Contraction ridge estimator: Simulation and Application to Economic Data

Adewale F. Lukman[1*†] and Essoham Ali[2,3†]

[1]Department of Mathematics and Statistics, University of North Dakota, Grand Forks, North Dakota, 58202 USA.
[2]Univ Bretagne Sud, CNRS UMR 6205, LMBA, Vannes, France.
[3]Institut de Mathématiques Appliquées, UCO, 49000, Angers, France.

*Corresponding author(s). E-mail(s): adewale.lukman@und.edu;
Contributing authors: essoham.ali@univ-ubs.fr ;
†These authors contributed equally to this work.

Abstract

This paper introduces a novel regularization technique known as the Contraction Ridge estimator (CRidge), designed to address the limitations of traditional Least Squares (LS) and other biased estimation methods such as Ridge, Liu, Kibria-Lukman (KL), and Contraction Least Squares (COLS) estimators, particularly in the presence of multicollinearity. The proposed estimator modifies the COLS by integrating ridge regression, enhancing its numerical stability and performance in high-dimensional and highly collinear settings. Theoretical comparisons are presented, establishing conditions under which CRidge outperforms other estimators based on Mean Squared Error Matrix (MSEM) and Scalar Mean Squared Error (SMSE) criteria. A comprehensive Monte Carlo simulation study further validates the theoretical findings, demonstrating the superiority of CRidge over LS, Ridge, Liu, KL and COLS estimators across various scenarios, including different levels of multicollinearity, sample sizes, and noise levels. In addition, an empirical application using the electricity data illustrates the practical utility of the CRidge estimator. The results show that CRidge consistently achieves lower SMSE, Prediction Mean Squared Error (PMSE) and Prediction Mean Absolute Error (PMAE) compared to other methods, indicating its robustness and effectiveness for regression analysis under multicollinearity. The contraction ridge estimator is recommended as a reliable and efficient tool to improve estimation accuracy and prediction stability in complex regression problems.

Keywords: Contraction Ridge Estimator; Ridge regression; Multicollinearity; Regularization Techniques; Prediction Accuracy.

1 Introduction

Linear regression model estimation is a fundamental area in statistics, extensively studied due to its applicability in various scientific fields. The standard Linear Regression Model (LRM) is commonly used to express a response variable as a function of multiple predictors, and the model parameters are generally estimated using the Least Squares (LS) method. However, this method is known for its limitations when predictors are highly correlated (multicollinearity), making the predictor matrix singular or nearly singular, thus producing unstable estimates with high variance [8].

The increasing complexity of modern data analysis has motivated the development of various estimation techniques aimed at enhancing robustness and accuracy in regression modeling. In particular, when dealing with high-dimensional data or datasets with strong multicollinearity, traditional methods such as Least Squares (LS) often fail to produce reliable estimates. This inadequacy has prompted researchers to explore alternative methods that incorporate regularization techniques to stabilize the estimation process and reduce the adverse effects of multicollinearity.

Moreover, the development of efficient estimators has become a prominent area of research, especially in fields where predictive accuracy and stability are essential. For example, in economics, environmental modeling, and biomedical research, multicollinearity is mostly inevitable due to the inherent relationships among explanatory variables. Addressing this challenge requires estimators that not only minimize prediction error but also maintain consistency under varying levels of multicollinearity. Recent advances, such as the Kibria-Lukman estimator [13] and other shrinkage-based methods, have shown the potential to improve estimation performance; however, there is still room for improvement.

Ridge regression, introduced by Hoerl and Kennard [8], is one of the most popular shrinkage-based methods. It involves adding a penalty term to the Gram matrix of the classical LS method to reduce instability caused by multicollinearity. However, this technique presents drawbacks related to the selection of the regularization parameter, as noted by Liu [15]. To address this issue, Liu [15] proposed the Liu estimator, which aims to improve the performance of ridge regression by appropriately adjusting this parameter.

More recently, [13] developed a new regularization technique called the Kibria-Lukman (KL) estimator. This estimator is based on the modification of the ridge estimator to enhance its robustness in multicollinearity. However, the improvement provided by this technique remains limited in certain situations, which justifies the search for alternative approaches.

A potential alternative is the contraction least squares (COLS) estimator, introduced by [22]. Although this concept has been mentioned in the literature, it has received relatively limited attention compared to the ridge estimator. In this paper, we propose a modified version of the contraction estimator, called the Contraction Ridge estimator (CRidge), which combines the characteristics of Ridge and Contraction estimators to improve numerical stability and estimation performance when multicollinearity is present.

The main objective of this paper is to evaluate the theoretical properties of the Contraction Ridge estimator and compare it to other existing estimators, including the least squares estimator, the Ridge estimator, the Liu estimator, the Kibria-Lukman estimator, and the classical contraction estimator. To achieve this, we used an approach based on the Mean Squared Error Matrix (MSEM) and the Scalar Mean Squared Error (SMSE) to compare the performance of these estimators.

The following sections of this paper are organized as follows: Section 2 presents the linear regression methods and model estimation, as well as the biased estimators proposed in the literature. Section 3 focuses on the theoretical comparison between the different estimators. Section 4 presents the simulation methodology and the results obtained. Section 5 provides an illustrative application based on a real dataset. As in other studies, the discussion and conclusions are provided at the end of this paper.

2 Linear regression and model estimation

The linear regression model (LRM) expresses a single response variable as a linear function of predictors. The general LRM is given by:

$$y = \theta_0 + \theta_1 z_1 + \cdots + \theta_q z_q + \varepsilon. \tag{1}$$

where y is the response variable, z are the predictors, $\theta_0, \theta_1, \ldots, \theta_q$ are the unknown regression parameters, and ε are the random error terms. In matrix notation, the model becomes:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_q \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Consequently,

$$\mathbf{y}_{(n\times 1)} = \mathbf{Z}_{(n\times(q+1))}\boldsymbol{\theta}_{((q+1)\times 1)} + \boldsymbol{\varepsilon}_{(n\times 1)}. \tag{2}$$

such that $E(\boldsymbol{\varepsilon}) = 0$, $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 I$ and $\mathrm{Cov}(\varepsilon_j, \varepsilon_k) = 0, \quad j \neq k$.

