Two parameter estimators: Biased and almost unbiased estimation for nonorthogonal problems

Muhammad Qasim[1*], Kristofer Månsson[1], Pär Sjölander[1], B.M. Golam Kibria[2]
[1]Department of Economics, Finance and Statistics, Jönköping University, Sweden.
[2]Department of Mathematics and Statistics, Florida International University, Miami, Florida, USA.

*Corresponding Author: muhammad.qasim@ju.se

Abstract

In this paper, we consider the estimation of the parameter (β) in a classical linear regression model by combining the ridge and Liu estimators. The biased and almost unbiased two-parameter estimators are proposed. The necessary and sufficient conditions for the superiority of the proposed estimators over the existing estimators in terms of matrix mean squared error are derived. Besides, we suggest the algorithm for choosing the shrinkage parameters $(k \,\&\, d)$ for newly developed estimators. The performance of the estimators is gauged through Monte Carlo simulation and empirical application.

Key Words: Linear regression model, Multicollinearity, Ridge regression estimator, Portland cement dataset, Biased-corrected two-parameter estimator

1. Introduction

The ordinary least squares estimator (OLSE) is used to estimate the unknown regression coefficients in the classical linear regression model (CLRM). Multicollinearity causes inflated variance in the model, thus making the OLSE unstable in the sense that these are very sensitive to minor changes in the model. Therefore, Hoerl and Kennard (1970) suggested the ridge regression estimator (RRE) minimize this problem by reducing the estimator's variance with the cost of accepting (a small) bias. Estimating the optimal ridge parameter k is a crucial issue for practitioners in the RRE. See, for instance, the following approaches to choose k; McDonald and Galarneau (1975), Gibbon (1981), Kibria (2003), Månsson et al. (2010), Lukman et al. (2017) Saleh et al. (2019), Amin et al. (2020) and among others. Many other directions have been taken in the literature to improve the RRE suggested by Hoerl and Kennard (1970). One method was proposed by Liu (1993), which is known as Liu regression estimator (LRE). Its main advantage over the RRE is that it is a linear function of the shrinkage parameter. Kaçiranlar et al. (1999) improved Liu's approach and introduced a restricted Liu estimator. Akdeniz and Erol (2003) compared some biased estimators in the linear regression in the sense of matrix mean squared error (MMSE). By combining the mixed estimator and the Liu estimator, Hubert and Wijekoon (2006) proposed the stochastic restricted Liu estimator, which outperforms Liu estimator and mixed estimator under certain conditions.

The primary aim of this article is to introduce a new two-parameter estimator that provides an alternative method to mitigate the problem of multicollinearity in the CLRM. This new method encompasses OLSE and RRE as exceptional cases. We also introduce an almost unbiased two-parameter estimator. We compare the MMSE properties analytically and prove the superiority of our new methods under certain conditions. Then, we show the superiority of the proposed estimator in finite samples using a Monte Carlo simulation study. Finally, we apply the methods on chemometric application. Regression models are widely used in chemistry to build efficient and robust prediction models. In the chemometric application, we use the classical Portland cement data that was also analyzed by Lukman et al. (2019). This example models the heat evolved after 180 days of curing cement, measured in calories per gram of cement by four highly correlated variables.

2.Model specification and the estimators

To describe the problem, we consider the following CLRM:

$$Y = X\beta + \varepsilon, \quad (1)$$

where Y is an $(n \times 1)$ vector of observations on the response variable, X is an $(n \times p)$ full ranked design matrix consisting of the explanatory variables, β is an $(p \times 1)$ a column vector of unknown regression coefficients and ε is a $(n \times 1)$ vector of random errors assumed to be normally distributed with $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon^t) = \sigma^2 \mathrm{I}_n$ where I_n is an $(n \times n)$ identity matrix. The OLSE of the unknown parametric vector β is:

$$\hat{\beta}_{OLS} = (X^t X)^{-1} X^t Y. \quad (2)$$

In the presence of multicollinearity, the estimated regression coefficient using $\hat{\beta}_{OLS}$ are too large in the form of absolute value. Therefore, Hoerl and Kennard (1970) and Liu (1993) proposed RRE and LRE as remedy, respectively when $\|\hat{\beta}_{OLS}\|$ is too large in the situation of multicollinearity. Furthermore, RRE and LRE have a smaller length than the OLSE, i.e., $\|\hat{\beta}_{RR}\| < \|\hat{\beta}_{OLS}\|$ and $\|\hat{\beta}_{LR}\| < \|\hat{\beta}_{OLS}\|$. The RRE is obtained by augmenting Eq. (1) with $0 = k^{1/2}\beta + \varepsilon^t$ to and then use the OLSE and derived following form of the estimator:

$$\hat{\beta}_{RR} = (X^t X + kI)^{-1} X^t Y, \ (k > 0) \quad (3)$$

The LRE defined by Liu (1993) as:

$$\hat{\beta}_{LR} = (X^t X + I)^{-1}(X^t X + dI)\hat{\beta}_{OLS}, \quad (4)$$

where d is known as the Liu parameter and it takes the values between zero and one. Liu (1993) obtained $\hat{\beta}_{LR}$ by augmenting $d\hat{\beta}_{OLS} = \beta + \varepsilon^t$ to Eq. (1) and then using the OLS estimate. Özkale and Kaçiranlar (2007) stated that as k becomes larger for the RRE, the distance between $k^{1/2}\beta$ and 0 increases and the RRE have an excessive amount of bias. Therefore, Özkale and Kaçiranlar (2007) proposed a two-parameter estimator (TPE) and it is defined as follows:

$$\hat{\beta}_{kd} = (X^t X + kI)^{-1}(X^t X + kdI)\hat{\beta}_{OLS}, \ k > 0, 0 < d < 1. \quad (5)$$

