Mathematics and Statistics 9(5): 630-638, 2021 http://www.hrpub.org
DOI: 10.13189/ms.2021.090502

A Monte Carlo Study for Dealing with Multicollinearity and Autocorrelation Problems in Linear Regression Using Two Stage Ridge Regression Method

Hussein Eledum[1,*], Hytham Hussein Awadallah[2]

[1]Department of Statistics, Faculty of Science, University of Tabuk, Saudi Arabia
[2]Department of Applied Statistics, Faculty of Science and Technology, Shendi University, Sudan

Received May 28, 2021; Revised July 20, 2021; Accepted August 22, 2021

Cite This Paper in the following Citation Styles

(a): *[1] Hussein Eledum, Hytham Hussein Awadallah , "A Monte Carlo Study for Dealing with Multicollinearity and Autocorrelation Problems in Linear Regression Using Two Stage Ridge Regression Method," Mathematics and Statistics, Vol. 9, No. 5, pp. 630 - 638, 2021. DOI: 10.13189/ms.2021.090502.*

(b): *Hussein Eledum, Hytham Hussein Awadallah (2021). A Monte Carlo Study for Dealing with Multicollinearity and Autocorrelation Problems in Linear Regression Using Two Stage Ridge Regression Method. Mathematics and Statistics, 9(5), 630 - 638. DOI: 10.13189/ms.2021.090502.*

Abstract In the multiple linear regression model, the problem of multicollinearity may come together with autocorrelation; therefore, several methods of estimation are developed to deal with this case; Two-Stage Ridge Regression (TR) is one of them. This article's main objective is to run a Monte Carlo simulation to investigate the impact of both problems, Multicollinearity and Autocorrelation, in multiple linear regression model on the performance of the TR. The simulation is carried out under different levels of multicollinearity, and sets of autocorrelation coefficient, taking into account different sample sizes. Some new properties for the TR method, including expectation, variance and mean square error, are droved. In contrast, the study also has developed some techniques to estimate the biasing parameter for the TR by modifying some popular techniques used in ridge regression (RR). Moreover, Mean Square Error is used as a base for evaluation and comparison. The empirical findings from the simulations revealed that the TR estimator performs better than the RR, and the values of the biasing parameter under the TR are always less than that under the RR. This paper contributes to the existing literature on developing new estimation methods used to overcome the presence of mixed problems in a linear regression model and studying their properties.

Keywords Autocorrelation, Multicollinearity, Ridge Regression, Two Stages, Monte Carlo Simulation

1. Introduction

Consider the multiple linear regression model in matrix form

$$y = X\beta + \varepsilon \tag{1}$$

where y is an $n \times 1$ vector of observations on the dependent variable, X is a full rank $n \times p$ non-stochastic matrix of observations on the explanatory variables, β is the $p \times 1$ vector of unknown coefficients, and ε is an $n \times 1$ random vector of disturbances with zero means.

For model (1) the Least Squares (LS) estimator of β is,

$$\hat{\beta}_{LS} = (X'X)^{-1}X'y \tag{2}$$

Both LS estimator and its covariance matrix heavily depend on the characteristics of the $X'X$ matrix. If $X'X$ is ill-conditioned, i.e., the column vectors of X are linearly dependent, the LS estimators are sensitive to a number of errors. For example, some of the regression coefficients may be statistically insignificant or have the wrong sign, which may result in wide confidence intervals for individual parameters. One of the most popular estimators dealing with multicollinearity is the Ridge Regression (RR) estimator proposed by Hoerl and Kennard [6,7] which is defined as,

$$\hat{\beta}_{RR} = (X'X + kI_P)^{-1}X'y = \left(I_p + k(X'X)^{-1}\right)^{-1}\hat{\beta}_{LS} \quad (3)$$

where k is a non-negative constant. It is called biasing or ridge parameter. It is observed that when $k = 0$, model (3) returns LS estimates.

For the sake of convenience, we assume that the matrix X and response variable y are standardized in such a way that $X'X$ is a non-singular correlation matrix, and $X'y$ is the correlation between X and y. Let Λ and V be the matrices of eigenvalues and eigenvectors of $X'X$, respectively, satisfy $V'(X'X)V = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, where λ_i is the i^{th} eigenvalue of $X'X$, and $V'V = I_P$ we obtain the canonical form of model (1) as

$$y = Za + \epsilon \quad (4)$$

where $Z = XV$ implies that $Z'Z = \Lambda$, and $a = V'\beta$. The LS estimator of a is given as [10,11]

$$\hat{a}_{LS} = \Lambda^{-1}Z'y$$

Therefore, LS estimator of β is given by

$$\hat{\beta}_{LS} = V\hat{a}_{LS}$$

and the mean square error of $\hat{a}_{LS}$ is given as

$$\text{MSE}(\hat{a}_{LS}) = \hat{\sigma}^2 \sum_{i=1}^{p} \lambda_i^{-1} \quad (5)$$

where $\hat{\sigma}^2$ is the estimates of σ^2 in the LS.

