR class estimator are, respectively,

$$\mathrm{MSE}(\hat{\beta}_r) = \sigma^2 T_r S_r^{-1} T_r' + [T_r S_r^{-1}\Lambda_r T_r' - I_p]\beta\beta'[T_r S_r^{-1}\Lambda_r T_r' - I_p]. \tag{3.8}$$

$$\begin{aligned} \mathrm{mse}(\hat{\beta}_r) &= \mathrm{mse}(\hat{\beta}_r(0,1)) \\ &= \sum_{i=1}^{r} \frac{\sigma^2}{\lambda_i} + \sum_{i=p-r}^{p} \alpha_i^2. \end{aligned} \tag{3.9}$$

Baye and Parker (1984) showed that $\mathrm{mse}(\hat{\beta}_r(k)) < \mathrm{mse}(\hat{\beta}_r)$ for $k > \frac{\sigma^2}{max\alpha_i^2} > 0$. But, when $d = 1-k$, $\mathrm{mse}(\hat{\beta}_r(k,1-k)) = \mathrm{mse}(\hat{\beta}_r(k))$. For this, we may state the following theorem.

Theorem 3.3. *For* $k > \frac{\sigma^2}{max\alpha_i^2} > 0$, *there exist* $d < 1-k$ *such that the* $mse(\hat{\beta}_r(k,d))$ *is better than* $mse(\hat{\beta}_r)$.

We can note from our theorems that the comparison results depend on the unknown parameters α and σ^2. In consequence of that, we cannot exclude that our results obtained in the theorems will be held and the results may be changeable. So, we replace them (α and σ^2) by their unbiased estimators. Since d depends on k and on the unknown parameters (α, σ^2) we replace k by its estimator (in this study k is estimated using the estimator that suggested by Hoerl and Kennard (1970a) and we denoted $\hat{k}_{HK} = \hat{\sigma}^2 / \sum_{i=1}^{r} \hat{\alpha}_i^2$). We like to mention here that there are a lot of estimator that suggestion by the researchers to estimate the ridge parameter k, therefore we refer to Kibria (2003) and Muniz and Kibria (2009) for more details about that. So, the estimated d_{opt} will be given as follows:

$$\hat{d}_{opt} = \frac{\sum_{i=1}^{r} \lambda_i(\hat{\alpha}_i^2 - \hat{\sigma}^2)/(\lambda_i + \hat{k}_{HK})(\lambda_i + 1)^2}{\sum_{i=1}^{r} \lambda_i(\lambda_i\hat{\alpha}_i^2 + \hat{\sigma}^2)/(\lambda_i + \hat{k}_{HK})^2(\lambda_i + 1)^2}. \tag{3.10}$$

4. The Monte Carlo Simulation

This section conducted a simulation study to compare the performance of the $(r - (k - d))$ class estimator with other estimators. To achieve different degrees of collinearity, following McDonald and Galarneau (1975), Gibbons (1981) and Kibria (2003), the explanatory variables are generated by using the following equation:

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

where z_{ij} are independent standard normal pseudo-random numbers, $p = 5$ is the number of the explanatory variables, $n = 20$, 50, and 100, and γ is specified so that the correlation between any two explanatory variables is given by γ^2. Four different sets of correlation are considered according to the value of $\gamma = 0.80$, 0.85, 0.95, and 0.99. Also, the explanatory variables are standardized so that $X'X$ will be in correlation form.

According to Alheety and Gore (2008) and Muniz and Kibria (2009), we consider the coefficient vector that corresponded to the largest eigenvalue of $X'X$ matrix. The n observations for the dependent variable are determined by the following equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + e_i,$$

where e_i are independent normal pseudo-random numbers with mean 0 and variance σ^2. In this study, β_0 is taken to be zero. The values of σ^2 are considered as 0.001, 0.05, 0.09, 0.1, 0.5, 1,5, and 25. The experiment is replicated 2,000 times by generating new error terms. The *MSE*s for the estimators are calculated as follows:

$$MSE(\hat{\beta}^*) = \frac{1}{2000} \sum_{r=1}^{2000} (\hat{\beta}^*_{(r)} - \beta)'(\hat{\beta}^*_{(r)} - \beta),$$

where $\hat{\beta}^*$ is any estimator that used in this study for making a comparison.