The regression parameters in equation (2) are estimated using the least squares (LS) method. The method minimizes the error sum of squares. LS is defined as follows:

$$\widehat{\boldsymbol{\theta}}_{LS} = (\mathbf{Z}^\top \mathbf{Z})^{-1}\mathbf{Z}^\top y \tag{3}$$

Model (1) can be written in canonical form as follows:

$$y = X\nu + \varepsilon, \tag{4}$$

where $X = ZH$, $\nu = H'\theta$, and H is the orthogonal matrix whose columns constitute the eigenvectors of $Z'Z$. Then

$$X'X = H'Z'ZH = E = \mathrm{diag}(e_1, \ldots, e_q),$$

where $e_1 \geq e_2 \geq \cdots \geq e_q > 0$ are the ordered eigenvalues of $Z'Z$. Hence, (3) is canonically represented as:

$$\hat{\nu_{LS}} = E^{-1}X'y \tag{5}$$

The properties of the estimator considered in this study are evaluated using the mean squared error matrix (MSEM) and the scalar mean squared error (SMSE). MSEM of an estimator $\tilde{\tau}$ is defined as

$$\mathrm{MSEM}(\tilde{\tau}) = V(\tilde{\tau}) + \mathrm{Bias}(\tilde{\tau})\mathrm{Bias}(\tilde{\tau})', \tag{6}$$

where $V(\tilde{\tau})$ is the variance-covariance matrix and $\mathrm{Bias}(\tilde{\tau}) = \mathbb{E}(\tilde{\tau}) - \tau$ is the bias vector. The SMSE is the trace of the MSEM, defined by:

$$\mathrm{SMSE}(\tilde{\tau}) = \mathrm{tr}(V(\tilde{\tau})) + \mathrm{Bias}(\tilde{\tau})'\mathrm{Bias}(\tilde{\tau}). \tag{7}$$

Based on the canonical transformation, the following relationships hold:

$$\tilde{\nu} = H'\tilde{\theta}, \quad \mathrm{SMSE}(\tilde{\nu}) = \mathrm{MSE}(\tilde{\theta}) \quad \text{and} \quad \mathrm{MSEM}(\tilde{\nu}) = H'\mathrm{MSEM}(\tilde{\theta})H. \tag{8}$$

where $\tilde{\nu}$ can be any estimators considered in this study. Hence, the MSEM and SMSE for ν_{LS} are given by:

$$MSEM(\hat{\nu}_{LS}) = \hat{\sigma}^2 E^{-1} \tag{9}$$

$$SMSE(\hat{\nu}_{LS}) = \hat{\sigma}^2 \sum_{j=1}^{q} \frac{1}{\lambda_j}, \tag{10}$$

where

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}.$$

2.1 Biased Estimators

Studies have revealed the limitation of LS when there are linear relationships among the predictors (multicollinearity). For example, when the predictors have perfect multicollinearity, the Gram matrix $\mathbf{Z}^\top\mathbf{Z}$ becomes non-invertible. Regularization techniques such as ridge regression are the potential alternatives. [8] developed ridge regression by incorporating a penalty into the $\mathbf{Z}^\top\mathbf{Z}$ matrix in (3) to reduce the influence of multicollinearity. The ridge estimator of ν is given by

$$\hat{\nu}_k = E_k^{-1}E\nu_{LS} \tag{11}$$

where $E_k^{-1} = (E + kI)^{-1}$ k is the regularization parameter. [9] defined the regularization parameter as follows:

$$k = \frac{p\sigma^2}{\sum_{j=1}^{q} v_j^2} \tag{12}$$

The MSEM and SMSE for $\hat{\nu}^k$ are defined as follows:

$$MMSE(\hat{\nu}_k) = \hat{\sigma}^2 E_k^{-1} E E_k^{-1} + k^2 E_k^{-1} \nu_{LS} \nu_{LS}^{'} E_k^{-1} \tag{13}$$

$$SMSE(\hat{\nu}_k) = \hat{\sigma}^2 \sum_{j=1}^{q} \frac{e_j}{(e_j + k)^2} + k^2 \sum_{j=1}^{q} \frac{\nu_j^2}{(e_j + k)^2} \tag{14}$$

Liu [15] argued that the ridge estimator has a drawback associated with its regularization parameter. To address this issue, [15] proposed the Liu estimator, which is defined as follows:

$$\hat{\nu}_d = E_d^{-1} E_{d+} \hat{\nu}_{LS}, \quad 0 \leq d < 1 \tag{15}$$

where $E_d^{-1} = (E + I)^{-1}$ and $E_{d+} = (E + dI)$. Hence, MSEM and SMSE for $\hat{\nu}_d$ are defined as follows:

$$MSEM(\hat{\nu}_d) = \hat{\sigma}^2 E_d^{-1} E_{d+} E^{-1} E_{d+} E_d^{-1} + (d-1)^2 E_d^{-1} \nu_{LS} \nu_{LS}' E_d^{-1} \tag{16}$$

$$SMSE(\hat{\nu}_d) = \hat{\sigma}^2 \sum_{j=1}^{q} \frac{(e_j + d)^2}{(e_j + 1)^2 e_j} + (d-1)^2 \sum_{j=1}^{q} \frac{\nu_j^2}{(e_j + 1)^2} \tag{17}$$

Kibria and Lukman [13] recently developed another regularization technique called the Kibria-Lukman estimator. The estimator is defined as follows:

$$\hat{\nu}_{KL} = E_k^{-1} E_a \hat{\nu}_{LS}, \tag{18}$$

where $E_a = (E - kI)$. Hence, MSEM and SMSE for $\hat{\nu}_{KL}$ are defined as follows:

$$MSEM(\hat{\nu}_{KL}) = \hat{\sigma}^2 E_k^{-1} E_a E^{-1} E_a E_k^{-1} + 4k^2 E_k^{-1} \nu_{LS} \nu_{LS}' E_k^{-1} \tag{19}$$

$$SMSE(\hat{\nu}_{KL}) = \hat{\sigma}^2 \sum_{j=1}^{q} \frac{(e_j - k)^2}{(e_j + k)^2 e_j} + 4k^2 \sum_{j=1}^{q} \frac{\nu_j^2}{(e_j + k)^2} \tag{20}$$

2.2 Contraction Ridge Estimator

Mayer and Willke [22] introduced the contraction estimator as an alternative to the widely studied ridge estimator. However, despite its potential, the contraction estimator has received limited attention in the literature. In this study, we propose a modified version of the contraction estimator and investigate its theoretical properties. According to Özkale and Kaciranlar [24], the simplified form of the contraction estimator is as follows:

$$\hat{\nu}_c = (1 + \rho)^{-1} \hat{\nu}_{LS}, \quad \rho > 0$$

This formulation effectively shrinks each component of the LS method by a factor of $(1+\rho)^{-1}$, offering a regularization approach similar to ridge regression. For computational convenience, we redefine ρ as k and determine its value using the ridge parameter of (12). Consequently, the contraction estimator is given by

$$\hat{\nu}_c = (1 + k)^{-1} \hat{\nu}_{LS}, \quad k > 0 \tag{21}$$

Notably, as $k \to 0$, the contraction estimator converges to the LS method:

$$\lim_{k \to 0} \hat{\nu}_c = \hat{\nu}_{LS}$$

The MSEM of $\hat{\nu}_c$ is as follows:

$$MSEM(v_c) = \hat{\sigma}^2 (1 + k)^{-2} E^{-1} + k^2 (1 + k)^{-2} v v' \tag{22}$$

$$SMSE(v_c) = \frac{\hat{\sigma}^2}{(1 + k)^2} \sum_{j=1}^{q} \frac{1}{e_j} + \frac{k^2}{(1 + k)^2} \sum_{j=1}^{q} \nu_j^2 \tag{23}$$

