Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.8, 2013

A Proposed Nth – Order Jackknife Ridge Estimator for Linear Regression Designs

Mbe Egom Nja
Dept. of Mathematics, Federal University Lafia, Nigeria.
e-mail address: mbe_nja@yahoo.com

Abstract
Several remediation measures have been developed to circumvent the problem of collinearity in General Linear Regression Designs. These include the Generalized Ridge, Jackknife Ridge, second- order Jackknife Ridge estimation procedures. In this paper, an nth-order Jackknife Ridge estimator is developed using canonical parameter transformation. Using the MATLAB version 7 software, parameter estimates, biases and variances of these estimators are computed to show their behavior and strengths. The results show that the parameter estimates are basically the same for all the methods. There is variance reduction at the Generalized Ridge estimator and at the ordered Jackknife Ridge estimators, though the Generalized Ridge estimator is slightly superior in this respect. As the order of Jackknife Ridge estimator increases, the variance decreases up to a certain nth-order and remains constant thereafter. Where variances of two consecutive estimators are the same or nearly so, the last but one estimator is considered optimal. This establishes a convergence criterion for the sequence of Jackknife Ridge estimators. It is shown from the five illustrative design matrices that higher order Jackknife Ridge estimators are superior to lower order Jackknife Ridge estimators in terms of bias. Thus further solving the problem of bias introduced by the Ridge estimator.
Keywords: Canonical transformation; collinearity; mean square error; positive definite matrix; squared bias; variance inflation.

1. INTRODUCTION

One of the major consequences of multicollinearity on the Ordinary Least Squares (OLS) method of estimation is that it produces large variances for the estimated regression coefficients (Batah et al, 2008; Batah, 2011; Khurana, chaubey et al, 2012). This leads to poor prediction in certain regions of the regression space (Lesaffre & Marx,1993) as one is unable to determine the effect of each predictor on the response. A set of variables is said to be collinear if one or more variables in the set can be expressed exactly or nearly as a linear combination of the others in the set. To remedy this situation, several measures have been offered. These include: the Ridge regression method (Hoerl & Kendal, 1970; Lesaffre & Marx, 1993, Carley & Natalia, 2004, Belsley, et al 1980; Hawkins & Yin, 2002; Kleinbaum, et al, 1998; Maddala, 1992; Ngo et al, 2003; Mardikyan & Cetin, 2008), the iterative principal component method (Marx & Smith,1990b).

Ridge regression seeks to find a set of regression coefficients that is more stable in the sense of having a small Mean Square Error (MSE) since multicollinearity (collinearity) usually results in ordinary least square estimators that may have extremely large variances (Nelder & Wedderburn, 1972, Carley et al 2004). The ridge technique enlarges the small eigenvalue(s), thus decreasing the MSE which is defined as

$MSE = \sigma^2 T_r[(X'X)^{-1}] = \sigma^2 \sum_{j=1}^{h} \frac{1}{\lambda_j}$ where λ_j is the jth eigenvalue of $X'X$.

Large MSE suggests that estimated parameters may be far from the true ones. That is the variances of parameter estimates are inflated. Khurana et al (2012) developed the second-order Jackknife Ridge estimator (J2R) to further reduce the bias associated with the Ordinary Ridge estimator. The J2R estimator is superior to the Jackknife ridge estimator (JRE) in terms of bias. They also showed that the Jackknife Ridge estimator (JRE) is superior to the Generalized Ridge estimator (GRE) which in turn has a lower bias than the Modified Jackknife Ridge estimator (MJRE).

In this paper a third-order Jackknife estimator is developed and a general nth-order Jackknife estimator is proposed as a result. The following design matrices each with collinearity among columns are used for illustration:

X_1 = [1,1,1,1; 2,3,4,5; 4,6,8,10; 15,22,19,40]
X_2 = [1,1,1,1; 2,3,2,4; 4,6,4,8; 15,22,19,40]
X_3 = [1,1,1,1; 1,2,3,4; 2,4,6,8; 15,22,19,40]
X_4 = [1,1,1,1; 3,4,3,5; 6,8,6,10; 15,22,19,40]
X_5 = [1,1,1,1; 4,5,4,6; 8,10,8,12; 15,22,19,40]

The result of the analysis shows that the bias of the Jackknife Ridge Estimators, beginning from the first order up to the fifth order reduces at the speed of 10^{-4} per order. For instance, using matrix 1, the second eigenvalue λ_2, and the 4^{th} category, the sequence of bias for the indicated estimators reduces at the speed of 10^{-4} per order and is as follows: 0.397×10^{-3}, 4.943×10^{-7}, 0.614×10^{-11}, 0.763×10^{-15}, 0.949×10^{-19}. The variances of the Jackknife estimators are basically constant but smaller than the variance of the Ordinary Least Squares estimator. Thus the advantage of the Ridge Regression estimator over the Ordinary Least Squares estimator is further improved by higher order Jackknife Ridge estimator. In section 1, the introduction is provided. In section 2, the model and Ridge estimators are introduced. Section 3 contains the proposed estimators. The comparison of the bias of the estimators is made in section 4, while section 5 contains the comparison of the MSE of the estimators. Section 6 contains the illustrative examples. The conclusion is made in section 7.