The simulated MSEs for all estimators are presented in Table 1. It is observed that the proposed AULE estimator performing well compare to others for small σ. It is also noted that the performance of the estimators depend on σ^2, γ,

Table 1
The estimated mean square errors of the OLS, PC, $(r-k)$, $(r-d)$, and $(r-(d-k))$ estimators

		$\gamma = 0.80$					$\gamma = 0.90$				
n	σ^2	OLS	PC	$(r-k)$	$(r-d)$	$(r-(k-d))$	OLS	PC	$(r-k)$	$(r-d)$	$(r-(k-d))$
	0.001	3.2883	0.2726	0.2721	0.2721	0.2718	5.3478	1.2416	1.239	0.1979	1.2372
	0.05	1.6612	3.263	1.7432	1.8811	1.1518	2.9216	4.7019	1.5891	1.7427	1.1637
	0.09	1.3595	2.7622	1.6137	1.6137	1.8566	2.0443	3.4784	1.0843	1.0843	2.1765
100	0.1	1.291	2.6411	1.1173	1.1569	1.0637	1.8814	3.2367	1.0566	1.079	1.2593
	0.5	0.8452	1.2616	1.0004	1.0027	1.0003	1.0027	1.3505	1.0003	1.0021	1.0597
	1	1.1511	1.1511	1.0001	1.002	1.0491	1.2474	1.2474	1.0001	1.0018	1.0313
	5	1.112	1.112	1	1.0017	1.0376	1.214	1.214	1	1.0017	1.0271
	25	1.1101	1.1101	1	1.0017	1.0369	1.2128	1.2128	1	1.0017	1.0268
	0.001	4.1968	0.6646	0.6633	0.6637	0.6627	0.7868	0.2693	0.2693	0.2693	0.2693
	0.05	3.2664	3.4214	1.8334	1.9049	1.1448	2.8207	4.5262	1.924	1.803	1.1407
	0.09	2.7578	2.892	1.8408	1.8408	1.9245	2.182	3.6154	1.211	1.211	1.1932
50	0.1	2.6364	2.7652	1.1376	1.1697	1.1244	2.0519	3.4187	1.1583	1.1644	1.4589
	0.5	1.3479	1.3838	1.001	1.0047	1.0009	1.2586	1.655	1.0058	1.0146	1.1503
	1	1.2846	1.2846	1.0004	1.004	1.0784	1.5417	1.5417	1.0022	1.0071	1.0643
	5	1.2547	1.2547	1.0003	1.0038	1.0685	1.5009	1.5009	1.0002	1.002	1.0506
	25	1.2544	1.2544	1.0003	1.0038	1.0684	1.4984	1.4984	0.9999	1.0011	1.049
	0.001	1.7516	0.7972	0.7963	0.7965	0.796	3.5603	2.2678	2.2638	2.2673	2.2636
	0.05	7.8903	6.2573	2.3133	3.5103	1.313	11.9725	1.4698	1.1634	1.0288	1.0584
	0.09	6.0119	4.6943	2.2822	2.2822	1.8068	8.0121	1.674	1.0939	1.0939	1.4833
20	0.1	5.6316	4.3866	1.2703	1.6139	1.105	7.3807	1.704	1.0833	1.0313	1.0977
	0.5	2.3683	2.0448	1.0233	1.0503	1.0217	3.1728	1.8611	1.0259	1.0316	1.3456
	1	1.9118	1.9118	1.0177	1.0318	1.1899	2.6477	1.8637	1.0233	1.0362	1.1652
	5	1.8613	1.8613	1.0155	1.0288	1.1778	2.6006	1.8619	1.0221	1.0318	1.1534
	25	1.8574	1.8574	1.0153	1.0283	1.1747	2.5973	1.8611	1.0219	1.031	1.1509

(*continued*)