The RR estimator of a is given as

$$\hat{a}_{RR} = \left(I_p + k\Lambda^{-1}\right)^{-1}\hat{a}_{LS}$$

and the mean square error of $\hat{a}_{RR}$ is given as

$$\text{MSE}(\hat{a}_{RR}) = Kk^2 \sum_{i=1}^{p} \frac{\hat{a}_{LS_i}^{2}}{(\lambda_i + \hat{K})^2} + \hat{\sigma}^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + \hat{K})^2} \quad (6)$$

Several techniques had been developed to estimate k, the most popular of which is developed by [6] (HK) that is

$$\hat{k}_{HK} = \frac{\hat{\sigma}^2}{max\left(\hat{a}_{LS_i}^2\right)} \quad (7)$$

Hoerl et al. [5] (HKB) proposed the following estimator

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \hat{a}_{LS_i}^2} \quad (8)$$

Thisted [13] (TH) suggested that the estimated value of k be

$$\hat{k}_T = \frac{(p-2)\hat{\sigma}^2}{\sum_{i=1}^{p} \hat{a}_{LS_i}^2} \quad (9)$$

Lawless and Wang [9] (LW) proposed estimator of k as

$$\hat{k}_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \lambda_i \hat{a}_{LS_i}^2} \quad (10)$$

In model (1) we assume the error term follows the first-order autoregressive scheme AR (1) given by

$$\epsilon_t = \rho\epsilon_{t-1} + v_t \qquad t = 1,2,\dots,n \quad (11)$$

where $|\rho| < 1$ and v_t is a random variable with zero mean, and

$$E(v_t v_{t-s}) = \begin{cases} \sigma^2, & s = 0 \\ 0, & s \neq 0 \end{cases} \text{ and } \text{Cov}(\epsilon) = \sigma^2\Sigma$$

Since $\sigma^2\Sigma$ is the variance-covariance matrix of the errors, Σ must be positive definite, so there exists an $n \times n$ symmetric matrix P, such that $P'P = \Sigma$ so that model (1) can be written as

$$Py = PX\beta + P\varepsilon$$

Let $y^* = Py$, $X^* = PX$ and $\varepsilon^* = P\varepsilon$ then $E(\varepsilon^*) = 0$ and $\text{Cov}(\epsilon^*) = \sigma^2 I$

Therefore, the transformation model

$$y^* = X^*\beta + \varepsilon^* \quad (12)$$

satisfies the assumption of error $\varepsilon^* \sim N(0, \sigma^2 I)$. For model (12) the LS estimator, which called the Two-stage Prais-Winsten (TP) [12] is

$$\hat{\beta}_{TP} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^* = (X'\Sigma X)^{-1}X'\Sigma y \quad (13)$$

where $\Sigma = P'P$ and

$$P_{nxn} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ 0 & 0 & -\rho & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

$\hat{\beta}_{TP}$ is the best linear unbiased estimator of β, where $\text{Cov}(\beta) = \sigma^2(X'\Sigma X)^{-1}$.

Since the rank of X^* is equal to that of X, then the multicollinearity still also affects the TP estimator. Eledum and Zahri [2] proposed the Two-stage Ridge estimator (TR) as:

$$\hat{\beta}_{TR} = (X'\Sigma X + kI_P)^{-1}X'\Sigma y \quad (14)$$

We assume that the transformation matrix X^* and response variable y^* are standardized in such a way that $X^{*\prime}X^* = X'\Sigma X$ is a non-singular correlation matrix and $X^{*\prime}y^* = X'\Sigma Y$ is the correlation between X^* and y^*. Let the matrix $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_p)$ be the diagonal matrix whose diagonal elements are the eigenvalues $\theta_1 > \theta_2 > \cdots > \theta_p > 0$ of $X'\Sigma X$ and G be the matrix consisting of eigenvectors of $X'\Sigma X$ satisfying $G'(X'\Sigma X)G = \Theta$ and $G'G = I_P$. The canonical form of model (12) is

$$y^* = Wa^* + \epsilon^* \quad (15)$$

where $W = X^*G$ implies that $W'W = \Theta$, and $a^* = G'\beta$. The TP estimator of a^* is given as

$$\hat{a}^*_{TP} = \Theta^{-1}W'y^* \quad (16)$$

Therefore, TP estimator of β is given by

$$\hat{\beta}_{TP} = G\hat{a}_{TP}$$

and the mean square error of $\hat{a}_{PW}$ is given as

$$\text{MSE}(\hat{a}_{TP}) = \hat{\sigma}^2_{TP} \sum_{i=1}^{p} \theta_i^{-1} \quad (17)$$

where $\hat{\sigma}^2_{TP}$ is the estimates of σ^2 in the TP method.

The TR estimator of a^* is given as

$$\hat{a}^*_{TR} = \left(\Theta + kI_p\right)^{-1} W'y^* \quad (18)$$

In this article, some properties were droved for TR, and some techniques for estimating the ridge regression parameter k in case of TR are considered. Mean Square Error is used as a base for evaluation and comparison.

The rest of the paper is outlined as follows. In Section 2, we derive some new properties for the Two-stage Ridge estimator $\hat{a}^*_{TR}$. Estimation of the TR parameter k is demonstrated in Section 3. Numerical example and Monte Carlo Simulation are given in Sections 4 and 5, respectively. Section 6 pertains to the results of the simulation. Finally, we provide some concluding remarks in Section 7.

2. Some New Properties for $\hat{a}^*_{TR}$

In this section, we present some interesting properties for the TR estimator $\hat{a}^*_{TR}$

Lemma 1. The Expected value of $\hat{a}^*_{TR}$ is given as

$$E(\hat{a}^*_{TR}) = a^* - kI_p\left(\Theta + kI_p\right)^{-1} a^*$$

Proof. Consider the TR estimator $\hat{a}^*_{TR}$ of (18) namely

$$\hat{a}^*_{TR} = \left(\Theta + KI_p\right)^{-1} W'y^*$$

substitute y^* in the above equation, by its value in (15) we get

$$\hat{a}^*_{TR} = \left(\Theta + kI_p\right)^{-1} W'(Wa^* + \epsilon^*)$$
$$= \left(\Theta + kI_p\right)^{-1} \Theta a^* + \left(\Theta + kI_p\right)^{-1} W'\epsilon^*$$

taking the expectation for the above equation and since $E(\epsilon^*) = 0$, we get

$$E(\hat{a}^*_{TR}) = \left(\Theta + kI_p\right)^{-1} \Theta a^*$$

adding and subtracting kI_p from the matrix Θ in the above equation, to get

$$E(\hat{a}^*_{TR}) = \left(\Theta + kI_p\right)^{-1}\left[\left(\Theta + kI_p\right) - kI_p\right] a^*$$

then the expectation of $\hat{a}^*_{TR}$ is given as:

$$E(\hat{a}^*_{TR}) = a^* - kI_p\left(\Theta + kI_p\right)^{-1} a^*$$

From equation above, we note that the $\hat{a}^*_{TR}$ is a biased estimator, where the term $-kI_p\left(\Theta + KI_p\right)^{-1} a^*$ represents the bias.