2. TH MODEL AND RIDGE ESTIMATORS

Consider a multiple linear regression model

$$Y = X\beta + \varepsilon \quad (2.1)$$

Where Y is an $(n \times 1)$ vector of observations, β is a $(p \times 1)$ vector of unknown regression coefficients, X is an $(n \times p)$ matrix of explanatory variables $X_1, X_2, \ldots\ldots, X_p$ and ε is an $(n \times 1)$ vector of errors, the elements of which are assumed to be independently and identically normally distributed with

$E(\varepsilon) = 0 \; and \; var(\varepsilon) = \sigma^2 I$

The ordinary least square estimator known as best linear unbiased estimator corresponding to (2.1) is given as

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2.2)$$

Although the OLS estimator in (2.2) is unbiased, it has the problem of inflated variance where collinearity is present which may result in estimators that are not in tune with the researcher's prior belief. As a remediation measure, Hoerl & Kennard (1970) proposed the Generalized Ridge estimator (GRE). The Ridge estimator solves the problem inflated variance but is biased. It is given as

$$\hat{\beta}_{RE} = (X'X + \theta I)^{-1}X'Y \quad (2.3)$$

for some biasing constant θ. To alleviate the problem of bias in Ridge regression, Singh et al (1986) proposed an Almost Unbiased Ridge estimator (AUGRR) using the Jackknife technique which was actually introduced by Quenouille(1956).The Ridge regression estimator has undergone several modifications over the years. Batah et al (2008) introduced the modified Jackknife Ridge estimator (MJRE) by combining the ideas of GRE and Jackknife Ridge estimators. Batah (2011) again introduced another variant of the Jackknife estimator known as the Generalized Jackknife Ridge estimator (GJR) together with its associated Generalized Jackknife Ordinary Ridge estimator (GOJR).The Generalized Ridge estimator (GRE) leads to a reduction in the sampling variance whereas the Jackknife Ridge estimator (JRE) leads to the reduction of bias. Khurana, Chaubey & Chandra (2012) suggested a new Ridge estimator called the second- order Jackknife Ridge estimator (J2R) to further reduce bias. In this paper, a generalization for the nth-order Jackknife Ridge estimator is proposed for further reduction of bias. A convergence criterion is designed to select the optimal n so that an estimator with minimum n is selected for the most reduced bias and minimum variance. In canonical form model (1) can be written as

$$Y = Z \propto + \varepsilon \quad (2.4)$$

Where $Z = XT, T$ is the matrix of eigenvectors of $X'X$.

$Z'Z = J = T'X'XT = J = diag(\lambda_1, \lambda_2, \ldots\ldots, \lambda_p$

where λ_i is the ith eigenvalue of $X'X$

$\propto = T'\beta$

$T'T = TT' = I_P$

Then OLS estimator of $\propto$ is given by

$\hat{\propto} = (Z'Z)^{-1}Z'Y = J^{-1}Z'Y$

Thus the OLS estimator of β is given as

$$\hat{\beta}_{OLS} = T\,\hat{\propto}_{OLS} \quad (2.5)$$

The ordinary Ridge regression estimator (ORE) of β is

$$\hat{\beta}_{ORE} = T\,\hat{\propto}_{ORE} = T(I - rA_r^{-1})\,\hat{\propto} = T(I - rA_r^{-1})J^{-1}Z'Y \quad (2.6)$$

Where $r_1 = r_2 = \cdots = r_p = r, r \geq 0$

(r = biasing constant) $= diag(r)$

$J = diag(\lambda_i)$

This is generalized to obtain the Generalized Ridge estimators of $\propto$ as

$$\hat{\propto}_{GRE} = [J + R]^{-1}Z'Y = [J + R]^{-1}J\,\hat{\propto} = [I - R(J + R)^{-1}] \propto = [I - RA^{-1}]\,\hat{\propto} \quad (2.7)$$

where J is a $(p \times p)$ diagonal element whose non-negative entries are eigenvalues of $X^I X$. R is a $(p \times p)$ diagonal matrix whose non-negative entries and biasing constants.

$$\hat{\beta}_{GRE} = T\,\hat{\propto}_{GRE} = T[I - RA^{-1}]\,\hat{\propto}_{GRE} \quad (2.8)$$