Table 1
Continued

		$\gamma = 0.95$					$\gamma = 0.99$				
n	σ^2	OLS	PC	$(r-k)$	$(r-d)$	$(r-(k-d))$	OLS	PC	$(r-k)$	$(r-d)$	$(r-(k-d))$
	0.001	0.8019	0.2858	0.2858	0.2868	0.2858	0.7989	0.3018	0.3019	0.3018	0.3018
	0.05	4.4775	6.4759	1.5929	2.505	2.3463	2.5422	1.0173	1.0168	1.0158	1.0011
	0.09	2.8387	4.2664	1.0751	1.0751	2.5438	5.7531	1.0122	1.0104	1.0104	1.0218
100	0.1	2.5711	3.883	1.0512	1.1661	1.4951	5.1652	1.0115	1.0095	1.0094	1.0004
	0.5	1.2453	1.3374	1.0019	1.0045	1.0076	3.0405	1.005	1.0034	1.0036	1.0039
	1	1.5007	1.2151	1.0014	1.0033	1.018	3.3193	1.0041	1.0028	1.0029	1.0048
	5	1.4613	1.1741	1.0011	1.0026	1.0142	3.2776	1.0034	1.0024	1.0023	0.0956
	25	1.4594	1.1722	1.0011	1.0025	1.0139	3.2755	1.0033	1.0023	1.0022	1.0139
	0.001	0.8085	0.2899	0.2899	0.2903	0.2899	0.7956	0.3024	0.3025	0.3024	0.3024
	0.05	5.4037	7.3294	1.8338	4.2434	2.4909	3.5697	0.9805	0.9809	0.982	0.9974
	0.09	3.6433	4.7665	1.1427	1.1427	3.0435	8.8971	0.9889	0.9897	0.9897	0.9858
50	0.1	3.3563	4.322	1.1051	1.6042	1.8202	8.3301	0.9902	0.9908	0.9912	0.9991
	0.5	1.9516	1.3889	1.0053	1.0129	1.0108	6.3538	0.9998	0.9988	0.9992	0.9993
	1	2.2092	1.2475	1.0023	1.0056	1.0224	6.6348	1.0011	0.9998	1.0002	1.0011
	5	2.166	1.1987	1.0004	1.0025	1.0152	6.5987	1.0021	1.0006	1.001	0.0954
	25	2.1637	1.196	1.0001	1.002	1.0142	6.5976	1.0023	1.0008	1.0011	1.0143
	0.001	4.4104	2.3195	2.3131	2.3172	2.3123	0.8046	0.3005	0.3008	0.3006	0.3005
	0.05	3.7613	4.4792	1.6203	2.1185	1.5807	9.5681	1.0583	1.0566	1.0529	1.0165
	0.09	3.3205	3.1679	1.17	1.17	2.4645	12.1668	1.0408	1.0371	1.0371	1.0672
20	0.1	3.2637	2.9599	1.1375	1.2341	1.4697	12.2073	1.0383	1.0346	1.0319	1.006
	0.5	3.2124	1.7152	1.0283	1.032	1.0212	12.2543	1.0174	1.0152	1.0125	1.0129
	1	3.3954	1.6592	1.0257	1.0318	1.1276	12.4843	1.0146	1.0127	1.0101	1.0116
	5	3.3809	1.6393	1.0249	1.0279	1.1209	12.5676	1.0123	1.0106	1.0081	0.0961
	25	3.3797	1.638	1.0248	1.0272	1.1195	12.7761	1.0119	1.0102	1.0078	1.1111

and the sample size n. For some values of d and k, the proposed estimator performed better than the rest. After a careful observation of Table 1, we may state that the simulation results are consistent with the theorems in Sec. 3.

5. An Example

To illustrate the theoretical results of this article we now consider a dataset on Portland cement originally due to Wood et al. (1932), and which has been widely analyzed since, (cf. e.g., Alheety and Gore, 2008; Hald, 1952; Kibria, 2003; Kaçiranlar et al., 1999; Liu, 2003; Muniz and Kibria, 2009; Mansson et al., 2010; Nomura, 1988). These data came from an experimental investigation of the heat evolved during the setting and hardening of Portland cements of varied composition and the dependence of this heat on the percentages of four compounds in the clinkers from which the cement was produced. In this example, the dependent variable Y is defined as heat evolved in calories per gram of cement. The independent variables are amounts of the following compounds: tricalcium aluminate (X1), tricalcium silicate (X2), tetracelcium alumino femrrite (X3), and dicalcium silicate (X4). The data set is given in the following Table 2.