2.1. Proposed estimators

The $\hat{\beta}_{kd}$ decreases the bias of RRE, but in the presence of severe multicollinearity, the performance of $\hat{\beta}_{kd}$ is still not satisfactory since the value of k may be too small which is then further pushed to zero by d. Also, the standard errors of the regression coefficients are higher under certain conditions. In many real-world chemometric problems, we expect a situation where the multicollinearity is high but imperfect, and the value of the ridge parameter k becomes too small and the performance of $\hat{\beta}_{kd}$ does not satisfactory. On the other hand, for large values of k, $\hat{\beta}_{kd}$ decreases the bias problem but the distance between $k^{1/2}\beta$ and 0 still increase (sometimes substantially). Therefore, we propose another class of two-parameter estimator (KQE) by augmenting Eq. (1) with $\left(\frac{-kd}{k^{1/2}}\right)\hat{\beta}_{OLS} = k^{1/2}\beta + \varepsilon', k \geq 0, 0 \leq d \leq 1$, and then apply the OLSE. The advantage of the new estimator over the existing estimators is that augmenting equation still gives a better fit by choosing appropriate values of k & d. Also, the proposed estimator will give minimum standard errors as compared to the $\hat{\beta}_{kd}$.

2.1.1. Biased two-parameter estimator

Following Hoerl and Kennard (1970), Liu (2003), and Kaciranlar et al. (1999), we proposed the new estimator as follows. Let $\hat{\beta}_{kq}$ be the new estimator of the vector β then we derive it from the following function:

Minimize $\left(Y - X\hat{\beta}_{kq}\right)^t \left(Y - X\hat{\beta}_{kq}\right)$

$$\text{subject to } \left(\hat{\beta}_{kq} + d\hat{\beta}_{OLS}\right)^t \left(\hat{\beta}_{kq} + d\hat{\beta}_{OLS}\right) = \hbar.$$

$$\left(Y - X\hat{\beta}_{kq}\right)^t \left(Y - X\hat{\beta}_{kq}\right) + k\left\{\left(\hat{\beta}_{kq} + d\hat{\beta}_{OLS}\right)^t \left(\hat{\beta}_{kq} + d\hat{\beta}_{OLS}\right) - \hbar\right\}. \quad (6)$$

The final form of the proposed estimator $\hat{\beta}_{kq}$ (KQE) is defined as:

$$\hat{\beta}_{kq} = (X^tX + kI)^{-1}\left(X^tY - dk\hat{\beta}_{OLS}\right) = \mathrm{A}_{kq}\hat{\beta}_{OLS}, k > 0, 0 \leq d \leq 1, \quad (7)$$

where k and d are the shrinkage parameters and $\mathrm{A}_{kq} = \{I - k(1+d)(X^tX + kI)^{-1}\}$. The proposed KQE can also be found as a solution to the linear stochastic restriction problem. By considering the prior information for β in the form of linear stochastic restriction as follows:

$$\left(\frac{-kd}{k^{1/2}}\right)\hat{\beta}_{OLS} = k^{1/2}\beta + \varepsilon^t,$$

where k and d are the shrinkage parameters, ε' is a $(p \times 1)$ vector of random errors with $E(\varepsilon^t) = 0$, $Var(\varepsilon^t) = \sigma^2\mathrm{I}$ and $E(\varepsilon\varepsilon^t) = 0$. The $\hat{\beta}_{kq}$ is the KQE which includes the following estimators as special cases:

$\lim_{k \to 0} \hat{\beta}_{kq} = \hat{\beta}_{OLS} = (X^tX)^{-1}X^tY$, the OLSE.

$\lim_{d \to 0} \hat{\beta}_{kq} = \hat{\beta}_{RR} = (X^tX + kI)^{-1}X^tY$, the RRE.

2.1.2. Almost unbiased two-parameter estimator

Since unbiasedness is a desirable property in real-world applications, so we derive the almost unbiased two-parameter estimator (AUKQE) with adjustment for the bias of the $\hat{\beta}_{kq}$. The bias of $\hat{\beta}_{kq}$ is:

$$\mathrm{Bias}\left(\hat{\beta}_{kq}\right) = -k(1+d)(X^tX + kI)^{-1}\beta.$$

Based on the bias of $\hat{\beta}_{kq}$, it is possible to derive the AUKQE, $\hat{\beta}_{aukq}$

$$\hat{\beta}_{aukq} = \hat{\beta}_{kq} - \mathrm{Bias}\left(\hat{\beta}_{kq}\right) = \hat{\beta}_{kq} + k(1+d)(X^tX + kI)^{-1}\beta.$$

The AUKQE may now be defined by following the methods in Ohtani (1986) where the parameter vector β is replaced with the KQE, $\hat{\beta}_{kq}$ as follows:

$$\hat{\beta}_{aukq} = \hat{\beta}_{kq} + k(1+d)(X^tX + kI)^{-1}\hat{\beta}_{kq}.$$

The final form of the AUKQE is defined as

$$\hat{\beta}_{aukq} = \left(2I - A_{kq}\right)A_{kq}\hat{\beta}_{OLS}. \quad (8)$$