The Jackknife Ridge Estimator for $\propto$ is

$$\widehat{\propto}_{JRE}=[I-(RA^{-1})^2]\,\widehat{\propto}_{OLS}=[I-R^*A^{-1}]\,\widehat{\propto}_{OLS} \quad (2.9)$$

The second-order Jackknife Ridge Estimator is

$$\widehat{\propto}_{J2R}=[I-(R^*A^{-1})^2]\,\widehat{\propto}_{OLS} \text{ Where } R^*=RA^{-1}R \quad (2.10)$$

3. THE PROPOSED ESTIMATOR

By extending the second-order Jackknife Ridge estimator to a proposed third-order Ridge estimator, a general nth-order Ridge estimator is proposed. The proposed third-order Jackknife Ridge estimator (J3R) denoted by $\widehat{\propto}_{J3R}$ is defined as

$$\widehat{\propto}_{J3R}=[I-(R^*A^{-1})^3]\,\widehat{\propto}_{OLS}$$
$$=[I-(R^2A^{-2})^3]\,\widehat{\propto}_{OLS} \quad (3.1)$$

The proposed nth-order Jackknife Ridge regression estimator (JnR) is defined as

$$\widehat{\propto}_{JnR}=[I-(R^*A^{-1})^n]\,\widehat{\propto}_{OLS} \quad (3.2)$$

where n=0,1,2,……….

It is observed that an optimal n is achieved when the sequence of the variances of the estimator converges to the variance of the ordinary Least Squares estimator.

that is $var(\widehat{\propto}_{JnR}) \to var(\widehat{\propto}_{OLS})$ *as* $n \to \infty$.

4. COMPARISON OF THE BIAS OF J2R, J3R AND JnR ESTIMATORS.

The following theorems compare the performance of J2R and J3R and that of J(n-1)R and JnR in terms of bias. this comparison is similar to that of khurana et al (2012).

Theorem 1

Let R be a $(p \times p)$ diagonal matrix with non-negative entries. Then the third-order Jackknife estimator $\widehat{\propto}_{J3R(i)}$, reduces the bias of the second-order Jackknife Ridge estimator $\widehat{\propto}_{J2R(i),}$ assuming $r_i > 0$.

Proof

The Jackknife Ridge Estimator for $\propto$ is

$$\widehat{\propto}_{JRE}=[I-(RA^{-1})^2]\,\widehat{\propto}_{OLS}=[I-R^*A^{-1}]\,\widehat{\propto}_{OLS}$$

The second-order Jackknife Ridge Estimator is

$$\widehat{\propto}_{J2R}=[I-(R^*A^{-1})^2]\,\widehat{\propto}_{OLS} \textit{ where } R^*=RA^{-1}R$$

$$\text{Bias}\left(\widehat{\propto}_{JRE}\right)=E\left(\widehat{\propto}_{JRE}\right)-\widehat{\propto}_{OLS} \quad (4.1)$$
$$=E(\widehat{\propto}-R^*A^{-1}\propto)-\widehat{\propto}_{OLS}$$
$$=-R^*A^{-1}\,\widehat{\propto}_{OLS}$$

$$\text{Similarly bias}\left(\widehat{\propto}_{J3R}\right)=E\left(\widehat{\propto}_{J3R}\right)-\widehat{\propto}_{OLS} \quad (4.2)$$
$$=-(R^*A^{-1})^2\,\widehat{\propto}_{OLS}$$

$$\text{and bias}\left(\widehat{\propto}_{J3R}\right)=E\left(\widehat{\propto}_{J3R}\right)-\widehat{\propto}_{OLS} \quad (4.3)$$
$$=-(R^*A^{-1})^3\,\widehat{\propto}_{OLS}$$

Let $|.|_i$ denote the absolute value of the $i-th$ component.

$$\text{Then}\left|Bias(\widehat{\propto}_{J2R})\right|_i=\frac{(r_i^*)^2}{(\lambda_i+r_i)^2}\mid\propto_i\mid=\frac{(r_i^2)^2}{(\lambda_i+r_i)^2}\times\frac{1}{(\lambda_i+r_i)^2}\mid\propto_i\mid$$
$$=\frac{r_i^4}{(\lambda_i+r_i)^4}\mid\propto_i\mid \quad (4.4)$$

$$\text{and}\left|Bias(\widehat{\propto}_{J3R})\right|_i=\frac{(r_i^*)^3}{(\lambda_i+r_i)^2}|\lambda_i|=\frac{r_i^6}{(\lambda_i+r_i)^6}\mid\propto_i\mid \quad (4.5)$$

$$\text{Thus }\left|Bias\left(\widehat{\propto}_{J2R}\right)\right|_i-\left|Bias\left(\widehat{\propto}_{J3R}\right)\right|_i$$
$$=\frac{r_i^4}{(\lambda_i+r_i)^4}|\propto_i|-\frac{r_i^6}{(\lambda_i+r_i)^6}\mid\propto_i\mid \quad (4.6)$$
$$=\frac{(\lambda_i^2r_i^4+2\lambda_ir_i^5)|\propto_i|}{(\lambda_i+r_i)^6}>0 \textit{ as } r_i>0$$