Since the contraction estimator exhibits similar behaviour to ridge regression, we propose a modification to enhance its applicability, particularly in multicollinearity. The original contraction estimator is a function of the LS estimator, which becomes unstable or noninvertible under perfect multicollinearity. To address this limitation, we replace the LS estimator with the ridge estimator, leading to the contraction ridge estimator:

$$\hat{\nu}_{ck} = (1+k)^{-1}\hat{\nu}k, \quad k > 0 \tag{24}$$

This modification ensures numerical stability and improves the estimator performance in highly collinear or high-dimensional settings. The properties of the contraction ridge estimator are defined as follows: *Bias* of v_{ck} is given as:

$$Bias(v_{ck}) = (1+k)^{-1}E_k^{-1}Av$$

where $A = (E - (1+k)E_k)$. The variance is given as:

$$V(v_{ck}) = \sigma^2(1+k)^{-2}E_k^{-1}EE_k^{-1}$$

Hence, MSEM is defined as

$$MSEM(v_{ck}) = \hat{\sigma}^2(1+k)^{-2}E_k^{-1}EE_k^{-1} + (1+k)^{-2}E_k^{-1}Avv'AE_k^{-1} \tag{25}$$

Consequently, the SMSE is as follows:

$$SMSE(v_{ck}) = \frac{\hat{\sigma}^2}{(1+k)^2}\sum_{j=1}^{q}\frac{e_j}{(e_j+k)^2} + \frac{k^2}{(1+k)^2}\sum_{j=1}^{q}\frac{(1+e_j+k)^2\nu_j^2}{(e_j+k)^2} \tag{26}$$

3 Theoretical Comparison

In this section, we compare the performance of the existing methods with the proposed method.

Lemma 1. (Farebrother [10]). *Let F be a positive definite matrix, namely $F > 0$, and let ν be some vector, then*

$$F - \nu\nu' \geq 0 \text{ if and only if } \nu' F^{-1}\nu \leq 1.$$

Lemma 2. (Trenkler and Toutenburg [25]). *Let $\hat{\theta}_j = C_j y$, $j = 1, 2$ be two competing estimators for θ. Suppose that $V = V(\hat{\theta}_1) - V(\hat{\theta}_2) > 0$, where $V(\hat{\theta}_j)$, $j = 1, 2$ denotes the variance-covariance matrix of $\hat{\theta}_j$. Then*

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

if and only if

$$m_2'(V + m_1m_1')^{-1}m_2 \leq 1,$$

where $MSEM(\hat{\theta}_j)$ and m_j denote the mean squared error matrix and bias vector of $\hat{\theta}_j$, respectively.

3.1 $\hat{\nu}_{LS}$ and $\hat{\nu}_{ck}$

We examined the difference between the MSEM of the LS method and the contraction ridge estimator as follows:

$$MSEM(\hat{\nu}_{LS}) - \text{MSEM}(\hat{\nu}_{ck}) = \sigma^2(E^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1}) - (1+k)^{-2}(E_k^{-1}A\nu_{LS}\nu_{LS}^{'}A^{'}E_k^{-1}). \tag{27}$$

Given that $k > 0$, we have the following theorem.

Theorem 1. *Given two linear estimators $\hat{\nu}_{LS}$ and $\hat{\nu}_{ck}$. If $k > 0$, $\hat{\nu}_{ck}$ dominates $\hat{\nu}_{LS}$ given that $MSEM(\hat{\nu}_{LS}) - MSEM(\hat{\nu}_{ck}) > 0$ if and only if*

$$\sigma^{-2}\nu_{LS}'A^{'}[(1+k)^2E_kEE_k - E]^{-1}A\nu_{LS} \leq 1.$$

Proof. Using the SMSE for both estimators in (10) and (26), we obtain

$$V(\hat{\nu}_{LS}) - V(\hat{\nu}_{ck}) = \sigma^2(E^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1})$$

$$= \sigma^2\text{diag}\left\{\frac{1}{e_j} - \frac{e_j}{(1+k)^2(e_j+k)^2}\right\}_{j=1}^{q}.$$

The variance-covariance difference $E^{-1}-(1+k)^{-2}E_k^{-1}EE_k^{-1}$ will be positive definite (pd) if and only if

$$(1+k)^2(e_j+k)^2-e_j^2>0 \quad \text{or} \quad (1+k)(e_j+k)-e_j>0.$$

For $k>0$, we can certainly see that $(1+k)(e_j+k)-e_j>0$. Therefore, $E^{-1}-(1+k)^2E_k^{-1}EE_k^{-1}$ is pd. The proof is completed by Lemma 2.

3.2 $\hat{\nu}_k$ and $\hat{\nu}_{ck}$

We examined the difference between the MSEM of the Ridge and the contraction ridge estimator as follows:

$$\begin{aligned} MSEM(\hat{\nu}_k)-\text{MSEM}(\hat{\nu}_{ck}) =& \sigma^2(E_k^{-1}EE_k^{-1}-(1+k)^{-2}E_k^{-1}EE_k^{-1})+k^2E_k^{-1}\nu_{LS}\nu_{LS}^{'}E_k^{-1} \\ &-(1+k)^{-2}(E_k^{-1}A\nu_{LS}\nu_{LS}^{'}A^{'}E_k^{-1}). \end{aligned} \tag{28}$$

Given that $k>0$, we have the following theorem.

Theorem 2. *Given two linear estimators $\hat{\nu}_k$ and $\hat{\nu}_{ck}$. If $k>0$, $\hat{\nu}_{ck}$ dominates $\hat{\nu}_k$ given that $MSEM(\hat{\nu}_k)-MSEM(\hat{\nu}_{ck})>0$ if and only if*

$$\sigma^{-2}\nu_{LS}'A^{'}[(1+k)^2E-E+\sigma^{-2}k^2(1+k)^2\nu_{LS}\nu_{LS}]A\nu_{LS}\leq 1.$$

Proof. Using the SMSE for both estimators in (14) and (26), we obtain

$$\begin{aligned} V(\hat{\nu}_k)-V(\hat{\nu}_{ck}) &= \sigma^2(E_k^{-1}EE_k^{-1}-(1+k)^{-2}E_k^{-1}EE_k^{-1}) \\ &= \sigma^2\text{diag}\left\{\frac{e_j}{(e_j+k)^2}-\frac{e_j}{(1+k)^2(e_j+k)^2}\right\}_{j=1}^{q}. \end{aligned}$$

The variance-covariance difference $E_k^{-1}EE_k^{-1}-(1+k)^{-2}E_k^{-1}EE_k^{-1}$ will be positive definite (pd) if and only if

$$(1+k)^2e_j-e_j>0 \quad \text{or} \quad e_j[(1+k)^2-1]>0.$$

For $k>0$, we can certainly see that $e_j[(1+k)^2-1]>0$. Therefore, $E_k^{-1}EE_k^{-1}-(1+k)^{-2}E_k^{-1}EE_k^{-1}$ is pd. The proof is completed by Lemma 2.