Lemma 2. The variance of $\hat{a}^*_{TR}$ is given as

$$Var(\hat{a}^*_{TR}) = \hat{\sigma}^2_{TP}\left(\Theta + kI_p\right)^{-2} WW'$$

Proof. We directly compute the variance of $\hat{a}^*_{TR}$ by taking variance for (18)

$$Var(\hat{a}^*_{TR}) = Var\left[\left(\Theta + kI_p\right)^{-1} W'y^*\right]$$
$$= \left(\Theta + KkI_p\right)^{-1} WW'\left(\Theta + kI_p\right)^{-1} Var(y^*)$$
$$= \hat{\sigma}^2_{TP}\left(\Theta + kI_p\right)^{-2} WW'$$

Lemma 3. The mean square error of $\hat{a}^*_{TR}$ is given as

$$MSE(\hat{a}^*_{TR}) = \left(\Theta + kI_p\right)^{-2} k^2 I_p a^{*\prime} a^*$$
$$+ \hat{\sigma}^2_{TP}\left(\Theta + kI_p\right)^{-2} WW'$$

Proof. We know that

$$MSE(\hat{\theta}) = E\left(\hat{\theta} - \theta\right)'\left(\hat{\theta} - \theta\right)$$

where $\hat{\theta}$ is an estimator of θ.

Therefore, $\hat{a}^*_{TR} - a^* = \left(\Theta + kI_p\right)^{-1} W'y^* - a^*$. Substitute y^* by its value in the (15), we obtain:

$$\hat{a}^*_{TR} - a^* = \left(\Theta + kI_p\right)^{-1} W'(Wa^* + \epsilon^*) - a^*$$
$$= \left(\Theta + kI_p\right)^{-1} \Theta a^* + \left(\Theta + kI_p\right)^{-1} W'\epsilon^* - a^*$$
$$= \left[\left(\Theta + kI_p\right)^{-1} \Theta - I_P\right] a^* + \left(\Theta + kI_p\right)^{-1} W'\epsilon^*$$

adding and subtracting kI_p from the first term of the right side of the above equation, we get

$$\hat{a}^*_{TR} - a^* = \left[\left(\Theta + kI_p\right)^{-1} \Theta + kI_p - kI_p - I_P\right] a^*$$
$$+ \left(\Theta + kI_p\right)^{-1} W'\epsilon^*$$
$$= \left[\left(\Theta + kI_p\right)^{-1}\left(\Theta + kI_p\right) - kI_p\left(\Theta + kI_p\right)^{-1} - I_P\right] a^*$$
$$+ \left(\Theta + kI_p\right)^{-1} W'\epsilon^*$$
$$= \left[I_P - kI_p\left(\Theta + kI_p\right)^{-1} - I_P\right] a^* + \left(\Theta + kI_p\right)^{-1} W'\epsilon^*$$
$$= -kI_p\left(\Theta + kI_p\right)^{-1} a^* + \left(\Theta + kI_p\right)^{-1} W'\epsilon^*$$
$$= \mathcal{M}$$

therefore, $\text{MSE}(\hat{a}^*_{TR}) = E(\mathcal{M}'\mathcal{M})$, since $E(\epsilon^*\epsilon^{*\prime}) = \hat{\sigma}^2_{TP}$ and the cross-product is zero because $E(\epsilon^*) = 0$, then

$$MSE(\hat{a}^*_{TR}) = \left(\Theta + kI_p\right)^{-2} k^2 I_p a^{*\prime} a^*$$
$$+ \hat{\sigma}^2_{TP}\left(\Theta + kI_p\right)^{-2} WW'$$

where the first term represents $\text{Bias}^2(\hat{a}^*_{TR})$ and the second one refers to $\text{Var}(\hat{a}^*_{TR})$.

3. Estimating the Two Stage Ridge TR Parameter k

This section discusses some techniques used to estimate k for the TR of the model (14) by modifying the techniques of (7,8, 9 and 10). The modified Hoerl and Kenard (MHK) $\hat{k}_{MHK}$, modified Hoerl, Kenard and Baldwin (MHKB) $\hat{k}_{MHKB}$, modified Thisted (MTH)

$\hat{k}_{MTH}$, and modified Lawless and Wang (MLW) $\hat{k}_{MLW}$ are respectively given as:

$$\hat{k}_{MHK} = \frac{\hat{\sigma}^2_{TP}}{max(\hat{a}^2{}_{TP_i})} \quad (19)$$

$$\hat{k}_{MHKB} = \frac{p\hat{\sigma}^2_{TP}}{\sum_{i=1}^{p} \hat{a}^2{}_{TP_i}} \quad (20)$$

$$\hat{k}_{MTH} = \frac{(p-2)\hat{\sigma}^2_{TP}}{\sum_{i=1}^{p} \hat{a}^2{}_{TP_i}} \quad (21)$$

$$\hat{k}_{MLW} = \frac{p\hat{\sigma}^2_{TP}}{\sum_{i=1}^{p} \lambda_i \hat{a}^2{}_{TP_i}} \quad (22)$$

Where $\hat{\sigma}^2_{TP}$ is the estimates of σ^2 in the TP method and $\hat{a}_{TP}$ is the TP estimator of the parameter vector a.