This proves that the third-order Jackknife reduces the bias of the second-order Jackknife ridge estimator.It has earlier been proven (Khurana et al 2012) that the second-order Jackknife reduces the bias of the Jackknife ridge estimator which in turn reduces the bias of the ordinary ridge estimator.This paper proposes a general extension of the third-order Jackknife Ridge estimator to an nth-order Jackknife Ridge estimator since it is shown that higher-order Jackknife Ridge estimators have lower biases.

Theorem 2

Let R be a (pxp) diagonal matrix with non-negative entries. Then the nth-order Jackknife estimator $\widehat{\propto}_{JnR(i)}$reduces the bias of the (n-1)th-order Jackknife Ridge estimator $\widehat{\propto}_{J(n-1)R(i)}$, assuming $r_i\epsilon(0,1)$.

Proof

The Jackknife, second-order Jackknife and third-order Jackknife Ridge Estimators for $\propto$ are given respectively as:

$\hat{\alpha}_{JRE} = [I - R^{*}A^{-1}]\,\hat{\alpha}_{OLS}$, $\hat{\alpha}_{J2R} = [I - (R^{*}A^{-1})^{2}]\,\hat{\alpha}_{OLS}$
and $\hat{\alpha}_{J3R} = [I - (R^{*}A^{-1})^{3}]\,\hat{\alpha}_{OLS}$
where $R^{*} = RA^{-1}R$
Thus the (n-1)th-order Jackknife Ridge estimator for $\propto$ can be defined as

$$\hat{\alpha}_{J(n-1)R} = \left[I - (R^{*}A^{-1})^{(n-1)}\right]\hat{\alpha}_{OLS} \quad (4.7)$$

and the nth-order Jackknife Ridge estimator is

$$\hat{\alpha}_{JnR} = [I - (R^{*}A^{-1})^{n}]\,\hat{\alpha}_{OLS}$$

$$Bias\left(\hat{\alpha}_{J(n-1)R}\right) = E\left(\hat{\alpha}_{J(n-1)R}\right) - \hat{\alpha}_{OLS} \quad (4.8)$$

$$= -(R^{*}A^{-1})^{n-1}\,\hat{\alpha}_{OLS}$$

Let $|.|_i$ denote the absolute value of the $i-th$ component

$$\text{Then } |Bias(\hat{\alpha}_{J(n-1)R})|_i = \frac{[r_i^{*}]^{(n-1)}}{(\lambda_i+r_i)^2}|\propto_i| = \frac{\left[r_i^{(n-1)}\right]^{(n-1)}}{[\lambda_i+r_i]^{(n-1)}} \times \frac{1}{[\lambda_i+r_i]^2}|\propto_i| = \frac{r^{(n-1)^2}}{[\lambda_i+r_i]^{2(n-1)}}|\propto_i| \quad (4.9)$$

$$|Bias(\hat{\alpha}_{JnR})|_i = \frac{(r_i^{*})^n}{(\lambda_i+r_i)^2} = \frac{r_i^{n^2}}{(\lambda_i+r_i)^{2n}}|\propto_i| \quad (4.10)$$

Hence $\left|Bias\left(\hat{\alpha}_{J(n-1)R}\right)\right|_i - \left|Bias\left(\hat{\alpha}_{JnR}\right)\right|_i$

$$= \frac{r_i^{(n-1)^2}}{[\lambda_i+r_i]^{2(n-1)}}|\propto_i| - \frac{r_i^{n^2}}{(\lambda_i+r_i)^{2n}}|\propto_i| \quad (4.11)$$

$$= |\propto_i|[\lambda_i + r_i]^{2(1-n)}\left[r_i^{(n-1)^2} - \{\lambda_i + r_i\}^{-2} r_i^{n^2}\right] > 0$$

since $r_i \epsilon (0,1)$
This proves that the $nth-order$ Jackknife estimator reduces the bias of the $(n-1)th-order$ Jackknife Ridge estimator.
The Bias of the ordered Jackknife Ridge estimators can be written as follows:

$\text{Bias}(\hat{\alpha}_{J2R}) = E[\hat{\alpha}_{J2R}] - \propto$
$= E[I - (R^2A^{-2})^2] \propto - \propto$
$= [I - (R^2A^{-2})^2]E(\hat{\alpha}) - \propto$
$= -(R^2A^{-2})^2 \propto = -(R^{*}A^{-1})^2 \propto$