Marqaards (1980) introduced the variance inflation factor (VIF) as a formal method of detecting multicollinearity which is now widely accepted. VIF measures how much the variances of the estimated regression coefficients are inflated as compared to the independent variables which are not linearly related.

When the model fitted by least squares method, the variances of the estimates $\hat{\beta}_1, \ldots, \hat{\beta}_p$ are

$$\mathrm{Var}\left(\hat{\beta}_i\right) = \mathrm{VIF}_i\left(\frac{\sigma^2}{S_{ii}}\right), \quad i = 1, 2, \ldots, p,$$

Table 2
Hald data

x_1	x_2	x_3	x_4	y
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

where

$$S_{ii} = \sum_{u=1}^{n} (x_{iu} - \bar{x}_i)^2$$

is the usual corrected sum of squares of x_i.

If a column x_i is orthogonal to all other columns of the X matrix, then $\mathrm{VIF}_i = 1$. Thus, VIF_i is a measure of how much σ^2/S_{ii} is inflated by the relationship of other columns of X with x_i. The VIF_i can be defined specifically in the following way. Suppose that R_i^2 is the coefficient of determination obtained when the i-th predictor variable x_i is regressed on all the remaining predictors x_j with $j \neq i$ in the model. Then

$$\mathrm{VIF}_i = \frac{1}{\left(1 - R_i^2\right)}. \tag{5.1}$$

From (5.1), if R_i^2 equals 0, then VIF_i will be 1. As R_i^2 approaches 1, VIF_i will approach infinity. Marqaards (1980) suggested that a VIF greater than 10 indicates the presence of strong multicollinearity.

In terms of correlation matrix $R_{X'X}$, VIF_i can be considered as the (i, i)th diagonal element of the inverse of $R_{X'X}$:

$$\mathrm{VIF} = diag\left(R_{X'X}^{-1}\right). \tag{5.2}$$

If we make the independent variables in the standard form then we use the variance inflation factor as a measure for existing the multicollinearity, for the Hald data we get $VIF_{x_1} = 38.49$, $VIF_{x_2} = 254.42$, $VIF_{x_3} = 46.86$, and $VIF_{x_4} = 282.51$ which indicate that there is a strong multicollinearity. The model includes the intercept item. The matrix $X'X$ has singular values $\lambda_1 = 44663.3027$, $\lambda_2 = 5965.3394$, $\lambda_3 = 809.9521$, $\lambda_4 = 10.267$, and $\lambda_5 = 105.4058$. The condition number of X is $\kappa = \lambda_{max}/\lambda_{min} = 6056.37$ and so X may be considered as being quite "ill-conditioned." The least squares estimator of the regression coefficients is:

$$\hat{\beta} = (X'X)^{-1}X'Y = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4]' = [62.4052, 1.5511, 0.5102, 0.1019, -0.1441]'.$$

Most authors recommend standardizing the data so that $X'X$ matrix is in the form of a correlation matrix. An advantage of standardization of the data is that the regression coefficients will then be expressed in comparable numerical units. The standardization is accomplished by transforming the linear model $Y = X\beta + \epsilon$ to $Y_s = X_s\beta_s + \epsilon$. Another advantage of standardizing the matrix is that it can show which variables are highly correlated. The corresponding least squares estimator is

$$\hat{\beta}_s = (X_s'X_s)^{-1}X_s'Y_s = [0.6065, 0.5277, 0.0434, -0.1603]'.$$

Since there are 13 observations and 4 parameters for the standardized data, we obtain the estimator of σ^2 as follows:

$$\hat{\sigma}_s^2 = \frac{(Y_s - X_s\hat{\beta}_s)'(Y_s - X_s\hat{\beta}_s)}{n - p} = 0.00196.$$