3. The MMSE properties

This section explains the necessary and sufficient conditions for the superiority of the proposed estimators over the existing estimators in the sense of MMSE. In addition, we illustrate the bias comparison between the KQE and the AUKQE. The MMSE of an estimator $\hat{\beta}$ of β can be defined as

$$MMSE(\hat{\beta}) = E(\hat{\beta}-\beta)^t(\hat{\beta}-\beta) = Cov(\hat{\beta}) + Bias(\hat{\beta})\{Bias(\hat{\beta})\}^t, \quad (9)$$

where $Cov(\hat{\beta})$ is represent the covariance matrix of $\hat{\beta}$ and $Bias(\hat{\beta}) = E(\hat{\beta}) - \beta$ is the bias vector. The MMSE of OLSE, RRE, LRE, TPE, KQE and AUKQE is defined as

$$MMSE(\hat{\beta}_{OLS}) = \sigma^2(X^tX)^{-1}. \quad (10)$$

$$MMSE(\hat{\beta}_{RR}) = \sigma^2 A_{RR}(X^tX)^{-1}(A_{RR})^t + (A_{RR}-I)\beta\beta^t(A_{RR}-I)^t, \quad (11)$$

$$MMSE(\hat{\beta}_{LR}) = \sigma^2 A_{LR}(X^tX)^{-1}(A_{LR})^t + (d-1)^2(X^tX+I)^{-1}\beta\beta^t\{(X^tX+I)^{-1}\}^t, \quad (12)$$

$$MMSE(\hat{\beta}_{kd}) = \sigma^2 A_{kd}(X^tX)^{-1}(A_{kd})^t + k^2(d-1)^2(X^tX+kI)^{-1}\beta\beta^t\{(X^tX+kI)^{-1}\}^t, \quad (13)$$

$$MMSE(\hat{\beta}_{kq}) = \sigma^2 A_{kq}(X^tX)^{-1}(A_{kq})^t + (A_{kq}-I)\beta\beta^t(A_{kq}-I)^t. \quad (14)$$

$$MMSE(\hat{\beta}_{aukq}) = \sigma^2(2I-A_{kq})A_{kq}(X^tX)^{-1}\{(2I-A_{kq})A_{kq}\}^t + (I-A_{kq})^2\beta\beta^t\left\{(I-A_{kq})^2\right\}^t. \quad (15)$$

where $A_{RR} = \{I + k(X^tX)^{-1}\}^{-1}$, $A_{LR} = (X^tX+I)^{-1}(X^tX+dI)$, $A_{kd} = (X^tX+kI)^{-1}(X^tX+kdI)$ and $(A_{RR}-I) = -k(X^tX+kI)^{-1}$.

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the two estimators of β, the estimator $\hat{\beta}_2$ is said to be superior to the estimator $\hat{\beta}_1$ if and only if

$$\Theta = MMSE(\hat{\beta}_1) - MMSE(\hat{\beta}_2) \geq 0. \quad (16)$$

We define a variety of lemmas to illustrate the MMSE properties of the proposed estimators:

Lemma 1: Let $\hat{\beta}_j = \mathcal{A}_j \mathscr{y}$, $j = 1,2$ be the two competing estimators of β. Suppose $\Theta = \mathrm{Cov}(\hat{\beta}_1) - Cov(\hat{\beta}_2) > 0$, where $Cov(\hat{\beta}_j)$, $j = 1,2$ denotes the covariance matrix of $\hat{\beta}_j$. Then $\Theta(\hat{\beta}_1, \hat{\beta}_2) = MMSE(\hat{\beta}_1) - MMSE(\hat{\beta}_2) \geq 0 \Leftrightarrow b_2^t(\sigma^2\Theta + b_1 b_1^t)^{-1} b_2 \leq 1$, where b_j denote the bias vector of $\hat{\alpha}_j$, $j = 1,2$.

Proof: See Trenkler and Toutenburg (1990) for more details.

Lemma 2: *Let M $(M > 0)$ be a positive definite (p.d.) matrix, α be a vector of nonzero constants, then $M - \alpha\alpha^t$ is a non-negative definite (n.n.d.) matrix if and only if $\alpha^t M^{-1}\alpha \leq$* 1.

Proof: See Farebrother (1976) for more details.

Theorem 1: *Let $k > 0$ and $0 \leq d \leq 1$ under the CLRM with correlated regressors, the $\hat{\beta}_{kq}$ is superior to the $\hat{\beta}_{OLS}$ in the MMSE sense, namely, $\Theta(\hat{\beta}_{kq}, \hat{\beta}_{OLS})$, if and only if $\beta^t(A_{kq}-I)^t[M_1]^{-1}(A_{kq}-I)\beta \leq \sigma^2$, where $M_1 = \left\{(X^tX)^{-1} - A_{kq}(X^tX)^{-1}(A_{kq})^t\right\}$.*

Proof: From Eq. (10) and Eq. (13), we find the difference between MMSEs as

$$\Theta_1 = \Theta(\hat{\beta}_{OLS}, \hat{\beta}_{kq}) = MMSE(\hat{\beta}_{OLS}) - MMSE(\beta_{kq})$$
$$= \sigma^2\left\{(X^tX)^{-1} - A_{kq}(X^tX)^{-1}(A_{kq})^t\right\} - (A_{kq}-I)\beta\beta^t(A_{kq}-I)^t. \quad (17)$$