Similarly,

$\text{Bias}(\hat{\alpha}_{J3R}) = -(R^2A^{-2})^3 \propto = -(R^{*}A^{-1})^3 \propto$
$\text{Bias}(\hat{\alpha}_{JnR}) = -(R^2A^{-2})^n \propto = -(R^{*}A^{-1})^n \propto$

Khurana, Chaubey & Chandra (2012) proved that the difference of total squared biases of the Jackknife Ridge Estimator (JRE) and second-order Jackknife Ridge Estimator (J2R) of β , given by

$$D_1 = \sum\{|Bias(\hat{\beta}_{JRE})|_i^2 - |Bias(\hat{\beta}_{J2R})|_i^2\}$$

$$= \beta^I\{G[(A^{-1}R^{*})^2 - (A^{-1}R^{*})^4]G^I\}\beta \quad (4.12)$$

where $GBias(\hat{\alpha}) = Bias(\hat{\beta})$ is non-negative. It is strictly positive if at least one $r_{i,}\ i = 1, \ldots, n$ *is positive*. In the theorem that follows, they further proved the same result for the difference between the modified Jackknife Ridge (MJR) and the Generalized Ridge Estimators (GRE).

Theorem 3

Let R be a (p×p) diagonal matrix with non-negative entries, then the difference of total squared biases of the modified Jackknife Ridge (MJR) and Generalized Ridge Estimators (GRE) of β as given by

$$D_2 = \sum\{\left|Bias(\hat{\beta}_{MJR})\right|_i^2 - \left|Bias(\hat{\beta}_{GRE})\right|_i^2\} \quad (4.13)$$

is positive.

Proof

$$\hat{\alpha}_{MJR} = [I - (RA^{-1})^2][I - RA^{-1}]\,\hat{\alpha} = [I - (A^{-1}\phi R)]$$

Where $\phi = (I + RA^{-1} - R^{*}A^{-1})$ and $R^{*} = RA^{-1}R$ (4.14)

$Bias(\hat{\alpha}_{MJR}) = -(A^{-1}\phi R \propto$
$\hat{\alpha}_{GRE} = [I - (RA^{-1})^2]\,\hat{\alpha}$
Thus $Bias(\hat{\alpha}_{GRE}) = -RA^{-1} \propto$

Component wise

$$|Bias(\hat{\alpha}_{MJR})|_i - |Bias(\hat{\alpha}_{GRE})|_i = \frac{r_i\phi_i}{\lambda_i+r_i}|\propto_i| - \frac{r_i}{\lambda_i+r_i}|\propto_i| \quad (4.15)$$

$$= \frac{r_i\left[1+\frac{r_i}{\lambda_i+r_i}-\frac{r_i^2}{(\lambda_i+r_i)^2}\right]}{\lambda_i+r_i}|\propto_i| - \frac{r_i}{\lambda_i+r_i}|\propto_i|$$

$$= \frac{\lambda_i r_i^2}{(\lambda_i+r_i)^3}|\propto_i|$$

which is positive.
Hence

$D_2 = \sum\{|Bias(\hat{\beta}_{MJR})|_i^2 - |Bias(\hat{\beta}_{GRE})|_i^2\}$ (4.16)
$= \beta^I\{G[(A^{-1}\phi R)^2 - (A^{-1}R)^2]G^I\}\beta$
$= \beta^I\{G[A^{-1}\phi RA^{-1}\phi R - A^{-1}RA^{-1}R]G^I\}\beta$
$= \beta^I\{GA^{-1}[(\phi R)^2 - R^2]A^{-1}G^I\}\beta$
$= \frac{\lambda_i r_i^2}{(\lambda_i+r_i)^3} | \propto_i |$

which is positive definite.
Singh et al (1986) proved that the total squared bias for the second-order Jackknife Ridge estimator is less than that of the Jackknife Ridge estimator. Putting this together with the foregoing theorem, we have the following ordering:

$SB(\hat{\beta}_{J2R}) < SB(\hat{\beta}_{JRE}) < SB(\hat{\beta}_{GRE}) < SB(\hat{\beta}_{MJR})$

where $SB = total$ squared bias.
Following the theorem that compares the bias of the second-order and third-order Jackknife Ridge estimators and that of the (n-1)th-order and nth-order Jackknife Ridge estimators stated and proven in this study, a general ordering for the total squared biases for the Modified Jackknife and ordered Jackknife Ridge estimators can be written as

$BS(\hat{\beta}_{JnR}) < BS(\hat{\beta}_{J(n-1)R}) < SB(\hat{\beta}_{J2R}) < SB(\hat{\beta}_{JRE}) < BS(\hat{\beta}_{GRE}) < BS(\hat{\beta}_{MJR})$

5. COMPARISON OF THE MSE OF J2R, J3R AND JnR ESTIMATORS

Batah et al (2008) proved that the Mean Square Error (MSE) of the Modified Jackknife Ridge (MJR) estimator is smaller than that of the Jackknife Ridge estimator. Khurana, Chaubey & Chandra (2012) proved by the theorem that follows that the JR estimator dominates the J2R estimator by MSE evaluation.