Table 3
Values of estimates and mse for the estimators

	mse
$\hat{\alpha}$	1.22
$d = 0.1$	
$\hat{\alpha}_r(d)$	0.162
$\hat{\alpha}_r(k, d)$	0.24
$\hat{d}^* = 0.84$	
$\hat{\alpha}_r(d)$	0.167
$\hat{\alpha}_r(k, d)$	0.158
$\hat{d}_{opt} = 0.928$	
$\hat{\alpha}_r(k)$	0.163
$\hat{\alpha}_r(k, d)$	0.153
$\hat{\alpha}_r$	0.165
$\hat{\alpha}_r(k, d)$	0.161

The four eigenvalues of $X_s'X_s$ are 26.8284, 18.9128, 2.2393, and 0.0195. Since the smallest eigenvalue is not zero, the factors do define a four-dimensional space in the mathematical sense. The 4×4 matrix V is the matrix of normalized eigenvectors, Λ is a 4×4 diagonal matrix of eigenvalues of $X_s'X_s$ such that $X_s'X_s = V\Lambda V'$. Then $Z = X_sV$ and $\alpha = V'\beta_s$ so that $Y_s = X_s\beta_s + \epsilon = X_sVV'\beta_s + \epsilon = Z\alpha + \epsilon$. In orthogonal coordinates the least squares estimator of the regression coefficients is:

$$\hat{\alpha} = \Lambda^{-1}Z'Y_s = [-0.65696, -0.00831, 0.30277, -0.38804]'.$$

Numerical results are summarized in Table 3 to compare our new proposed estimator with OLS estimator, $(r-k)$ estimator, PCR estimator, and $(r-d)$ estimator. Values of d_{opt}, d^*, and mse estimates are obtained by replacing in the corresponding theoretical expressions all unknown model parameters by their OLS estimates.

(i) By using the formula (3.10), $\hat{d}_{opt} = 0.9284$, $\hat{k}_{HK} = 0.00374$, and $\hat{d}^* = 0.84$. Now if we look at Table 3, we see that $\hat{\alpha}_r(d)$ has a smaller estimated mse value than $\hat{\alpha}_r(k, d)$ for $0 < d < \hat{d}^*$ where $\sum_{i=1}^{r} \frac{(\alpha_i^2 - \sigma^2)}{(\lambda_i + k)(\lambda_i + 1)^2} > 0$. For example,

$$mse(\hat{\alpha}_r(d = 0.1)) = 0.1623 < mse(\hat{\alpha}_r(\hat{k} = 0.00374, d = 0.1)) = 0.2385.$$

And this is will be the same theoretical result in Theorem 3.1, part a(1). Also $\hat{\alpha}_r(k, d)$ has a smaller estimated mse value than $\hat{\alpha}_r(d)$ for $d < d^*$. For example,

$$mse(\hat{\alpha}_r(k = 0.00374, d = 0.80)) = 0.158 < mse(\hat{\alpha}_r(d = 0.80)) = 0.167.$$

And this is will be the same theoretical result in Theorem 3.1, part a(2).

(ii) By comparing $mse(\hat{\alpha}_r(k = 0.00374, d = 0.9284) = 0.161$ with $mse(\hat{\alpha}_r(\hat{k} = 0.00374)) = 0.163$, we see that $\hat{\alpha}_r(k, d)$ has a smaller estimated mse value than $\hat{\alpha}_r(k)$ and this is will also be the same as the theoretical result of Theorem 3.2.