3.3 $\hat{\nu}_d$ and $\hat{\nu}_{ck}$

We examined the difference between the MSEM of the Liu and the contraction ridge estimator as follows:

$$\begin{aligned} MSEM(\hat{\nu}_d)-\text{MSEM}(\hat{\nu}_{ck}) = & \sigma^2\Big(E_d^{-1}E_{d+}E^{-1}E_{d+}E_d^{-1}-(1+k)^{-2}E_k^{-1}EE_k^{-1}\Big) \\ &+(d-1)^2E_d^{-1}\nu_{LS}\nu_{LS}^{'}E_d^{-1} \\ &-(1+k)^{-2}\Big(E_k^{-1}A\nu_{LS}\nu_{LS}^{'}A^{'}E_k^{-1}\Big). \end{aligned} \tag{29}$$

Given that $k>0$, and $d>0$ we have the following theorem.

Theorem 3. *Given two linear estimators $\hat{\nu}_d$ and $\hat{\nu}_{ck}$. If $k>0$, $\hat{\nu}_{ck}$ dominates $\hat{\nu}_d$ given that $MSEM(\hat{\nu}_d)-MSEM(\hat{\nu}_{ck})>0$ if and only if*

$$(1+k)^{-2}\nu_{LS}'A^{'}E_k^{-1}[\sigma^2(E_d^{-1}E_{d+}E^{-1}E_{d+}E_d^{-1}-(1+k)^{-2}E_k^{-1}EE_k^{-1}+(d-1)^2E_d^{-1}\nu_{LS}\nu_{LS}^{'}E_d^{-1}]^{-1}E_k^{-1}A\nu_{LS}\leq 1.$$

Proof. Using the SMSE for both estimators in (17) and (26), we obtain

$$\begin{aligned} V(\hat{\nu}_d)-V(\hat{\nu}_{ck}) &= \sigma^2(E_d^{-1}E_{d+}E^{-1}E_{d+}E_d^{-1}-(1+k)^{-2}E_k^{-1}EE_k^{-1}) \\ &= \sigma^2\text{diag}\left\{\frac{(e_j+d)^2}{e_j(e_j+1)^2}-\frac{e_j}{(1+k)^2(e_j+k)^2}\right\}_{j=1}^{q}. \end{aligned}$$

The variance-covariance difference $E_d^{-1}E_{d+}E^{-1}E_{d+}E_d^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1}$ will be positive definite (pd) if and only if

$$(1+k)^2(e_j+d)^2(e_j+k)^2 - e_j^2(e_j+1)^2 > 0 \quad \text{or} \quad (1+k)(e_j+d)(e_j+k) - e_j(e_j+1) > 0.$$

For $k > 0$, we can certainly see that $(1+k)(e_j+d)(e_j+k) - e_j(e_j+1) > 0$. Therefore, $E_d^{-1}E_{d+}E^{-1}E_{d+}E_d^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1}$ is pd. The proof is completed by Lemma 2.

3.4 $\hat{\nu}_{KL}$ and $\hat{\nu}_{ck}$

We examined the difference between the MSEM of the Liu and the contraction ridge estimator as follows:

$$\begin{aligned} MSEM(\hat{\nu}_{KL}) - \text{MSEM}(\hat{\nu}_{ck}) = \sigma^2 & \Big(E_k^{-1}E_aE^{-1}E_aE_k^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1}\Big) \\ & + 4k^2E_k^{-1}\nu_{LS}\nu_{LS}^{'}E_k^{-1} \\ & - (1+k)^{-2}\Big(E_k^{-1}A\nu_{LS}\nu_{LS}^{'}A^{'}E_k^{-1}\Big). \end{aligned} \tag{30}$$

Given that $k > 0$, we have the following theorem.

Theorem 4. *Given two linear estimators $\hat{\nu}_{KL}$ and $\hat{\nu}_{ck}$. If $k > 0$, $\hat{\nu}_{ck}$ dominates $\hat{\nu}_{KL}$ given that $MSEM(\hat{\nu}_{KL}) - MSEM(\hat{\nu}_{ck}) > 0$ if and only if*

$$(1+k)^{-2}\nu_{LS}'A^{'}E_k^{-1}[\sigma^2(E_k^{-1}E_aE^{-1}E_aE_k^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1} + 4k^2E_k^{-1}\nu_{LS}\nu_{LS}^{'}E_k^{-1}]^{-1}E_k^{-1}A\nu_{LS} \leq 1.$$

Proof. Using the SMSE for both estimators in (20) and (26), we obtain

$$\begin{aligned} V(\hat{\nu}_{KL}) - V(\hat{\nu}_{ck}) &= \sigma^2(E_k^{-1}E_aE^{-1}E_aE_k^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1}) \\ &= \sigma^2\text{diag}\left\{\frac{(e_j-k)^2}{e_j(e_j+k)^2} - \frac{e_j}{(1+k)^2(e_j+k)^2}\right\}_{j=1}^{q}. \end{aligned}$$

The variance-covariance difference $E_k^{-1}E_aE^{-1}E_aE_k^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1}$ will be positive definite (pd) if and only if

$$(1+k)^2(e_j-k)^2 - e_j^2 > 0 \quad \text{or} \quad (1+e_j-k) > 0.$$

For $k > 0$, there is possibility that $(1+e_j-k) > 0$. Therefore, $E_k^{-1}E_aE^{-1}E_aE_k^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1}$ is pd. The proof is completed by Lemma 2.

3.5 $\hat{\nu}_c$ and $\hat{\nu}_{ck}$

We examined the difference between the MSEM of the Contraction method and the contraction ridge estimator as follows:

$$\begin{aligned} MSEM(\hat{\nu}_c) - \text{MSEM}(\hat{\nu}_{ck}) = \sigma^2 & ((1+k)^{-2}E^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1}) \\ & + k^2(1+k)^{-2}\nu_{LS}\nu_{LS} - k^2(1+k)^{-2}(E_k^{-1}A\nu_{LS}\nu_{LS}^{'}A^{'}E_k^{-1}). \end{aligned} \tag{31}$$

Given that $k > 0$, we have the following theorem.

Theorem 5. *Given two linear estimators $\hat{\nu}_c$ and $\hat{\nu}_{ck}$. If $k > 0$, $\hat{\nu}_{ck}$ dominates $\hat{\nu}_c$ given that $MSEM(\hat{\nu}_c) - MSEM(\hat{\nu}_{ck}) > 0$ if and only if*

$$(1+k)^{-2}\nu_{LS}'A^{'}E_k^{-1}[\sigma^2(1+k)^{-2}(E^{-1} - E_k^{-1}EE_k^{-1} + k^2\nu_{LS}\nu_{LS}]^{-1}E_k^{-1}A\nu_{LS} < 1.$$

Proof. Using the SMSE for both estimators in (23) and (26), we obtain

$$V(\hat{\nu}_c) - V(\hat{\nu}_{ck}) = \sigma^2(1+k)^{-2}(E^{-1} - E_k^{-1}EE_k^{-1})$$

$$= \sigma^2(1+k)^{-2}\text{diag}\left\{\frac{1}{e_j} - \frac{e_j}{(e_j+k)^2}\right\}_{j=1}^{q}.$$

The variance-covariance difference $E^{-1} - E_k^{-1}EE_k^{-1}$ will be positive definite (pd) if and only if

$$(e_j+k)^2 - e_j^2 > 0 \quad \text{or} \quad (e_j+k) - e_j > 0.$$

For $k > 0$, we can certainly see that $(e_j+k) - e_j = k > 0$. Therefore, $E^{-1} - E_k^{-1}EE_k^{-1}$ is pd. The proof is completed by Lemma 2.