4. Numerical Example

To show the properties of TR estimator we consider a real data set. The data set is the product in manufacturing sector in Iraq, this data set is used by [2] to evaluate the RR under the existence of AR(1). The data consists of three explanatory variables and 31 observations. The dependent variable represents the product value in the manufacturing sector and the explanatory variables refer to the value of the imported intermediate, imported capital commodities and the value of imported raw materials, respectively. The condition number of this data set is 600.25, which suggests the presence of multicollinearity. Durbin Watson test statistic is 0.9047 with p-value 0.0001996 indicates that there exist a positive AR(1) scheme. The estimate of the autocorrelation coefficient ρ is computed to be 0.5358. Table 1 shows values of k using the four techniques of determining k for each RR and TR, while table 2 demonstrates the MSE for LS, TP, RR, and TR.

Based on the outputs in table 2, we observe that MSE for TP (0.08117) is less than that of LS (0.11615). On the other hand, MSEs for TR that were computed using each of the four techniques of choosing k (0.07314, 0.07363, 0.07541, 0.07374) are less than MSEs for RR that were calculated using the corresponding techniques of estimating k (0.09482, 0.09091, 0.10298, 0.09164).

Table 1. Values of k using the four techniques of choosing k for RR and TR for the manufacturing data set

RR				
Technique	HK	HKB	TH	LW
Value	0.02586	0.03803	0.01268	0.03491
TR				
Technique	MHK	$MHKB$	MTH	MLW
Value	0.08604	0.02868	0.08806	0.08604

Table 2. MSE of LS, TP, RR and, TR for the manufacturing data set

Method	$K=0$	HK	HKB	TH	LW	MHK	$MHKB$	MTH	MLW
LS	0.11615	-	-	-	-	-	-	-	-
TP	0.08117	-	-	-	-	-	-	-	-
RR	0.11615	0.09482	0.09091	0.10298	0.09164	-	-	-	-
TR	0.08117	-	-	-	-	0.07314	0.07363	0.07541	0.07374

5. Monte Carlo Simulation

The simulation is carried out under different degrees of multicollinearity and different sets of autocorrelation coefficient taking into account different sample sizes. Computer programs using R have been designed. Based on [8] the explanatory variables were generated using

$$x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip+1},$$
$$i = 1,2,\dots,n \qquad j = 1,2,\dots,p$$

where z_{i1}, z_{i2}, $\dots, z_{ip+1}$ are independent standard normal pseudo-random numbers, and γ is specified so that the correlation between any two explanatory variables is given by γ^2.

Following [8], three different sets of correlations are considered corresponding to γ^2= 0.70, 0.80, and 0.99, which refer to a low, medium and high multicollinearity, respectively. Observations on the dependent variable are computed by

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$
$$i = 1,2,\dots,n \tag{23}$$

where $\epsilon_t \sim N(0, \sigma^2 \Omega)$, $\epsilon_t = \rho \epsilon_{t-1} + V_t$ and $V_t \sim N(0, \sigma_V^2)$. The true value of the vector coefficient β is set to $\beta = (2, 0.4, 1.3, 2.5)$.

Fuller [3] suggested the following method to generate AR(1)

$$\epsilon_t \sim N(0, \frac{\sigma_V^2}{1-\rho^2})$$

Eledum [1] modified Fuller method by

If $t = 1$:

$$\epsilon_t \sim N\left(0, sd_{\epsilon_1}\right), sd_{\epsilon_1} = \frac{sd_{v_t}}{\sqrt{(1-\rho^2)}} \quad \text{and} \quad \rho\epsilon_{t-1} = \epsilon_1 - v_1$$

if $t > 1$:

$$\rho\epsilon_{t-1} = \rho\epsilon_t \text{ and } \epsilon_t = \rho\epsilon_{t-1} + v_t$$

The Fuller method after modification is used to generate AR(1). Based on George et al. [4], three different sets of autocorrelation coefficient are considered corresponding to $\rho = 0.5, 0.7,$ and 0.99, which refers to a low, medium and high autocorrelation, respectively. To investigate the effect of the sample sizes on the performance of the estimator under study TR, we have chosen n to be equal to 10, 30, 70, and 100, which refers to small, medium, large, and extra-large sample sizes, respectively. The experiment is replicated 10000 times.

6. Results of the Simulation

The results of the simulation are explained in tables 3, 4, 5, 6 and 7. Table 3 represents the MSE for LS and TP methods under different sample sizes, different sets of autocorrelation coefficient and different degrees of correlation between explanatory variables. However, tables 4 and 5 demonstrate the MSE for RR and TR for different values of k, under different sample sizes, different autocorrelation coefficient and different degrees of correlation between explanatory variables. Moreover, tables 6 and 7 show k's values of RR and TR under different sample sizes, sets of autocorrelation, and degrees of the correlation between explanatory variables, respectively.