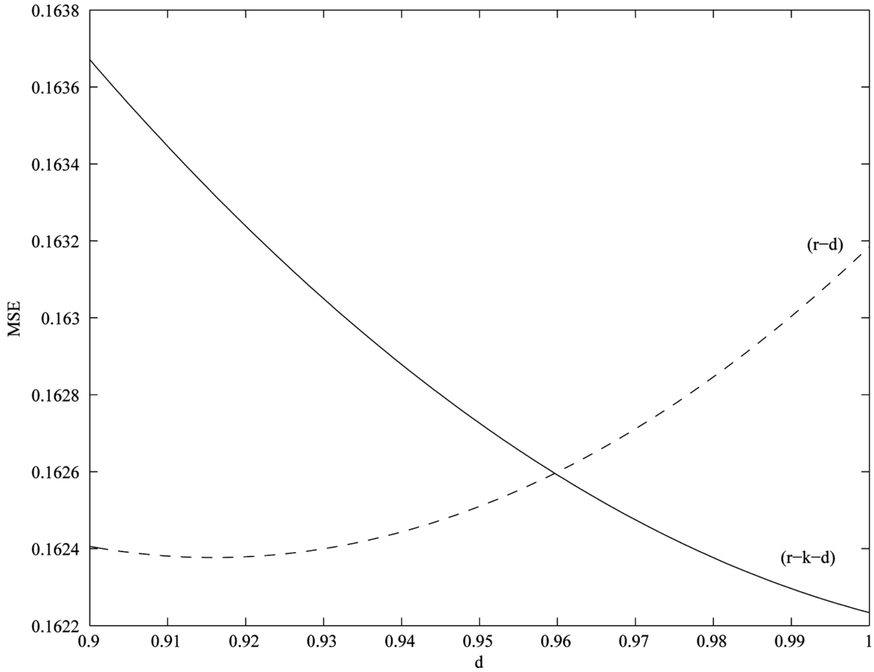

Figure 1. Plot of $MSE(\hat{\beta}_r(k, d))$ and $\hat{\beta}_r(d)$ vs. d when k is fixed at $k = 0.00374$.

(iii) Let $k = 0.005$ and $d = 0.95 < 1 - 0.005$. By comparing $mse(\hat{\alpha}_r(k = 0.005, d = 0.95) = 0.162$ with $mse(\hat{\alpha}_r = 0.1632$,we see that $\hat{\alpha}_r(k, d)$ has a smaller estimated mse value than $\hat{\alpha}_r$, which supported the theoretical result of Theorem 3.3.

The plot of $MSE(\hat{\beta}_r(k, d))$ and $MSE(\hat{\beta}_r(d))$ vs. d in the interval [0, 1] when k is fixed at $k = 0.00374$ has been presented in Fig. 1. This figure indicates that $MSE(\hat{\beta}_r(k, d))$ decreases as d increases and large value of d, $\hat{\beta}_r(k, d)$ dominates $\hat{\beta}_r(d)$. On the other hand, $MSE(\hat{\beta}_r(d))$ increases as d increases slowly and $\hat{\beta}_r(d)$ dominates the estimator $\hat{\beta}_r(k, d)$ for larger space of d. This figure has supported the results in Sec. 3.1.

The plot of $MSE(\hat{\beta}_r(k, d))$ and $MSE(\hat{\beta}_r(k))$ vs. k in the interval [0, 1] when d is fixed at $d = 0.9284$ has been presented in Fig. 2. This figure indicates that both $MSE(\hat{\beta}_r(k, d))$ and $MSE(\hat{\beta}_r(k))$ increase as k increases. The estimator $\hat{\beta}_r(k, d)$ dominates $\hat{\beta}_r(k)$ when $k > 0.10$. This figure has supported the results in Sec. 3.2.

The plot of $MSE(\hat{\beta}_r(k, d))$ and $MSE(\hat{\beta}_r)$ vs k in the interval [0, 0.05] when d is fixed at $d = 0.928$ has been presented in Fig. 3. It is evident from Fig. 3 that both estimators dominate each other for some values of k. For small value of k, the proposed estimator dominates the PCR and for large values of k PCR dominates the proposed estimator. Fig. 3 has supported the results in Sec. 3.3

6. Summary and Concluding Remarks

A new Liu-type estimator, namely $(r - (k - d))$ class estimator, has been proposed in this article. The ordinary least squares estimator (OLS), ordinary ridge regression estimator (ORR), Liu estimator (LE), $(k - d)$ class estimator, principal components