3.6 Theoretical Validation

We validate the practical applicability of our theoretical findings using real-world data, ensuring their relevance beyond simulated scenarios. A comprehensive description of the data is provided in Section 5.

Table 1: Theoretical conditions and computed values for different theorems.

Theorem	Theoretical Conditions	Values
1	$\sigma^{-2}\nu_{LS}'A'[(1+k)^2E_kEE_k - E]^{-1}A\nu_{LS} \leq 1$	0.0772
2	$\sigma^{-2}\nu_{LS}'A'[(1+k)^2E - E + \sigma^{-2}k^2(1+k)^2\nu_{LS}\nu_{LS}]A\nu_{LS} \leq 1$	0.1282
3	$(1+k)^{-2}\nu_{LS}'A'E_k^{-1}[\sigma^2(E_d^{-1}E_{d+}E^{-1}E_{d+}E_d^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1} + (d-1)^2E_d^{-1}\nu_{LS}\nu_{LS}'E_d^{-1}]^{-1}E_k^{-1}A\nu_{LS} \leq 1$	5.4e-05
4	$(1+k)^{-2}\nu_{LS}'A'E_k^{-1}[\sigma^2(E_k^{-1}E_aE^{-1}E_aE_k^{-1} - (1+k)^{-2}E_k^{-1}EE_k^{-1} + 4k^2E_k^{-1}\nu_{LS}\nu_{LS}'E_k^{-1}]^{-1}E_k^{-1}A\nu_{LS} \leq 1$	0.1291
5	$(1+k)^{-2}\nu_{LS}'A'E_k^{-1}[\sigma^2(1+k)^{-2}(E^{-1} - E_k^{-1}EE_k^{-1} + k^2\nu_{LS}\nu_{LS}]^{-1}E_k^{-1}A\nu_{LS} \leq 1$	0.9691

4 The Monte Carlo simulation

This section presents a simulation study comparing the performance of different estimators considered in this study. The goal is to provide a comprehensive analysis under varying conditions, including sample size, multicollinearity level, noise level and the number of predictors. In addition, the evaluation criteria for the different estimation methods are also examined.

4.1 The design of the experiment

The regressors are generated following the approaches of [20], [12], [4] and [1], where:

$$x_{ij} = (1-\gamma^2)^{1/2}t_{ij} + \gamma t_{ip}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, q. \tag{32}$$

Here, t_{ij} are independent standard normal pseudo-random numbers, and γ is chosen to control the correlation between two regressors, where the correlation is given by γ^2. The variables are standardized so that $X'X$ is in correlation form. We consider five different correlation levels: $\gamma^2 = 0.80, 0.90, 0.95, 0.99, 0.999$. The number of regressors is set to $p = 4$ and $p = 8$ when $n = 30, 50, 100, 200$.

Newhouse and Oman [23] noted that the MSE is minimized subject to the constraint $\nu'\nu = 1$. We select the coefficients $\nu_1, \ldots, \nu_p$ as the normalized eigenvector corresponding to the largest eigenvalue of $X'X$. The response variable is generated by:

$$y_i = \nu_0 + \nu_1 x_{i1} + \cdots + \nu_q x_{iq} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{33}$$

where ε_i are independent normal pseudo-random numbers with mean zero and variance σ^2. The intercept ν_0 is set to zero for simplicity. The values of σ^2 considered are 5 and 10. For each choice of γ, n, and p, we generate the data and perform simulations 2,000 times with new error terms in each iteration. After sampling, we compute the Estimated MSE for each of the estimators as follows:

$$EMSE(\hat{\nu}) = \frac{1}{2000}\sum_{i=1}^{2000}(\hat{\nu}_{(i)} - \nu)'(\hat{\nu}_{(i)} - \nu) \tag{34}$$

where $\hat{\nu}_i$ represents each estimator under consideration for 2000 replication. The results are summarized in Tables 2 and 3.

4.2 Simulation results

The results in Tables 2 and 3 present the Mean Squared Errors (MSE) obtained by different estimation methods (LS, Ridge, Liu, KL, COLS, CRidge) for different values of $p = 4$ and $p = 8$, $n = 30, 50, 100, 200$, and $\sigma = 5$ and $\sigma = 10$. These results allow us to assess the performance of each method according to the complexity of the problem (increasing p) and the sample size.

Generally, for a given value of p and σ, we observe that increasing n leads to a significant reduction in the Mean Squared Error for all methods. This is particularly notable for traditional methods such as LS, which are especially sensitive to small sample sizes. For $n = 30$, the MSEs are significantly higher than for $n = 200$. However, this improvement is also observed for regularized methods such as Ridge, CRidge, and Liu, which confirms their increased robustness to variations in sample size.

When p increases from 4 to 8, a significant degradation in the performance of the methods is observed, particularly for LS which displays extremely high MSEs when p is also high. This is due to the fact that the non-regularized LS model becomes increasingly unstable as p increases, especially when n is relatively small. The regularized methods Ridge, CRidge, and Liu show better resistance to this increase in complexity, although their performance also decreases with increasing p.

Comparison of Methods

i. LS exhibits the worst performance in terms of MSE in almost all scenarios, especially when σ is high or when γ values are close to 1 (strong correlations).
ii. Ridge and CRidge show a clear improvement over OLS, with CRidge consistently outperforming Ridge. This is particularly evident for large samples where CRidge presents the lowest MSE values.
iii. Liu is competitive compared to Ridge but generally performs less than CRidge, although its results remain satisfactory compared to LS.
iv. KL and COLS exhibit interesting behaviours. KL shows very low MSEs for weak correlations $\gamma = 0.8$ but explodes for high correlations. COLS appears effective for small γ values but becomes unstable for stronger correlations.

Increasing the standard deviation σ naturally worsens the MSEs for all methods. However, regularized methods seem less affected by this increase than traditional methods, which confirms their robustness.

Overall, regularized methods, particularly CRidge, show good performance even when p increases or when correlations are high. LS is unsuitable for complex scenarios or small sample sizes. Therefore, the choice of method strongly depends on the sample size, model complexity (number of predictors), and the correlation structure between variables.

The Figures 1 and 2 present the evolution of the Mean Squared Error (MSE) as a function of the sample size for different estimators under a multicollinearity level of $\gamma = 0.8$. Specifically, Figure 1 corresponds to a standard deviation of $\sigma = 5$, while Figure 2 corresponds to $\sigma = 10$.

The results confirm that increasing the sample size consistently reduces the MSE for all estimators, indicating improved performance with more data. This trend is particularly noticeable for traditional estimators such as LS, which are highly sensitive to small sample sizes and exhibit significantly higher MSEs for smaller samples. However, this improvement with increasing sample size is also evident for regularized methods such as Ridge, CRidge, and Liu, demonstrating their enhanced robustness to variations in sample size.