We rewrite Eq. (17) as:

$$\Theta_1 = \sigma^2 \text{diag}\left\{\frac{1}{\lambda_j} - \frac{(\lambda_j - kd)^2}{(\lambda_j + k)^2 \lambda_j}\right\}_{j=1}^{p} - (A_{kq} - I)\beta\beta^t (A_{kq} - I)^t,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ are the eigenvalues of the design matrix X^tX. It is clear that when $k > 0$ and $0 \leq d \leq 1$, the matrix $(X^tX)^{-1} - A_{kq}(X^tX)^{-1}(A_{kq})^t$ is p.d. if $(\lambda_j + k)^2 > (\lambda_j - kd)^2$. Therefore, using Lemma 2, Θ_1 is n.n.d. if and only if $\beta^t(A_{kq} - I)^t[M_1]^{-1}(A_{kq} - I)\beta \leq \sigma^2$.

Theorem 2: *Under the CLRM with correlated regressors, if* $(b_{kq})^t[M_2]^{-1}b_{kq} < 1$ 0 for $k > 0$ and $0 \leq d \leq 1$, then $MMSE(\hat{\beta}_{kd}) - MMSE(\hat{\beta}_{kq}) > 0$, *where* $b_{kq} = (A_{kq} - I)\beta$ and $M_2 = \sigma^2\left\{A_{kd}(X^tX)^{-1}(A_{kd})^t - A_{kq}(X^tX)^{-1}(A_{kq})^t\right\}$.

Proof: The difference between the MMSE functions of $\hat{\beta}_{kd}$ and $\hat{\beta}_{kq}$ is obtained as

$$\Theta_2 = \Theta(\hat{\beta}_{kd}, \hat{\beta}_{kq}) = MMSE(\hat{\beta}_{kd}) - MMSE(\hat{\beta}_{kq})$$

$$\Theta_2 = \sigma^2\left\{A_{kd}(X^tX)^{-1}(A_{kd})^t - A_{kq}(X^tX)^{-1}(A_{kq})^t\right\} + k^2(d-1)^2(X^tX + kI)^{-1}\beta\beta^t\{(X^tX + kI)^{-1}\}^t - b_{kq}{b_{kq}}^t. \quad (18)$$

We can write the expression (18) as;

$$= \sigma^2 \text{diag}\left\{\frac{(\lambda_j + kd)^2}{\lambda_j(\lambda_j + k)^2} - \frac{(\lambda_j - kd)^2}{(\lambda_j + k)^2\lambda_j}\right\}_{j=1}^{p} + k^2(d-1)^2(X^tX + kI)^{-1}\beta\beta^t\{(X^tX + kI)^{-1}\}^t - b_{kq}{b_{kq}}^t.$$

$$= \sigma^2 \text{diag}\left\{\frac{4\lambda_j kd}{\lambda_j(\lambda_j + k)^2}\right\}_{j=1}^{p} + k^2(d-1)^2(X^tX + kI)^{-1}\beta\beta^t\{(X^tX + kI)^{-1}\}^t - b_{kq}{b_{kq}}^t.$$

Since $b_{kq}{b_{kq}}^t$ is n.n.d., then it is noticeable that $\left(A_{kd}(X^tX)^{-1}(A_{kd})^t - A_{kq}(X^tX)^{-1}(A_{kq})^t\right) + k^2(d-1)^2(X^tX + kI)^{-1}\beta\beta^t\{(X^tX + kI)^{-1}\}^t$ will be p.d. It can be easily shown that $Cov(\hat{\beta}_{kd}) - Cov(\hat{\beta}_{kq})$ is a p.d. matrix for $k > 0$ and $0 \leq d \leq 1$. Hence, we can state that $\hat{\beta}_{kq}$ has a smaller sampling variance and covariance matrix than the $\hat{\beta}_{kd}$. Thus, the proof is completed through Lemmas 1 and 2.

Theorem 3: *Let* $k > 0$ *and* $0 \leq d \leq 1$ *under the CLRM with correlated regressors, then* $MMSE(\hat{\beta}_{RR}) - MMSE(\hat{\beta}_{kq}) > 0$ *if* $\beta^t(A_{kq} - I)^t[M_3]^{-1}(A_{kq} - I)\beta \leq 1$, *where* $M_3 = \sigma^2\left\{A_{RR}(X^tX)^{-1}(A_{RR})^t - A_{kq}(X^tX)^{-1}(A_{kq})^t\right\}$.

Theorem 4: Let us consider two biased competing estimators, namely $\hat{\beta}_{LR}$ and $\hat{\beta}_{kq}$ of β. *If* $k > 0$ *and* $0 \leq d \leq 1$ *under the CLRM with correlated regressors, the estimator* $\hat{\beta}_{kq}$ *is superior to the estimator* $\hat{\beta}_{LR}$ *in the MMSE form, namely* $MMSE(\hat{\beta}_{LR}) - MMSE(\hat{\beta}_{kq}) > 0$ *if* $\beta^t(A_{kq} - I)^t[M_4]^{-1}(A_{kq} - I)\beta < 1$, *where* $M_4 = \sigma^2\left\{A_{RR}(X^tX)^{-1}(A_{RR})^t - A_{kq}(X^tX)^{-1}(A_{kq})^t\right\}$.