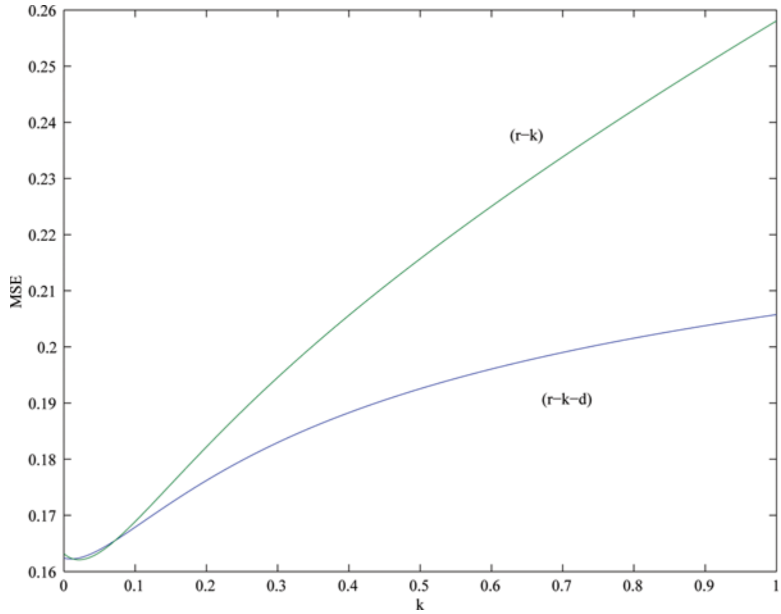

Figure 2. Plot of $MSE(\hat{\beta}_r(k, d))$ and $\hat{\beta}_r(k)$ vs. k when d is fixed at $d = 0.928$. (color figure available online)

regression (PCR), $(r - d)$ class estimator, and $(r - k)$ class estimator are the special cases of the proposed estimator. We observed that under some conditions on d, the proposed estimator performed well compared to others. A simulation study was conducted to compare the perfromance of the estimators. We found

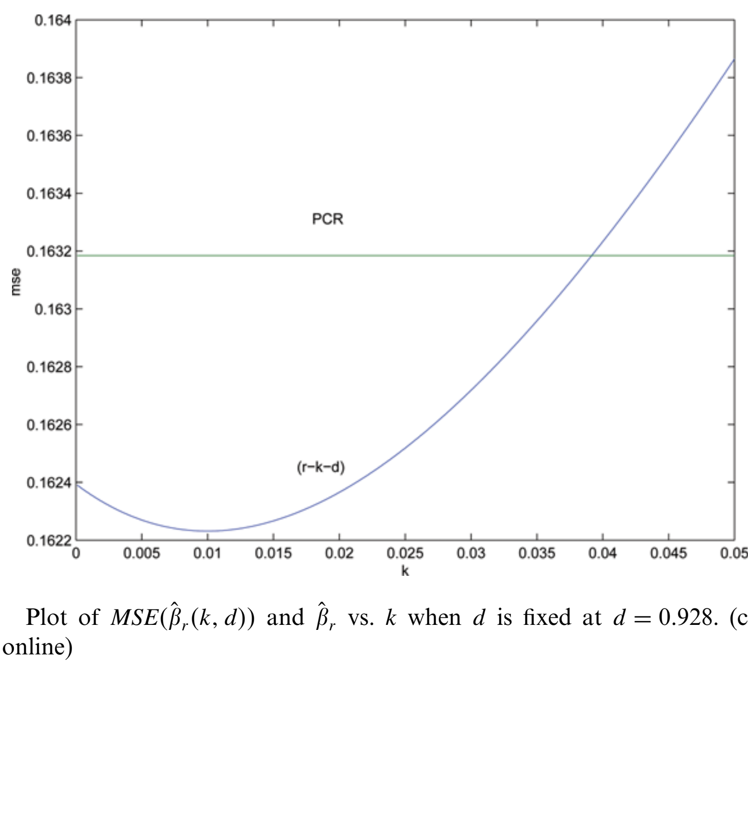

Figure 3. Plot of $MSE(\hat{\beta}_r(k, d))$ and $\hat{\beta}_r$ vs. k when d is fixed at $d = 0.928$. (color figure available online)

that the simulation results are supported the theoretical results of the article. A numerical example based on Portland cement data has been analyzed to illustrate the theoretical results of the article. We believe that the proposed estimator would be useful for the practitioners.