Theorem 4

Let R be a (p×p) diagonal matrix with non-negative entries. Then the difference of the MSE matrix of the second-order Jackknife estimator and the Jackknife Ridge estimator,

$\Delta_1 = MSE(\hat{\propto}_{J2R}) - MSE(\hat{\propto}_{JRE})$

is a positive definite matrix if and only if the following inequality is satisfied:

$\propto^I \{L^{-1}[\sigma^2 m + (A^{-1}R^*)^2 \propto\propto^I (R^*A^{-1})^2]L^{-1}\}^{-1} \propto \leq 1$

where $L = A^{-1}R^*, R^* = RA^{-1}R$ and (4.17)

$m = [I - A^{-1}R^*]J^{-1}[I - A^{-1}R^*]'\{[I + A^{-1}R^* + I]A^{-1}R^*\}$ (4.18)

Proof

Using the expression of the variance-covariance matrix of the least squares estimator, that is
$Var(\hat{\propto}) = \sigma^2(Z'Z)^{-1} = \sigma^2 J^{-1}$ and the expression for JRE as earlier stated, it can be written that (4.19)

$Var(\hat{\propto}_{JRE}) = E[(\hat{\propto}_{JRE} - E(\hat{\propto}_{JRE}))(\hat{\propto}_{JRE} - E(\hat{\propto}_{JRE}))^1] = [I - A^{-1}R^*]Var(\hat{\propto})[I - A^{-1}R^*]'$
$= [I - A^{-1}R^*]\sigma^2 J^{-1}[I - A^{-1}R^*]'$ (4.20)

Further using the expression for $Bias(\hat{\propto}_{JRE})$ from (4.1), it can be written that

$MSE(\hat{\propto}_{JRE}) = [I - A^{-1}R^*]\sigma^2 J^{-1}[I - A^{-1}R^*]' + A^{-1}R^* \propto\propto^{-1} R^*A^{-1}$ (4.21)

Similarly using the expression for $Var(\hat{\propto}_{J2R})$ as given by

$Var(\hat{\propto}_{J2R}) = [I - (A^{-1}R^*)^2]\sigma^2 J^{-1}[I - (A^{-1}R^*)^2]'$

and the expression for$Bias(\hat{\propto}_{J2R})$ from(4.2) we have

$MSE(\hat{\propto}_{J2R}) = [I - (A^{-1}R^*)^2]\sigma^2 J^{-1}[I - (A^{-1}R^*)^2]^1 + (A^{-1}R^*)^2 \propto\propto' (R^*A^{-1})^2$ (4.22)

from (4.21) and (4.22),Δ1 becomes

$\Delta_1 = \sigma^2 m + (A^{-1}R^*)^2 \propto\propto' (R^*A^{-1})^2 - A^{-1}R^* \propto\propto' R^*A^{-1}$

where

$m = [I - (A^{-1}R^*)^2]J^{-1}[I - (A^{-1}R^*)^2]' - [I - A^{-1}R^*]J^{-1}[I - A^{-1}R^*]'$
$= [I - A^{-1}R^*]J^{-1}[I - A^{-1}R^*]'\{[I + A^{-1}R^*][I + A^{-1}R^*]' - I\}$
$= [I - A^{-1}R^*]J^{-1}[I - A^{-1}R^*]'\{[I + A^{-1}R^* + I][I + A^{-1}R^* - I]\}$
$= [I - A^{-1}R^*]J^{-1}[I - A^{-1}R^*]'\{[I + A^{-1}R^* + I]A^{-1}R^*\}$

But $[I - A^{-1}R^*]$ is a diagonal matrix with positive entries and
$\{[I + A^{-1}R^* + I]A^{-1}R^*\}$ is also positive definite
Thus M is also positive definite and hence the difference Δ_1 is positive definite if and only if $L^{-1}\Delta_1 L^{-1}$ is positive definite.