3.1. Bias comparison of AUKQE and KQE

This subsection compares the bias of the KQE and AUKQE. Predominately, the almost unbiased estimator always provides a smaller bias than biased estimator, but it does not give a minimum variance of the regression coefficient. Let $\hat{\gamma}$ be any type of an estimator of the parameter γ, then the squared bias (SB) of $\hat{\gamma}$ is specified as $\mathrm{SB}(\hat{\gamma}) = \{Bias(\hat{\gamma})\}^2$. Therefore, the bias and the SB of $\hat{\beta}_{kq}$ can be defined as:

$$SB(\hat{\beta}_{kq}) = (A_{kq} - I)\beta\beta^t(A_{kq} - I)^t = k^2(1+d)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2}, \tag{19}$$

where α_j^2 is defined as the *jth* element of $\Gamma^t\hat{\beta}_{OLS}$ and Γ is the eigenvector of the matrix X^tX such that $X^tX = \Gamma^t\Lambda\,\Gamma$, where $\Lambda = \mathrm{diag}(\lambda_j)$. Using the Eq. (19), the bias and SB of the AUKQE are defined as:

$$\begin{aligned}\mathrm{Bias}(\hat{\beta}_{aukq}) &= E(\hat{\beta}_{aukq}) - \beta \\ &= \{(2I - A_{kq})A_{kq} - I\}\beta \\ &= -k^2(1+d)^2(X^tX + kI)^{-2}\beta\end{aligned}$$

$$\mathrm{SB}(\hat{\beta}_{aukq}) = (I - A_{kq})^2\beta\beta^t\left((I - A_{kq})^2\right)^t = k^4(1+d)^4 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^4}.$$

One can compare the SB of the estimators by considering the SB differences between the estimators as $\Theta_1 = SB(\hat{\beta}_{kq}) - \mathrm{SB}(\hat{\beta}_{aukq}) > 0$:

$$\Theta_1 = k^2(1+d)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2} - k^4(1+d)^4 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^4}$$

$$\Theta_1 = \sum_{j=1}^{p} \frac{k^2(1+d)^2\alpha_j^2\{(\lambda_j + k)^2 - k^2(1+d)^2\}}{(\lambda_j + k)^4}.$$

where λ_j is the eigenvalue of the matrix X^tX, k and d are the shrinkage parameters. Reduction of bias in AUKQE is observed once we consider $\left|Bias(\hat{\beta}_{kq})_j\right| - \left|Bias(\hat{\beta}_{aukq})_j\right| = \frac{k(1+d)(\lambda_j - kd)}{(\lambda_j + k)^2}|\alpha_j|$. It can be easily seen that Θ_1 is positive since the expression $(\lambda_j + k)^2 - k^2(1+d)^2 > 0$ when $k > 0$ and $0 \le d \le 1$. Therefore, we can define that $SB(\hat{\beta}_{kq}) - \mathrm{SB}(\hat{\beta}_{aukq}) > 0$. Hence, the bias of KQE is higher than the bias of AUKQE. Therefore, based on the theoretical comparison, we can define the following theorem:

Theorem 5: *Under the CLRM, we have* $\left\|Bias(\hat{\beta}_{aukq})\right\|^2 < \left\|Bias(\hat{\beta}_{kq})\right\|^2$ *for* $k > 0$ and $0 \le d \le 1$ *if* $(\lambda_j + k)^2 > k^2(1+d)^2$.

4. Estimating methods for selecting k and d

The performance of the proposed estimator depends on the suitable value of shrinkage parameters *k* and *d*. Therefore, we derive optimal values of *k* and *d* and suggest an algorithm for the determination of *k* and *d*. The optimal value of *d* is obtained by taking the derivatives of $MSE(\hat{\beta}_{kq}) = \sum_{j=1}^{p} \frac{\sigma^2(1-\Phi_j)^2 + \lambda_j\Phi_j^2\alpha_j^2}{\lambda_j}$, where $\Phi_j = kd^*/(\lambda_j + k)$ and $d^* = d + 1$ with respect to *d* for fixed *k* as follows:

$$\frac{\partial MSE(\hat{\beta}_{kq})}{\partial d} = \sum_{j=1}^{p} \frac{2\alpha_j^2k^2\lambda_j(1+d) - 2k\sigma^2(\lambda_j - kd)}{\lambda_j(\lambda_j + k)^2}.$$

For $\frac{\partial MSE(\hat{\beta}_{kq})}{\partial d} = 0$, simplifying the numerator of the above expression and solving for *d* as:

$$d = \frac{\sum_{j=1}^{p}(\sigma^2 - \alpha_j^2k)}{\sum_{j=1}^{p}\left(\frac{k\sigma^2}{\lambda_j} + \alpha_j^2k\right)},$$

where σ^2 and α_j^2 are the unknown parameters and we replace these unknown parameters with their unbiased estimators and propose the following estimator:

$$\hat{d} = min\left(1, \frac{(\hat{\sigma}^2 - \hat{\alpha}_{min}^2 k)}{\left(\frac{k\hat{\sigma}^2}{\lambda_{min}} + \hat{\alpha}_{min}^2 k\right)}\right). \tag{20}$$

The condition $\lambda_j\hat{\sigma}^2 - \hat{\alpha}_j^2 k\lambda_j > 0$ should hold for the value of $\hat{d}$ to be positive and therefore, we propose the following restriction for $\hat{d}$ as

$$k^* = min\left(\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}\right)_{j=1}^{p} \tag{21}$$