Acknowledgments

The authors are thankful to the referees for their valuable and constructive comments which certainly improved the presentation and quality of the article. This article was completed while the second author was on sabbatical leave (2010–2011). He is grateful to the Florida International University for awarded him the sabbatical leave, which gives him excellent research facilities.

References

Alheety, M. I., Gore, S. D. (2008). A new estimator in multiple linear regression. *Model Ass. Statist. Applic.* 3(3):187–200.

Baye, M. R., Parker, D. F. (1984). Combining ridge and principal component regression. *Commun. Statist. Theor. Meth.* 13(2):197–205.

Alkhamisi, M., Khalaf, G., Shukur, G. (2006). Some modifications for choosing ridge parameters. *Commun. Statist. Theor. Meth.* 35:2005–2020.

Alkhamisi, M., Shukur, G. (2008). Developing ridge parameters for SUR model. *Commun. Statist. Theor. Meth.* 37(4):544–564.

Alkhamisi, M. (2010). Simulation study of new estimators combining the SUR ridge regression and the restricted least squares methodologies. *Statist. Pap.* 51(3):651–672.

Gibbons, D. G. A. (1981). Simulation study of some ridge estimators. *J. Amer. Statist. Assoc.* 76:131–139.

Hald, A. (1952). *Statistical Theory with Engineering Applications*. New York: Wiley.

Hoerl, A. E., Kennard, R. W. (1970a). Ridge regression: Biased estimation for non-orthogonal problem. *Technometrics* 12:55–67.

Hoerl, A. E., Kennard, R. W. (1970b). Ridge regression: Application for non-orthogonal problem. *Technometrics* 12:69–82.

Hoerl, A. E., Kennard, R. W. (1981). Ridge regression – 1980: Advances, algorithms, and applications. *Amer. J. Mathemat. Manage. Sci.* 1:5–83.

Kaçiranlar, S., Sakallioğlu, S., Akdeniz, F., Styan, G. P. H., Werner, H. J. (1999). A new biased estimator in linear regression and detailed analysis of the widely-analysed dataset on Portland Cement. *Sankhya B* 61:443–459.

Kaçiranlar, S., Sakallioğlu, S. (2001). Combining the Liu estimator and the principal component regression estimator. *Commun. Statist. Theor. Meth.* 30(12):2699–2705.

Khalaf, G., Shukur, G. (2005). Choosing ridge parameters for regression problems. *Commun. Statist. Theor. Meth.* 34:1177–1182.

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Commun. Statist. Simul. Computat.* 32:419–435.

Liu, K. (1993). A new class of biased estimate in linear regression. *Commun. Statist. Theor. Meth.* 22:393–402.

Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Commun. Statist. Theor. Meth.* 32(5):1009–1020.

Mansson, K., Shukur, G., Kibria, B. M. G. (2010). A simulation study of some ridge regression estimators under different distributional assumptions. *Commun. Statist. Simul. Computat.* 39(8):1639–1670.

Marqaards, D. W. (1980). You should standardize the predictor variables in your regression models. *J. Amer. Statist. Assoc.* 75:87–91.

McDonald, G. C., Galarneau, D. I. A. (1975). Monte carlo evaluation of some Ridge-type estimators. *J. Amer. Statist. Assoc.* 70:407–416.

Muniz, G., Kibria, B. M. G. (2009). On some ridge regression estimators: An Empirical Comparisons. *Commun. Statist. Simul. Computat.* 38(3):621–630.

Nomura, M. (1988). On the almost unbiased ridge regression estimator. *Commun. Statist. Simul. Computat.* 17:729–743.

Sakallioglu, S., Kaciranlar, S. (2008). A new biased estimator based on ridge estimation. *Statist. Pap.* 49:669–689.

Saleh, A. K. Md., Kibria, B. M. G. (1993). Performances of some new preliminary test ridge regression estimators and their properties. *Commun. Statist. Theor. Meth.* 22:2747–2764.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proc. Third Berkeley Sympo. Math. Statist. Probab.* Universiy of California, Berkeley, pp. 197–206.

Wood, H., Stenior, H. H., Starke, H. R. (1932). Effect of composition of Portland cement on heat evolved during hardening. *Industr. Eng. Chem.* 24:1207–1214.