From table 3, the following conclusions can be drawn:

1. As the sample size increases, MSE decreases for the two methods of the estimation LS and TP; this is true for different sets of autocorrelation coefficient and different degrees of multicollinearity. For example, the values of MSE for LS when n=10,30,70,100, $\gamma^2 = 0.70$ and $\rho = 0.5$ are 0.0948, 0.0218, 0.0087 and 0.0060 respectively, whilst the corresponding values for TP method are 0.0764, 0.0153, 0.0058, and 0.0039 respectively. This can also be seen in figure 1.
2. MSE for TP method is less than that for LS for all samples size, different sets of autocorrelation coefficient and different degrees of multicollinearity. For example, when n=10,30,70,100, $\gamma^2 = 0.99$ and $\rho = 0.99$, the values of MSE for the (LS,TP) are (3.5793, 2.9723),(1.5704, 0.3897),(1.0062, 0.1076),and (0.8132, 0.0682), respectively. (See figure 2).
3. Whatever method of estimation LS or TP, sample size and autocorrelation coefficient, as the degree of multicollinearity increases, the MSE increases. For example, at n=100 and $\rho = 0.99$, MSE for LS when $\gamma^2 = 0.7, 0.8,$ and 0.99 is 0.0316, 0.0448 and 0.8132, respectively. (See figure 3).
4. For data that suffer from multicollinearity, whatever sample size and degree of multicollinearity, as the autocorrelation coefficient increases, MSE also increases. For example, at n=100 and $\gamma^2 = 0.99$, MSE for TS when $\rho = 0.5, 0.7,$ and 0.99 is 0.1464, 0.2030 and 0.8132, respectively. (See figure 4).

From tables 4 and 5 we observe that:

1. For all techniques of estimating k in the two methods RR and TR, different sets of autocorrelation coefficient and different degrees of multicollinearity, as the sample size increases MSE decreases. For example MSE when n=10,30,70,100, ρ= 0.5 and γ^2= 0.70 for Hoerl and Kenard original and modified (HK, MHK) is (0.0692, 0.0585),(0.0206,0.0147),(0.0086,0.0057) and (0.0059, 0.0039), respectively.(See figures 5 and 6).
2. Whatever sample size, autocorrelation coefficient and degree of multicollinearity, the TR estimator has minimum MSE comparing with the RR. For example, for n=10,30,70,100, ρ= 0.70 and γ^2= 0.80, MSE for Thisted original and modified (TH, MTH) is (0.1024, 0.0804), (0.036, 0.0189), (0.016, 0.0069), and (0.0113, 0.0047), respectively. (See figures 7).

Based on the results in tables 6 and 7, we see that whatever sample size, autocorrelation coefficient, and degree of multicollinearity, the techniques of estimating k in RR (HK, HKB, TH and LW) is greater in values than their corresponding one of TR (MHK, MHKB, MTH and MLW). For example, for n=10, ρ= 0.99 and γ^2= 0.80 the values of K (HK, HKB, TH, LW) is (0.3391, 0.6716, 0.2239, 0.4588) and the corresponding (MHK, MHKB, MTH, MLW) is (0.2608, 0.5181, 0.1727, 0.3413). (See figure 8).

Table 3. MSE of LS and TP for n=10,30,70,100 , γ^2= 0.70,0.80,0.99 and ρ= 0.5,0.70,0.99

ρ	γ^2= 0.70		γ^2= 0.80		γ^2= 0.99	
	$\hat{\beta}_{LS}$	$\hat{\beta}_{TP}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{TP}$	$\hat{\beta}_{LS}$	$\hat{\beta}_{TP}$
Sample size n= 10						
0.5	0.0948	0.0764	0.1289	0.1045	2.3830	1.9116
0.7	0.1066	0.0780	0.1467	0.1068	2.6646	1.9181
0.99	0.1407	0.1159	0.2029	0.1622	3.5793	2.9723
Sample size n= 30						
0.5	0.0218	0.0153	0.0302	0.0212	0.5282	0.3710
0.7	0.0283	0.0143	0.0394	0.0198	0.6943	0.3495
0.99	0.0623	0.0160	0.0890	0.0222	1.5704	0.3897
Sample size n= 70						
0.5	0.0087	0.0058	0.0120	0.0079	0.2117	0.1396
0.7	0.0119	0.0051	0.0165	0.0070	0.2908	0.1233
0.99	0.0393	0.0045	0.0556	0.0062	1.0062	0.1076
Sample size n= 100						
0.5	0.0060	0.0039	0.0083	0.0054	0.1464	0.0948
0.7	0.0083	0.0034	0.0116	0.0047	0.2030	0.0818
0.99	0.0316	0.0028	0.0448	0.0039	0.8132	0.0682

Table 4. MSE for RR under the four techniques of estimating k when n=10,30,70,100 , γ^2=0.70,0.80,0.99 and ρ= 0.5,0.70,0.99

ρ	γ^2= 0.70				γ^2= 0.80				γ^2= 0.99			
	HK	HKB	TH	LW	HK	HKB	TH	LW	HK	HKB	TH	LW
Sample size n= 10												
0.5	0.0692	0.0610	0.0737	0.0685	0.0864	0.0755	0.0933	0.0871	0.9306	0.9520	0.9996	1.6418
0.7	0.0749	0.0661	0.0801	0.0749	0.0943	0.0828	0.1024	0.0959	1.0338	1.0586	1.1061	1.8813
0.99	0.0902	0.0799	0.0974	0.0916	0.1163	0.1040	0.1273	0.1242	1.3616	1.4018	1.4422	2.6826
Sample size n= 30												
0.5	0.0206	0.0196	0.0209	0.0204	0.0277	0.0259	0.0283	0.0275	0.2660	0.2384	0.3037	0.2904
0.7	0.0261	0.0245	0.0266	0.0258	0.0350	0.0321	0.0360	0.0346	0.3321	0.3060	0.3776	0.3914
0.99	0.0496	0.0443	0.0522	0.0485	0.0653	0.0569	0.0698	0.0644	0.6785	0.6573	0.7430	1.0856
Sample size n= 70												
0.5	0.0086	0.0084	0.0086	0.0085	0.0117	0.0113	0.0118	0.0116	0.1375	0.1158	0.1525	0.1374
0.7	0.0116	0.0113	0.0117	0.0115	0.0158	0.0152	0.0160	0.0157	0.1719	0.1470	0.1940	0.1741
0.99	0.0331	0.0299	0.0345	0.0324	0.0440	0.0389	0.0464	0.0433	0.4491	0.4266	0.4983	0.6517
Sample size n= 100												
0.5	0.0059	0.0059	0.0060	0.0059	0.0082	0.0080	0.0082	0.0081	0.1048	0.0883	0.1138	0.1044
0.7	0.0082	0.0080	0.0082	0.0081	0.0112	0.0109	0.0113	0.0112	0.1325	0.1113	0.1468	0.1323
0.99	0.0272	0.0248	0.0282	0.0267	0.0364	0.0324	0.0382	0.0358	0.3706	0.3495	0.4128	0.5212