Furthermore, the Ridge and CRidge estimators effectively reduce prediction error under severe multicollinearity, as expected by their design. The CRidge estimator consistently outperforms the other methods, presenting the lowest MSE values across different sample sizes. This superior performance of CRidge is particularly evident for large samples, highlighting its effectiveness in terms of numerical stability and predictive accuracy.

The Figures illustrate that the improvement achieved by the CRidge estimator is most pronounced when the sample size increases. This confirms that CRidge is a promising alternative to classical and other biased estimation methods when dealing with multicollinearity in regression analysis.

These findings support the theoretical conclusions drawn in previous sections and justify the proposed estimator's applicability under various sample sizes and multicollinearity levels.

Table 2: Simulated result in terms of MSE when $p = 4$

		$\sigma = 5$					$\sigma = 10$				
n	Estimator	0.8	0.9	0.95	0.99	0.999	0.8	0.9	0.95	0.99	0.999
30	LS	8.937	15.714	29.356	138.368	1361.430	35.748	62.854	117.425	553.472	5445.720
	Ridge	4.764	6.968	10.387	24.226	55.437	18.726	27.528	41.216	96.675	221.680
	Liu	7.014	10.511	15.274	27.435	85.119	28.024	41.998	61.049	109.602	340.488
	KL	2.246	2.540	2.943	15.591	586.102	8.633	9.920	11.693	62.674	2344.873
	COLS	1.347	2.425	5.516	49.110	898.741	3.510	7.955	20.486	195.295	3594.343
	CRidge	1.027	1.428	2.369	9.409	38.418	2.217	3.992	7.967	36.709	153.388
50	LS	4.035	7.499	14.562	71.664	717.549	16.141	29.995	58.247	286.655	2870.195
	Ridge	2.785	4.432	7.082	18.571	46.554	10.859	17.386	27.955	73.982	186.114
	Liu	3.700	6.365	10.711	24.510	54.930	14.788	25.430	42.813	97.724	218.901
	KL	1.805	2.275	2.576	2.687	200.749	6.799	8.684	9.991	10.901	803.516
	COLS	0.854	1.135	2.096	18.756	409.877	1.363	2.615	6.625	73.716	1638.774
	CRidge	0.811	0.945	1.355	5.553	28.594	1.173	1.857	3.691	21.058	113.978
100	LS	1.904	3.292	6.090	28.421	278.670	7.614	13.168	24.361	113.684	1114.679
	Ridge	1.512	2.368	3.835	11.266	35.852	5.860	9.212	15.008	44.693	143.230
	Liu	1.832	3.074	5.362	16.917	37.345	7.324	12.284	21.423	67.596	149.037
	KL	1.173	1.613	2.151	2.550	32.274	4.371	6.054	8.175	10.060	129.560
	COLS	0.785	0.826	1.009	4.974	122.369	0.942	1.194	2.050	18.315	488.522
	CRidge	0.780	0.796	0.888	2.409	17.326	0.911	1.071	1.572	8.131	68.636
200	LS	1.205	2.307	4.528	22.337	222.707	4.820	9.226	18.112	89.347	890.827
	Ridge	1.010	1.753	3.022	9.560	32.198	3.899	6.788	11.782	37.863	128.596
	Liu	1.175	2.195	4.107	14.517	33.275	4.699	8.774	16.415	58.076	132.794
	KL	0.833	1.281	1.845	2.531	19.869	3.088	4.766	6.952	9.918	79.909
	COLS	0.786	0.797	0.894	3.553	90.448	0.881	1.007	1.518	12.553	360.776
	CRidge	0.785	0.783	0.828	1.910	14.415	0.870	0.950	1.261	6.047	56.912

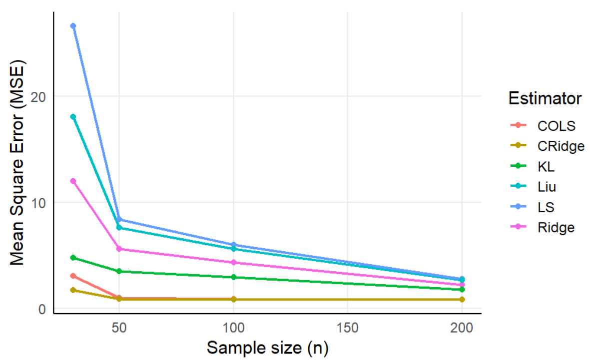

Figure 1: MSE against sample size for $\gamma = 0.8$ and $\sigma = 5$ using the data from Table 2

Table 3: Simulated result in terms of MSE when $p = 8$

		$\sigma = 5$					$\sigma = 10$				
n	Estimator	0.8	0.9	0.95	0.99	0.999	0.8	0.9	0.95	0.99	0.999
30	LS	26.620	48.353	91.343	430.023	4196.574	106.482	193.413	365.373	1720.093	16786.295
	Ridge	12.015	17.985	26.908	62.266	144.953	47.693	71.579	107.298	248.842	579.741
	Liu	18.026	25.935	34.779	46.066	69.133	72.107	103.724	139.182	184.575	276.727
	KL	4.774	6.120	9.387	77.441	2097.549	18.821	24.329	37.549	310.097	8390.647
	COLS	3.084	7.156	9.387	168.786	2920.924	10.560	27.024	72.322	674.172	11683.202
	CRidge	1.734	3.046	9.387	25.450	103.361	5.195	10.660	22.297	101.101	413.262
50	LS	8.370	14.369	26.406	122.379	1198.138	33.481	57.475	105.622	489.518	4792.553
	Ridge	5.619	8.553	13.404	35.806	96.928	22.150	33.840	53.214	142.866	387.565
	Liu	7.595	12.184	19.744	43.700	40.117	30.374	48.730	78.979	174.749	160.684
	KL	3.492	4.427	5.261	5.026	273.475	13.488	17.231	20.630	20.089	1094.406
	COLS	1.003	1.441	2.840	26.945	631.559	1.915	3.808	9.583	106.473	2525.513
	CRidge	0.907	1.129	1.761	8.531	53.388	1.530	2.573	5.301	32.964	213.148
100	LS	6.017	11.757	23.281	115.464	1151.590	24.067	47.026	93.123	461.854	4606.359
	Ridge	4.328	7.411	12.409	35.092	95.585	17.017	29.283	49.233	140.002	382.194
	Liu	5.621	10.306	18.098	43.275	37.027	22.478	41.213	72.384	173.385	148.300
	KL	2.944	4.157	5.221	4.139	243.422	11.317	16.136	20.440	16.522	974.205
	COLS	0.898	1.224	2.405	24.528	598.754	1.404	2.872	7.793	96.781	2394.281
	CRidge	0.852	1.025	1.588	8.099	52.067	1.223	2.089	4.559	31.205	207.855
200	LS	2.756	5.291	10.387	51.191	510.131	11.025	21.165	41.548	204.763	2040.525
	Ridge	2.239	3.912	6.787	21.689	72.665	8.753	15.367	26.792	86.342	290.432
	Liu	2.675	4.994	9.286	31.305	45.096	10.698	19.975	37.141	125.206	180.295
	KL	1.779	2.756	4.003	5.389	43.358	6.765	10.570	15.508	21.251	173.829
	COLS	0.824	0.871	1.120	7.129	204.677	0.959	1.267	2.423	26.934	817.783
	CRidge	0.817	0.837	0.971	3.416	30.732	0.930	1.134	1.839	12.159	122.283