The optimal value of k is determined by differentiating $MSE(\hat{\beta}_{kq})$ for k and equating it to be zero;

$$\frac{\partial MSE(\hat{\beta}_{kq})}{\partial k} = \sum_{j=1}^{p} \frac{2(1+d)\left[\left\{d\sigma^2 + \left(\alpha_j^2 d + \alpha_j^2\right)\lambda_j\right\}k - \lambda_j\sigma^2\right]}{(\lambda_j + k)^3},$$

$$k_j = \frac{\lambda_j\sigma^2}{d\sigma^2 + \alpha_j^2(1+d)\lambda_j}. \tag{22}$$

When $d = 0$, the expression, $k_j = \lambda_j\sigma^2/\left(d\sigma^2 + \alpha_j^2(1+d)\lambda_j\right)$ reduces to $k_j = \sigma^2/\alpha_j^2$, which is suggested by Hoerl and Kennard (1970) to estimate the ridge parameter k. It can be noted that the value of k_j is always positive. The expression in Eq. (22) depends on the unknown parameters σ^2 and α_j^2, and we replaced them by their corresponding unbiased estimators and proposed the following ridge estimator as:

$$\hat{k}_{opt} = min\left[\frac{\lambda_j\hat{\sigma}^2}{\hat{d}\hat{\sigma}^2 + \hat{\alpha}_j^2(1+\hat{d})\lambda_j}\right].$$

Following, Månsson et al. (2012) and Qasim et al. (2018 and 2020a & b), we propose the following estimators to estimate the value of *d*.

$$\hat{d}_{opt} = \sum_{j=1}^{p}\left(\frac{\sigma^2 - \alpha_j^2 k^*}{\frac{k^*\sigma^2}{\lambda_{max}} + \alpha_{max}^2 k^*}\right)\Big/p. \tag{23}$$

Finally, we define the following algorithm to determine the value of the biasing parameters, k and *d*:

Step 1: Compute the value of k^* using Eq. (21).

Step 2: Estimate $\hat{d}$ from Eq. (20) by using k^* in step 1.

Step 3: Calculate $\hat{d}_{opt}$ by substituting in the value of $\hat{k}_{opt}$ and k^*.

5. Simulation study

A simulation study is carried out to compare the finite sample properties of the proposed estimators with the existing estimators in different empirically relevant situations. The correlated explanatory variables are generated by following Gibbon (1981) and Kibria (2003) as follows:

$$x_{ij} = (1-\rho^2)^{1/2}\mathcal{Z}_{ij} + \rho\mathcal{Z}_{i(j+1)},$$

where $\mathcal{Z}_{ij}$ are the independent standard normal pseudo-random numbers and ρ is the degree of correlation between two regressors which is given by ρ^2. The performance of the proposed estimators depends on different factors such as sample size (n), degree of correlation (ρ) and value of the residual variance (σ^2). In the design of simulation, four different values of $\rho = 0.75, 0.90, 0.95, 0.99$, two different values of $\sigma^2 = 1{,}2$, and $n =$

25,50,100,200 are considered to judge the performance of the estimators. The response variable is generated as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_4 x_{i4} + \varepsilon_i, i = 1,2, \ldots, n$$

where y_i represent the nth observations of the dependent variable β_j are the regression coefficients and ε_i is the independent identically normally distributed error term with mean zero and variance σ^2. The values of β are chosen such that $\beta^t\beta = 1$ (see, e.g., Kibria, 2003). The AMSE is minimized when β is the normalized eigenvector corresponding to the largest eigenvalue of the matrix X^tX. Therefore, we selected slope parameters $\beta_j(\beta_1, \ldots, \beta_4)$ as the normalized eigenvector corresponding to the largest eigenvalue of the matrix X^tX. Besides, we assume zero intercept without loss of any generality. Then the variables are standardized so that X^ty represents the vector of correlations between the explanatory variables and the dependent variable. We use average mean squared error (AMSE) as a performance criterion. The AMSE of the estimator is determined based on 5000 replications and the entire process executed 5000 times to compute the simulated AMSE as follows:

$$AMSE(\hat{\beta}) = \frac{\sum_{r=1}^{5000}\left(\left(\hat{\beta} - \beta\right)^t\left(\hat{\beta} - \beta\right)\right)_r}{5000}$$

where $\hat{\beta}$ is any of the estimators of β in the rth replication. The simulation results are summarized in Figures 1-2 To demonstrate the finite sample properties of the estimators. We computed the AMSE of the OLSE, RRE, LRE, TPE, KQE and AUKQE under different situations that are common in a real-world application by changing the sample size (n), population variance (σ^2) and degree of correlation (ρ). Figure 1 shows the AMSE against different values of ρ and the AMSE against different values of n is shown in Figure 2. In almost all cases, the proposed class of KQE performed well. Though in some instances, the performance of the LRE is reasonably fair when n is small, and there is a limited number of explanatory variables. The LRE does not perform well when the n, ρ and σ^2 increase. As the parameters n, ρ^2 and σ^2 are increased in size; the relative performance of KQE is substantially improved. From simulation results, it can be seen that multicollinearity has a positive impact on AMSE. As the value of ρ increases, the AMSE is also increased. While the AMSE is decreased when n is increased. Based on the simulation results, we conclude that the KQE is performed considerably than the existing estimators.