Table 5. MSE for TR under the four methods of estimating k when n=10,30,70,100 , γ^2=0.70,0.80,0.99 and ρ= 0.5,0.70,0.99

ρ	γ^2= 0.70				γ^2= 0.80				γ^2= 0.99			
	MHK	$MHKB$	MTH	MLW	MHK	$MHKB$	MTH	MLW	MHK	$MHKB$	MTH	MLW
Sample size n= 10												
0.5	0.0585	0.0521	0.0618	0.0581	0.0738	0.0649	0.0791	0.0742	0.8062	0.8263	0.8672	1.4604
0.7	0.0592	0.0527	0.0625	0.0592	0.0750	0.0660	0.0804	0.0754	0.7939	0.8122	0.8546	1.4573
0.99	0.0776	0.0690	0.0833	0.0789	0.0991	0.0880	0.1077	0.1043	1.1089	1.1457	1.1788	2.1804
Sample size n= 30												
0.5	0.0147	0.0142	0.0149	0.0146	0.0200	0.0190	0.0203	0.0198	0.2056	0.1797	0.2330	0.2148
0.7	0.0138	0.0133	0.0139	0.0137	0.0187	0.0178	0.0189	0.0186	0.1942	0.1684	0.2200	0.2018
0.99	0.0152	0.0146	0.0154	0.0151	0.0206	0.0194	0.0210	0.0205	0.1998	0.1723	0.2284	0.2089
Sample size n= 70												
0.5	0.0057	0.0056	0.0057	0.0057	0.0078	0.0076	0.0078	0.0078	0.1013	0.0855	0.1095	0.1010
0.7	0.0050	0.0050	0.0050	0.0050	0.0069	0.0068	0.0069	0.0069	0.0919	0.0779	0.0987	0.0916
0.99	0.0044	0.0044	0.0045	0.0044	0.0061	0.0060	0.0061	0.0061	0.0818	0.0693	0.0874	0.0816
Sample size n= 100												
0.5	0.0039	0.0039	0.0039	0.0039	0.0053	0.0053	0.0054	0.0053	0.0750	0.0647	0.0795	0.0748
0.7	0.0034	0.0034	0.0034	0.0034	0.0047	0.0046	0.0047	0.0047	0.0665	0.0579	0.0700	0.0663
0.99	0.0028	0.0028	0.0028	0.0028	0.0039	0.0039	0.0039	0.0039	0.0569	0.0501	0.0595	0.0568

Table 6. Values of k of RR for n=10,30,70,100 , γ^2=0.70,0.80,0.99 and ρ= 0.5,0.70,0.99

ρ	γ^2= 0.70				γ^2= 0.80				γ^2= 0.99			
	HK	HKB	TH	LW	HK	HKB	TH	LW	HK	HKB	TH	LW
Sample size n= 10												
0.5	0.2281	0.4621	0.1540	0.2971	0.2173	0.4398	0.1466	0.2678	0.1017	0.1916	0.0639	0.2478
0.7	0.2543	0.5144	0.1715	0.3350	0.2484	0.4981	0.1660	0.3125	0.1075	0.2018	0.0673	0.2944
0.99	0.3515	0.7014	0.2338	0.4857	0.3391	0.6716	0.2239	0.4588	0.1246	0.2346	0.0782	0.4147
Sample size n= 30												
0.5	0.2270	0.4821	0.1607	0.2733	0.2260	0.4783	0.1594	0.2556	0.1640	0.3111	0.1037	0.2286
0.7	0.3058	0.6440	0.2147	0.3685	0.3033	0.6368	0.2123	0.3444	0.1973	0.3694	0.1231	0.3084
0.99	0.8671	1.7583	0.5861	1.0892	0.8413	1.6947	0.5649	1.0332	0.3329	0.6269	0.2090	0.8872
Sample size n= 70												
0.5	0.2269	0.4885	0.1628	0.2716	0.2253	0.4846	0.1615	0.2530	0.1997	0.3920	0.1307	0.2277
0.7	0.3213	0.6900	0.2300	0.3842	0.3205	0.6841	0.2280	0.3596	0.2652	0.5107	0.1702	0.3237
0.99	1.7672	3.6188	1.2063	2.1697	1.7001	3.4245	1.1415	2.0317	0.7401	1.3859	0.4620	1.8011
Sample size n= 100												
0.5	0.2261	0.4892	0.1631	0.2705	0.2267	0.4894	0.1631	0.2546	0.2105	0.4203	0.1401	0.2283
0.7	0.3254	0.7015	0.2338	0.3887	0.3266	0.7020	0.2340	0.3665	0.2878	0.5665	0.1888	0.3282
0.99	2.3301	4.7446	1.5815	2.8521	2.2442	4.5351	1.5117	2.6660	1.0016	1.8832	0.6277	2.3705

Table 7. Values of of TR for n=10,30,70,100 , ^{2}=0.70,0.80,0.99 and ρ= 0.5,0.70,0.99