$L^{-1}\Delta_1 L^{-1} = L^{-1}[\sigma^2 m + (A^{-1}R^*)^2 \propto\propto' (R^*A^{-1})^2]L^{-1} - \propto\propto'$

The matrix $[\sigma^2 m + (A^{-1}R^*)^2 \propto\propto' (R^*A^{-1})^2]$ in the above equation is symmetric positive definite. Therefore, it is concluded that $L^{-1}\Delta_1 L^{-1}$ is positive definite if and only if

$\propto' \{L^{-1}[\sigma^2 m + (A^{-1}R^*)^2 \propto\propto' (R^*A^{-1})^2]L^{-1}\}^{-1} \propto \leq 1$

$= [I-(A^{-1}R^*)^2]J^{-1}[I-(A^{-1}R^*)^2]'\{[I + A^{-1}R^* + I][I + A^{-1}R^* - I]\}$

$= [I-(A^{-1}R^*)^2]J^{-1}[I-(A^{-1}R^*)^2]'\{[I + A^{-1}R^* + I]A^{-1}R^*\}$ (4.23)

The matrix M is positive definite and hence the difference Δ_1 is positive definite if and only if $L^{-1}\Delta_1 L^{-1}$ is positive definite.

$L^{-1}\Delta_1 L^{-1} = L^{-1}[\sigma^2 m + (A^{-1}R^*)^3 \propto\propto' (R^*A^{-1})^3]L^{-1} - \propto\propto'$ (4.24)

The matrix $[\sigma^2 m + (A^{-1}R^*)^3 \propto\propto' (R^*A^{-1})^3]$ *in* (4.24) is symmetric positive definite. Hence using lemma 1 we conclude that $L^{-1}\Delta_1 L^{-1}$is positive definite if and only if the following inequality is satisfied.

$\propto' \{L^{-1}[\sigma^2 m + (A^{-1}R^*)^3 \propto\propto' (R^*A^{-1})^3]L^{-1}\}^{-1} \propto \leq 1$

Theorem 5

Let R be a (p×p) diagonal matrix with non-negative entries. Then the difference of the MSE matrix of the third order Jackknife estimator and the second-order Jackknife Ridge estimator,

$\Delta_2 = MSE(\propto_{J3R}) - MSE(\propto_{J2R})$ (4.25)

is a positive definite matrix if and only if the following inequality is satisfied:

$\propto' \{L^{-1}[\sigma^2 M + (A^{-1}R^*)^3) \propto\propto' (R^*A^{-1})^3]L^{-1}\}^{-1} \propto \leq 1$ (4.26)

where $L = A^{-1}R^*$, $R^* = RA^{-1}R$ and (4.27)

$A^{-1}R^*\}M = [L-(A^{-1}R^*)^2]J^{-1}[L-(A^{-1}R^*)^2]'\{[I + A^{-1}R^* + I]$ (4.28)

Proof

Using the expression of the variance – covariance matrix of the least squares estimator, that is $var(\hat{\propto}) = \sigma^2(Z'Z)^{-1}$ and the expression for $J2R$ given in (2.10), we have.

$var(\hat{\propto}_{J2R}) = E[(\hat{\propto}_{J2R} - E(\propto_{J2R}))(\hat{\propto}_{J2R} - E(\hat{\propto}_{J2R})')']$

$= [I-(A^{-1}R^*)^2]var(\hat{\propto})[I-(A^{-1}R^*)^2]'$

$= [I - A^{-1}R^*]\sigma^2 J^{-1}[I - A^{-1}R^*]'$ (4.29)

Also using the expression for Bias $(\hat{\propto}_{J2R})$ from (4.2), we obtain

$MSE(\hat{\propto}_{J2R}) = [I-(A^{-1}R^*)^2]\sigma^2 J^{-1}[I-(A^{-1}R^*)^2]' + (A^{-1}R^*)^2 \propto\propto' (R^*A^{-1})^2$ (4.30)

Similarly using the expression for Bias $(\hat{\propto}_{J3R})$ given by (4.3) we have

$var(\hat{\propto}_{J3R}) = [I-(A^{-1}R^*)^3\sigma^2 J^{-1}[-(A^{-1}R^*)^3]'$ (4.31)

$MSE(\hat{\propto}_{J3R}) = [I-(A^{-1}R^*)^3]\sigma^2 J^{-1}[I-(A^{-1}R^*)^3]' + (A^{-1}R^*)^3 \propto\propto' (R^*A^{-1})^3$ (4.32)

Thus

$\Delta_1 = MSE(\propto_{J3R}) - MSE(\propto_{J2R})$ (4.33)

$= \sigma^2 m + (A^{-1}R^*)^3 \propto\propto' (R^*A^{-1})^3 - (A^{-1}R^*)^2 \propto\propto' (R^*A^{-1})^2$

where

$m = [I-(A^{-1}R^*)^3]J^{-1}[I-(A^{-1}R^*)^3]' - [I-(A^{-1}R^*)J^{-1}[I-(A^{-1}R^*)^2]'$

$= [I-(A^{-1}R^*)^2]J^{-1}[I - A^{-1}R^*]'\{[I + A^{-1}R^*]' - I$

$= [I-(A^{-1}R^*)^2]J^{-1}[I-(A^{-1}R^*)^2]'\{[I - A^{-1}R^*][I - A^{-1}R^*]' - I$ (4.34)

These proofs (theorems 3 and 4) show that the $J2R$ does not improve on JRE and $J3R$ does not improve on $J2R$ in terms of MSE and by extension we conclude that JnR does not improve on $J(n-1)R$ in terms of MSE.