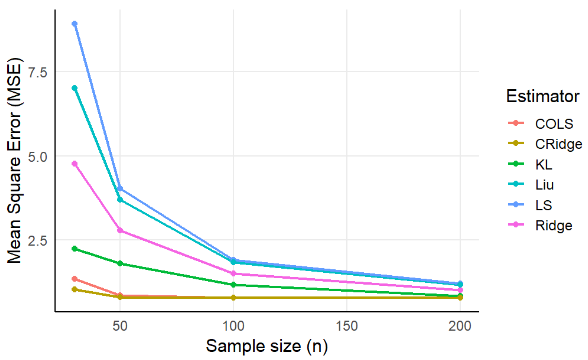

Figure 2: MSE against sample size for $\gamma = 0.8$ and $\sigma = 10$ using the data from Table 2

5 Illustrative applications

The Electricity dataset contains cost function data for 145 US electricity producers in 1955, with an additional 14 observations representing aggregate statistics (Greene, 2003). For statistical analysis, only the first 145 observations should be used. The data set comprises eight variables:

- **cost**: Total production cost.
- **output**: Total output of electricity.
- **labor**: Wage rate of labor.
- **laborshare**: Cost share for labor.
- **capital**: Capital price index.
- **capitalshare**: Cost share for capital.
- **fuel**: Fuel price.
- **fuelshare**: Cost share for fuel.

The total production cost serves as the response variable, modeled as a function of seven standardized predictors. To evaluate the adequacy of the linear regression model, we employed the Ramsey RESET test, which yielded a p-value of 1, indicating no evidence of model misspecification. The model was estimated using the least squares (LS) method, and diagnostic checks were performed. The variance inflation factor (VIF), presented in Figure 1, along with a condition number of 23.37, suggests multicollinearity.

To evaluate the predictive performance of the estimator, we used scalar mean squared error (SMSE), prediction mean squared error (PMSE) and prediction mean absolute error (PMAE). The data set was randomly partitioned into 80% training set and a 20% test set. The model parameters were estimated on the training data and the performance metrics were calculated using the test data to evaluate the generalizability.

The application result in Table 4 rigorously evaluates the performance of the estimators considered in this study, with a particular focus on the stability of the estimation and the predictive precision. The analysis compares Least Squares (LS), Ridge Regression (RIDGE), Liu Estimator (LIU), Kibria-Lukman Estimator (KL), Contraction Least Squares (COLS), and the newly proposed contraction ridge estimator (CRidge). Performance is assessed using the Scalar Mean Squared Error (SMSE), Prediction Mean Squared Error (PMSE) and Prediction Mean Absolute Error (PMAE).

Although LS remains an unbiased estimator, it suffers from high variance in the presence of multicollinearity, making it unsuitable for reliable parameter estimation. The CRidge estimator demonstrates superior performance across multiple criteria, striking an optimal balance between bias and variance. It achieves an SMSE of 11.0217, significantly lower than OLS (52.6319) and comparable to other shrinkage-based estimators. Notably, the KL estimator attains the lowest SMSE (2.1535), reflecting its strong regularization capabilities.

Predictive accuracy, assessed through PMSE, highlights the robustness of CRidge, which attains one of the lowest values (15.7785), outperforming LS (134.8210), RIDGE (130.8600), and Liu (129.0342). The COLS estimator also demonstrates strong predictive performance (PMSE = 16.1763), reinforcing the effectiveness of contraction-based methods. Although KL achieves a competitive PMSE (126.9678), it does not surpass CRidge, further supporting the advantages of the proposed estimator.

PMAE, which provides an alternative measure of predictive consistency by capturing absolute deviations, confirms CRidge's superior predictive stability (PMAE = 2.6232), closely followed by COLS (PMAE = 2.6412). Traditional estimators, including LS (6.7044), RIDGE (6.6107), and Liu (6.5648), exhibit significantly higher PMAE values, reinforcing their limitations in handling multicollinearity. Although KL performs reasonably well (6.5170), it remains outperformed by CRidge.

Overall, the proposed CRidge estimator emerges as the most effective approach, offering substantial improvements in estimation stability and predictive accuracy. While COLS exhibits strong predictive performance, its higher SMSE suggests the need for fine-tuning in some applications. The KL, Ridge, and Liu estimators remain competitive alternatives, but their slightly higher PMSE and PMAE values indicate that CRidge is a more robust and globally applicable choice.

6 Concluding Remarks

Linear regression remains a fundamental tool for modeling continuous response variables, yet multicollinearity poses significant challenges to the accuracy and reliability of least squares (LS) estimates.

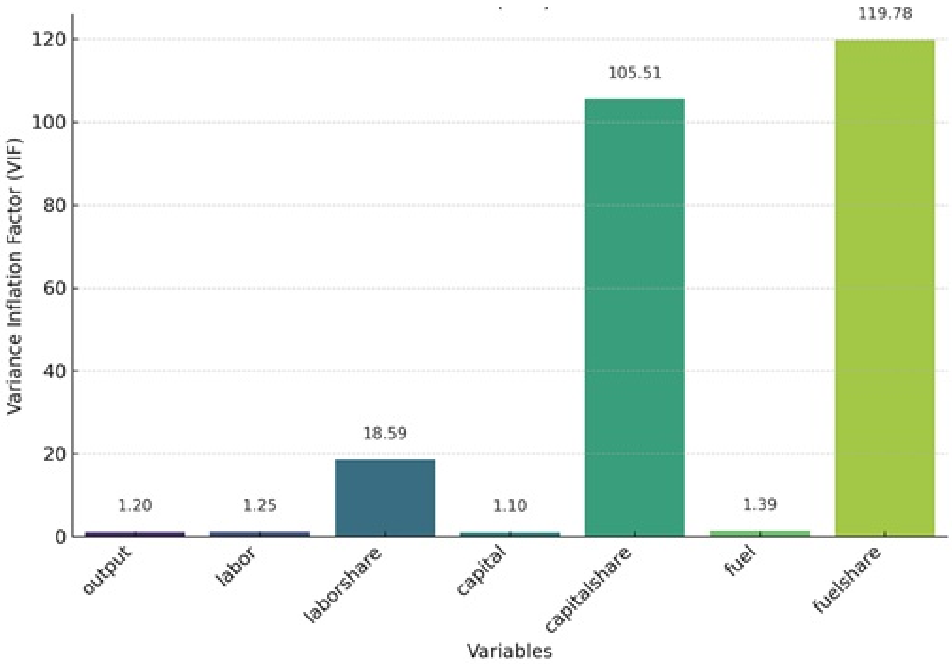

Figure 3: Variance Inflation Factor (VIF) for Predictor Variables

Table 4: Regression Estimates and Model Validation Using PMSE and PMAE

Coef.	LS	Ridge	LIU	KL	COLS	CRidge
Intercept	13.9470	13.8682	13.8280	13.7894	8.4066	8.3591
X1	20.7966	20.6483	20.5738	20.4999	12.5352	12.4458
X2	1.1389	1.1821	1.2028	1.2254	0.6864	0.7125
X3	1.5159	1.3405	1.2931	1.1652	0.9137	0.8080
X4	0.8668	0.8781	0.8815	0.8895	0.5224	0.5293
X5	0.4957	0.1355	0.0658	-0.2247	0.2988	0.0817
X6	1.6994	1.6618	1.6438	1.6242	1.0243	1.0017
X7	0.3326	-0.0276	-0.0927	-0.3877	0.2005	-0.0161
SMSE	52.6319	13.0051	10.9094	2.1535	106.1460	11.0217
PMSE	134.8210	130.8600	129.0342	126.9678	16.1763	15.7785
PMAE	6.7044	6.6107	6.5648	6.5170	2.6412	2.6232