ρ	^{2}= 0.70				^{2}= 0.80				^{2}= 0.99			
	MHK	$MHKB$	MTH	MLW	MHK	$MHKB$	MTH	MLW	MHK	$MHKB$	MTH	MLW
Sample size n= 10												
0.5	0.1740	0.3548	0.1183	0.2265	0.1669	0.3393	0.1131	0.2048	0.0818	0.1533	0.0511	0.1900
0.7	0.1737	0.3551	0.1184	0.2281	0.1698	0.3446	0.1149	0.2093	0.0857	0.1594	0.0531	0.1953
0.99	0.2657	0.5358	0.1786	0.3592	0.2608	0.5181	0.1727	0.3413	0.1161	0.2186	0.0729	0.3086
Sample size n= 30												
0.5	0.1523	0.3260	0.1087	0.1840	0.1523	0.3252	0.1084	0.1725	0.1222	0.2347	0.0782	0.1551
0.7	0.1401	0.3004	0.1001	0.1696	0.1404	0.3006	0.1002	0.1593	0.1173	0.2261	0.0754	0.1443
0.99	0.1579	0.3387	0.1129	0.1917	0.1584	0.3400	0.1133	0.1801	0.1323	0.2557	0.0852	0.1611
Sample size n= 70												
0.5	0.1444	0.3125	0.1042	0.1734	0.1429	0.3090	0.1030	0.1608	0.1342	0.2702	0.0901	0.1450
0.7	0.1252	0.2715	0.0905	0.1505	0.1253	0.2711	0.0904	0.1410	0.1198	0.2424	0.0808	0.1274
0.99	0.1093	0.2371	0.0790	0.1314	0.1092	0.2363	0.0788	0.1231	0.1059	0.2180	0.0727	0.1103
Sample size n= 100												
0.5	0.1414	0.3068	0.1023	0.1695	0.1416	0.3069	0.1023	0.1593	0.1366	0.2783	0.0928	0.1429
0.7	0.1222	0.2652	0.0884	0.1465	0.1224	0.2655	0.0885	0.1377	0.1183	0.2432	0.0811	0.1226
0.99	0.1001	0.2174	0.0725	0.1202	0.1007	0.2185	0.0728	0.1133	0.0993	0.2061	0.0687	0.1015

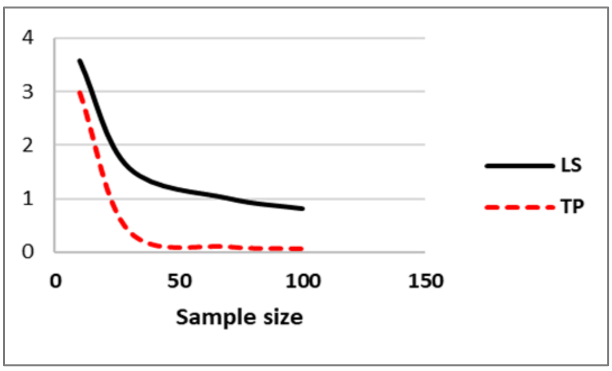

Figure 1. MSE of LS and TP for n=10,30,70,100 ,γ^2= 0.99 and ρ= 0.99

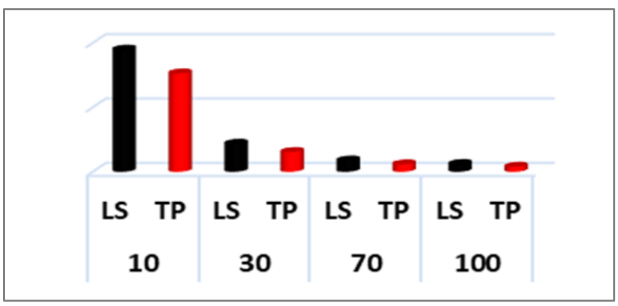

Figure 2. MSE of LS and TP for n=10,30,70,100 ,γ^2= 0.70 and ρ= 0.50

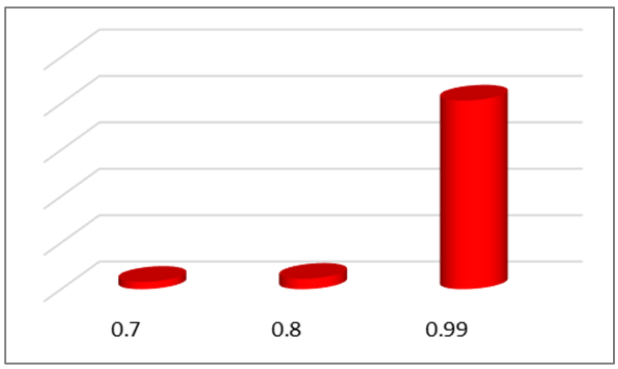

Figure 3. MSE of LS for n=100, ρ= 0.99 and ^{2}= 0.70, 0.08, 0.99

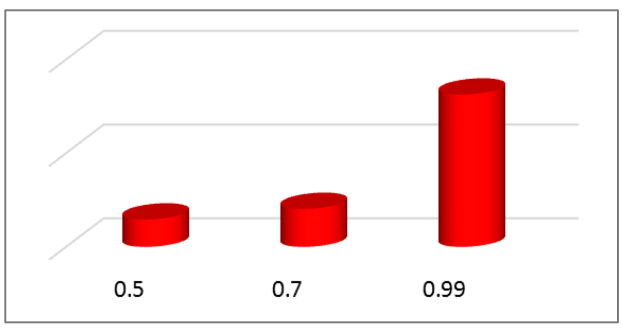

Figure 4. MSE of LS and TP for n=100, γ= 0.99 and ρ=0.5, 0.7, 0.99

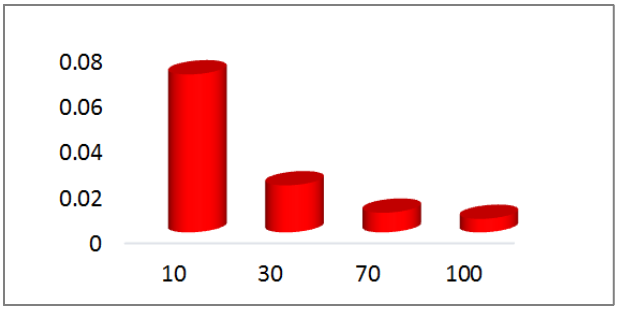

Figure 5. MSE for Hoerl and Kenard technique when n=10,30,70,100 , ρ= 0.50 and γ^2= 0.70