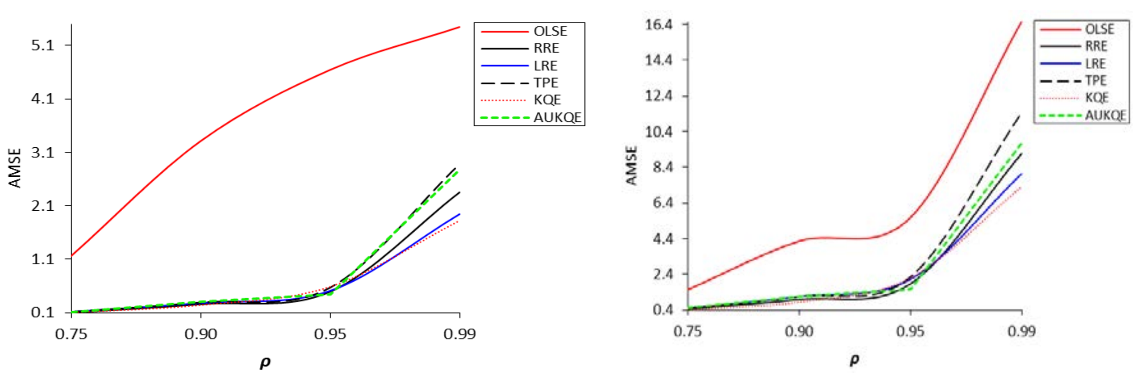

Figure 1: Left. Estimated AMSE of the estimators vs ρ when $n = 100$, $\sigma^2 = 1$. Right. Estimated AMSE of the estimators vs ρ when $n = 100$, $\sigma^2 = 2$

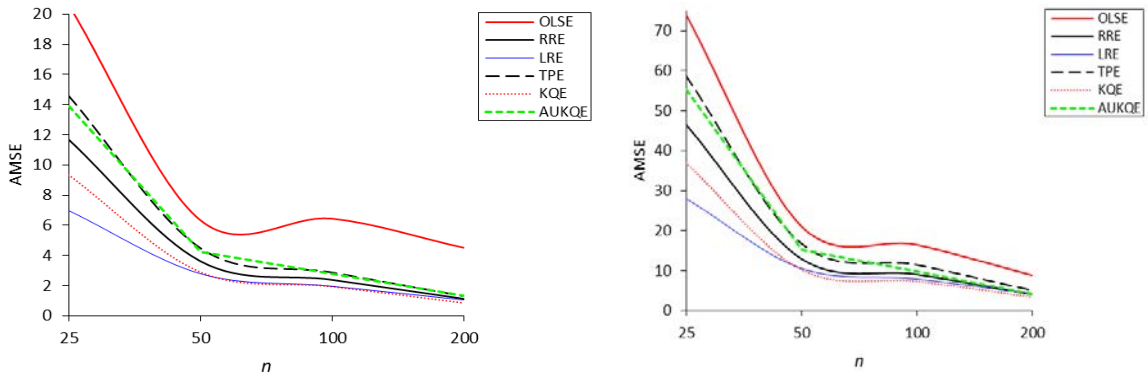

Figure 2: Left. Estimated AMSE of the estimators vs n when $\rho = 0.99$, $\sigma^2 = 1$. Right. Estimated AMSE of the estimators vs n when $\rho = 0.99$, $\sigma^2 = 2$

6. Numerical example

The Portland dataset, which was initially adopted by Woods et al. (1932) and also used in Lukman et al. (2019) is used as a first illustration in this paper to demonstrate the performance of the new estimator. The dependent variable is defined as the heat evolved after 180 days of curing measured in calories per gram of cement. This variable is modelled using four correlated explanatory variables corresponding to x_1 that represents tricalcium aluminate, x_2 that represents tricalcium silicate, x_3 that represents tetracalcium aluminoferrite, and x_4 that means β-dicalcium silicate. The condition index defined as $\sqrt{\lambda_{max}/\lambda_{min}}$ equals 6056, indicating a severe multicollinearity problem. Therefore, we conclude that the Portland dataset has a multicollinearity problem. The estimated coefficients and scalar MSE values are displayed in Table 1. We can see that among the unbiased and the almost unbiased estimators the OLSE performs the worst. There is a substantial decrease in the scalar MSE using the AUKQE as compared to the OLSE. Among the biased estimators, the LRE and TPE show the worst while the KQE has the lowest MSE. This result is in line with the simulated result since the LRE did not perform well when the σ^2 is large (in this application it is 5.98). Hence, the KQE and AUKQE are the best options among the biased and the almost unbiased estimators, respectively.

Table 1: Estimated coefficients and MSE of Portland cement dataset†

Estimators	$\hat{\beta}_o$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	MSE
OLSE	62.4054	1.5511	0.5102	0.1019	-0.1441	4912.09
RRE	42.9860	1.7509	0.7103	0.3062	0.0521	2706.36
LRE	49.9266	1.6767	0.6394	0.2312	-0.0176	3298.65
TPE	27.4575	1.9106	0.8703	0.4696	0.2090	4333.39
AUKQE	38.0793	1.8013	0.7609	0.3579	0.1586	2694.80
KQE	27.4575	1.9106	0.8703	0.4696	0.2090	2171.01

†Note: The eigenvalues of the matrix X^tX: 44676.21, 5965.42, 809.95, 105.42, 0.0012; The OLSE of σ^2: $\hat{\sigma}^2 = 5.98$; The condition index (CI): $CI = \sqrt{\lambda_{max}/\lambda_{min}} = 6056.3443$.