While regularization techniques such as ridge regression and the Liu estimator have been extensively studied, contraction least squares have received comparatively less attention. We introduce the contraction ridge estimator (CRidge) as a novel alternative that effectively addresses multicollinearity while preserving predictive accuracy. We rigorously derived the statistical properties of CRidge and established theoretical conditions under which it outperforms existing regularization methods. These conditions were validated empirically using the electricity dataset, confirming their practical relevance. Additionally, extensive simulations across varying sample sizes, predictor counts, multicollinearity levels, and error variances demonstrated the superiority of CRidge, as evidenced by its consistently lower mean squared error (MSE). A real-world application using electricity consumption data further reinforced these findings, showing that CRidge yields the lowest predicted mean squared error (PMSE) and predicted mean absolute error (PMAE) among competing shrinkage estimators. Diagnostic tests confirmed the appropriateness of the model, highlighting its robustness in real-world settings. The proposed estimator offers a significant advancement in regression modeling, particularly in fields where multicollinearity is prevalent, such as economics, environmental sciences, finance, and biomedical research. Given its strong theoretical foundation and empirical performance, CRidge presents a compelling alternative to conventional shrinkage methods. Future research will explore a robust contraction Ridge estimator, extending the methodology to accommodate both multicollinearity and the presence of outliers, ensuring broader applicability across diverse disciplines and datasets.

Author Contributions

Conceptualization: E. A. and A. F. L.; Methodology: E. A. and A. F. L.; Software: E. A. and A. F. L.; Formal Analysis: E. A. and A. F. L.; Resources: E. A. and A. F. L.; Writing—Original Draft Preparation: E. A. and A. F. L.; Review and Editing: E. A. and A. F. L.; Visualization: E. A. and A. F. L.

Conflict of interest

The authors have no conflicts of interest to declare.

Funding

This research received no external funding.

Data Availability Statement

The data will be made available upon request from the corresponding author

References

[1] Alreshidi, N. A., Alrasheedi, M. A., Lukman, A. F., Alrweili, H., & Farghali, R. A. Modified Kibria–Lukman Estimator for the Conway–Maxwell–Poisson Regression Model: Simulation and Application. *Mathematics*, 13(5), 794, 2025.

[2] Algamal, Z. Y., Lukman, A. F., Abonazel, M. R., & Awwad, F. A. Performance of the Ridge and Liu Estimators in the zero-inflated Bell Regression Model. *Journal of Mathematics*, 2022(1), 9503460, 2022.

[3] Amin, M., Akram, M. N., & Majid, A. On the estimation of Bell regression model using ridge estimator. *Communications in Statistics - Simulation and Computation*, 52(3), 854-867,2023.

[4] Bulut, Y. M., Lukman, A. F., Işılar, M., Adewuyi, E. T., & Algamal, Z. Y. Modified ridge estimator in the Bell regression model. *Journal of Inverse and Ill-posed Problems*, 32(6), 1081-1091, 2024.

[5] Bulut, Y. M., Lukman, A. F., Işılar, M., Adewuyi, E. T., & Algamal, Z. Y. Modified ridge estimator in the Bell regression model. *Journal of Inverse and Ill-posed Problems*, 2024. Available from: https://doi.org/10.1515/jiip-2022-0069

[6] Cessie, S. L., and Houwelingen, J. V. "Ridge estimators in logistic regression." *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 41, no. 1, pp. 191–201, 1992.

[7] Greene, W.H. Econometric Analysis, 5th edition. Upper Saddle River, NJ: Prentice Hall, 2003.

[8] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67, 1970.

[9] Hoerl, E., Kennard, R.W. and Baldwin, K.F. Ridge Regression: Some Simulations. *Communications in Statistics*, 4, 105-123, 1975.

[10] Farebrother, R. W. Further results on the mean square error of ridge regression. *J. Roy. Statist. Soc.* B 38:248–250, 1976.

[11] Kibria, B. G., Månsson, K., & Shukur, G. Some ridge regression estimators for the zero-inflated Poisson model. *Journal of Applied Statistics*, 40(4), 721-735, 2013.

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

[13] Kibria, B. G., & Lukman, A. F. Some ridge regression estimators for the zero-inflated Poisson model. *Journal of Applied Statistics*, 40(4), 721-735, 2020.

[14] Le Cessie, S., & Van Houwelingen, J. C. Ridge estimators in logistic regression. *Applied Statistics*, 41(1), 191-201, 1992.

[15] Liu, K. A New Class of Biased Estimators in Linear Regression. *Communications in Statistics - Theory and Methods*, 22(2), 393-402, 1993.

[16] Lukman, A. F., Adewuyi, E., Månsson, K., & Kibria, B. M. G. A new estimator for the multicollinear Poisson regression model: simulation and application. *Scientific Reports*, 11(1), 3732, 2021.

[17] Lukman, A. F., Aladeitan, B., Ayinde, K., & Abonazel, M. R. Modified ridge-type for the Poisson regression model: simulation and application. *Journal of Applied Statistics*, 49(8), 2124–2136, 2022.

[18] Månsson, K., & Shukur, G. A Poisson ridge regression estimator. *Economic Modelling*, 28(4), 1475-1481, 2011.

[19] Månsson, K. On ridge estimators for the negative binomial regression model. *Economic Modelling*, 29(2), 178-184, 2012.

[20] McDonald, G.C. & Galarneau, D.I. A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association*, 70, 407-416, 1975.

[21] Montgomery, D. C., Peck, E. A., & Vining, G. G. Introduction to Linear Regression Analysis. *Wiley*, 2012.

[22] Mayer, L. S., Willke, T. A. On biased estimation in linear models. *Technometrics* 15:497–508, 1973.

[23] Newhouse, J. P., Oman, S. D. An evaluation of ridge estimators. *Rand Corporation*, R-716-PR, 1–28, 1971.

[24] Özkale, M. R., & Kaciranlar, S. The restricted and unrestricted two-parameter estimators. *Communications in Statistics—Theory and Methods*, 36(15), 2707-2725, 2007.

[25] Trenkler, G., Toutenburg, H. (1990). Mean squared error matrix comparisons between biased estimators: an overview of recent results. *Statst. Pap.* 31:165–179.