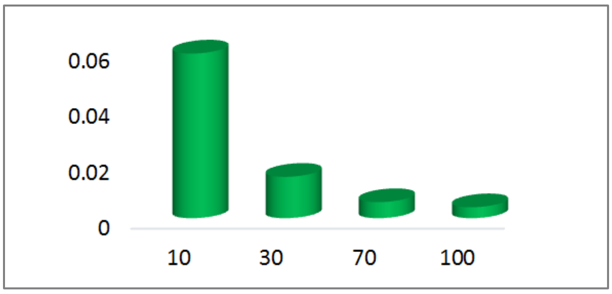

Figure 6. MSE for Modified Hoerl and Kenard technique when n=10,30,70,100 , ρ= 0.50 and ^{2}= 0.70

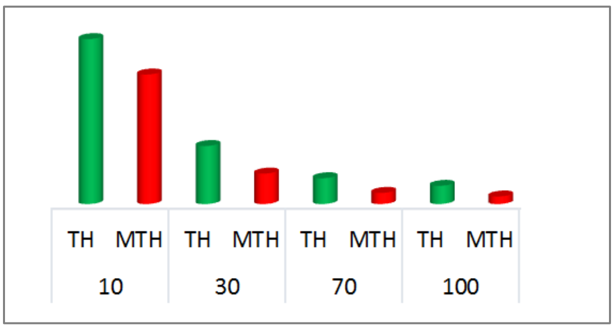

Figure 7. MSE for RR using TH and TR using MTH when n=10,30,70,100 , ρ= 0.70 and ^{2}= 0.80

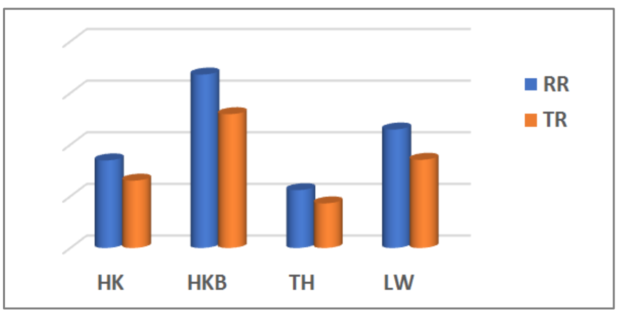

Figure 8. MSE of RR and TR for the four techniques of estimating K when n=10,30,70,100 , γ^2=0.7 and ρ= 0.99

7. Conclusions

In this paper, Two-Stage Ridge Regression TR is used to deal with the two problems, autocorrelation and multicollinearity, simultaneously. Some new properties were droved for TR, and some techniques for estimating the ridge regression parameter k are considered. The performances of the TR estimator is evaluated through Monte-Carlo Simulation, where levels of multicollinearity, sample sizes, and autocorrelation coefficients have been varied. The performance evaluation was done using the mean square error. The simulation showed that the TR estimator performs better than RR, and the value of k for TR is always less than that of RR.

REFERENCES

[1] H. Eledum. A first book of R, Noor Publishing, Germany, ISBN-13 : 978-3330840010, 2016.

[2] Eledum H, Zahri M. Relaxation method for two stages ridge regression estimator. International Journal of Pure and Applied Mathematics. 2013;85(4):653-67.

[3] WA. Fuller. Introduction to statistical time series. John Wiley & Sons; 2009 Sep 25.

[4] S. D. George, C. M. George. A simulation study of least squares and ridge estimators for normal and non normal Autocorrelated disturbances, Journal of Statistical Computation and Simulation, 47(1-2), 49-66, 2007.

[5] A. E. Hoerl, R. W. Kennard, K. F. Baldwin. Ridge regression: Some simulation. Communications in Statistics – Simulation and Computation, 4(2), 105–123, 1975.

[6] A.E. Hoerl, R.W. Kennard. Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12 (1), 55–67. doi:10.1080/00401706.1970.10488634.,1970a.

[7] A.E.Hoerl, R.W. Kennard. Ridge regression: applications to nonorthogonal problems. Technometrics, 12 (1), 69–82. doi: 10.2307/1267352, 1970b.

[8] B. M. G. Kibria. Performance of some new ridge regression estimators. Communications in Statistics-Simulation and Computation, 32, 419-435, 2003.

[9] J. F. Lawless, P.Wang. A simulation study of ridge and other regression estimators. Communications in Statistics A, 5, 307-323, 1976.

[10] D. C. Montgomery, E. A. Peck, G. G. Vining. Introduction to Linear Regression Analysis. (5th ed.). John Wiley and Sons. ISBN: 978-0-470-54281-1, 2012.

[11] Myers, Raymond H., and Raymond H. Myers. Classical and modern regression with applications. Vol. 2. Belmont, CA: Duxbury press, 1990.

[12] S. Safi. The Efficiency of OLS In The Presence Of Auto-Correlated Disturbances in Regression Models. Thesis of PhD in Statistics, American University library, Washington D.C.20016, 2004.

[13] R. A. Thisted. Ridge Regression, Minimax Estimation and Empirical Bayes Methods. *Technical Report 28, Division of Biostatistics*, Stanford University, California, 1976.