7. Conclusions

This article introduces biased and almost unbiased two-parameter estimators. Proposed estimator includes the OLSE and RRE as special cases to be used to achieve the minimum bias. The proposed estimators are compared theoretically with the OLSE, RRE, LRE and TPE in the sense of MMSE. Our proposed estimator has an advantage over the existing estimators since it exhibits the minimum MMSE under certain conditions. The KQE exhibit a minimum variance and scalar MSE compared to the TPE suggested by Özkale and Kaciranlar (2007) under certain conditions. Though, the performance of the proposed estimator depends on the appropriate selection of shrinkage parameters *k* and *d*. Therefore, we suggest an algorithm for selecting the shrinkage parameters. Based on the theoretical comparisons, simulation results, and empirically relevant real-world application, we conclude that the KQE is performed considerably better than the OLSE, RRE, LRE and the TPE. Therefore, this estimator can be recommended for practitioners.

References

1. Akdeniz, F., & Erol, H. (2003). Mean squared error matrix comparisons of some biased estimators in linear regression. Communications in Statistics-Theory and Methods, 32(12), 2389-2413.
2. Amin, M., Qasim, M., Amanullah, M., & Afzal, S. (2020). Performance of some ridge estimators for the gamma regression model. Statistical papers, 61(3), 997-1026.
3. Farebrother, R. W. (1976). Further results on the mean square error of ridge regression. Journal of the Royal Statistical Society. Series B (Methodological), 38(3), 248-250.
4. Gibbons, D. G. (1981). A simulation study of some ridge estimators. Journal of the American Statistical Association, 76(373), 131-139.
5. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.
6. Hubert, M. H., & Wijekoon, P. (2006). Improvement of the Liu estimator in linear regression model. Statistical Papers, 47(3), 471.
7. Hubert, M. H., & Wijekoon, P. (2006). Improvement of the Liu estimator in linear regression model. Statistical Papers, 47(3), 471.
8. Kaçiranlar, S., Sakallioğlu, S., Akdeniz, F., Styan, G. P., & Werner, H. J. (1999). A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on Portland Cement. Sankhyā: The Indian Journal of Statistics, Series B, 443-459.
9. Kaçiranlar, S., Sakallioğlu, S., Akdeniz, F., Styan, G. P., & Werner, H. J. (1999). A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on Portland Cement. Sankhyā: The Indian Journal of Statistics, Series B, 443-459.
10. Kejian, L. (1993). A new class of blased estimate in linear regression. Communications in Statistics-Theory and Methods, 22(2), 393-402.
11. Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. Communications in Statistics-Simulation and Computation, 32(2), 419-435.
12. Liu, K. (2003). Using Liu-type estimator to combat collinearity. Communications in Statistics-Theory and Methods, 32(5), 1009-1020.
13. Lukman, A. F., Ayinde, K., & Ajiboye, A. S. (2017). Monte Carlo study of some classification-based ridge parameter estimators. Journal of Modern Applied Statistical Methods, 16(1), 24.
14. Lukman, A. F., Ayinde, K., Binuomote, S., & Clement, O. A. (2019). Modified ridge-type estimator to combat multicollinearity: Application to chemical data. Journal of Chemometrics, 33(5), e3125.

15. Månsson, K., Kibria, B. M. G., & Shukur, G. (2012). On Liu estimators for the logit regression model. Economic Modelling, 29(4), 1483-1488.
16. Månsson, K., Shukur, G., & Kibria, B. M. G. (2010). A simulation study of some ridge regression estimators under different distributional assumptions. Communications in Statistics-Simulation and Computation, 39(8), 1639-1670.
17. McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association, 70(350), 407-416.
18. Ohtani, K. (1986). On small sample properties of the almost unbiased generalized ridge estimator. Communications in Statistics-Theory and Methods, 15(5), 1571-1578.
19. Özkale, M. R., & Kaciranlar, S. (2007). The restricted and unrestricted two-parameter estimators. Communications in Statistics—Theory and Methods, 36(15), 2707-2725.
20. Qasim, M., Amin, M., & Amanullah, M. (2018). On the performance of some new Liu parameters for the gamma regression model. Journal of Statistical Computation and Simulation, 88(16), 3065-3080.
21. Qasim, M., Amin, M., & Omer, T. (2020a). Performance of some new Liu parameters for the linear regression model. Communications in Statistics-Theory and Methods, 49(17), 4178-4196.
22. Qasim, M., Kibria, B. M. G., Månsson, K., & Sjölander, P. (2020b). A new Poisson Liu Regression Estimator: method and application. Journal of Applied Statistics, 47(12), 2258-2271.
23. Saleh, A. K. Md. E., Arashi, M., and Kibria, B. M. G. (2019). Theory of Ridge Regression Estimation with Applications. Wiley, New York.
24. Trenkler, G., & Toutenburg, H. (1990). Mean squared error matrix comparisons between biased estimators—An overview of recent results. Statistical Papers, 31(1), 165.
25. Woods, H., Steinour, H. H., & Starke, H. R. (1932). Effect of composition of Portland cement on heat evolved during hardening. Industrial & Engineering Chemistry, 24(11), 1207-1214.
26. Wu, J., & Yang, H. (2013). Efficiency of an almost unbiased two-parameter estimator in linear regression model. Statistics, 47(3), 535-545.