Variance

The $var(\hat{\propto}_{OLS}) = \sigma^2(Z'Z)^{-1} = \sigma^2 J^{-1}$ (4.35)

$\sigma^2 = \frac{Y'Y - \hat{\propto}'_{OLS}Z'Y}{n-p-1}, \hat{\propto} = (Z'Z)^{-1}Z'Y \text{ or } \hat{\sigma}^2 = \frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{n-p}$

$var(\hat{\propto}_{JRE}) = [I - A^{-1}R^*]\sigma^2 J^{-1}[I - A^{-1}R^*]'$ (4.36)

Where

$(I - A^{-1}R)^s = diag(\frac{\lambda_1^s}{(\lambda_1 + r_1)^s}, \dots, \frac{\lambda_i^s}{(\lambda_1 + r_1)^s} \dots, \frac{\lambda_p^s}{\lambda_p + r})$ (4.37)

$var(\hat{\propto}_{J2R}) = [I-(A^{-1}R^*)^2\sigma^2 J^{-1}[I-(A^{-1}R^*)^2]'$ (4.38)

$var(\hat{\propto}_{J3R}) = [I-(A^{-1}R^*)^3]\sigma^2 J^{-1}[I-(A^{-1}R^*)^3]'$ (4.39)

$var(\hat{\propto}_{JnR}) = [I-(A^{-1}R^*)^n]\sigma^2 J^{-1}[I-(A^{-1}R^*)^n]'$ (4.40)

These variances are now computed and compared using illustration examples.

6. ILLUSTRATIVE EXAMPLES

The biases and variances of the OLS, GR, JR, J2R and J3R estimators were computed for five design matrices each having two collinear explanatory variables. Results are reported for the $\propto$ and β estimates in each of the five design cases. Comparison is therefore made for these estimators in terms of bias and variances. The results are tabulated in the tables shown in the appendix.

7. DISCUSSION

The proposed third – order Jackknife estimator is superior to the second – order Jackknife which in turn is superior to the ordinary Jackknife estimator in terms of bias (table 3). It is also proven that the nth – order Jackknife Ridge estimator is superior to the (n-1)th Jackknife estimator by bias evaluation. It can then be concluded that higher order Jackknife Ridge estimators have lower biases which approach zero as the order n increases. In terms of variance, the higher – order Jackknife estimators are superior to the ordinary least square estimator but they converge to the ordinary least estimator with increasing n (table 2). The Jackknife Ridge and the ordered Jackknife Ridge estimators are basically the same in value but differ from the Ordinary Least Square estimators in all the five illustrative examples (table 1). This shows that the proposed higher order Jackknife estimators are superior to the existing Ordinary Ridge, Generalized Ridge and second-order Jackknife Ridge estimators.

REFERENCES

1. Batah, F.S.M, Ramanathan, T.V., Gore, S.D.(2008). The Efficiency of Modified Jackknife and Ridge Type Regression Estimators: A comparison. Surveys in Mathematics and its applications, 3 111-122
2. Batah, F.S.M.(2011). A new Estimator By Generalized Modified Jackknife Regression Estimator: Journal of Basarah Researches (Sciences), 37(4) 138-149.
3. Belsley, D, A; Edwin, K; Welsch, R.E. (1980). Regression Diagnostics: Identifying influential data and sources of collinearity, Wiley, New York, 1980
4. Carley, Kathleen M., Natalia Y. K. (2004). A Network Optimization Approach for Improving Organizational Design: Carnegie Mellon University, School of computer science, Institute for Software Research International, Technical Report CMU-ISRI-04-102.
5. Hawkins, D.M; Yin, X. (2002). A faster algorithm for ridge regression. Computational statistics and data analysis, 40, 253-262.
6. Hoerl, A. E and Kenard, R.W. (1970). Ridge regression: Biased estimation for non-orthogonal problems. Communications in Statistics- Theory and methods, 4 105-123
7. Khurana, M., Chaubey, Y.P.and Chandra, S. (2012). Jackknifing the Ridge Regression Estimator: A Revisit: Technical Report 12(1) 1-19
8. Kleinbaun, D.G.; Kupper,L. L; Muller, K.E. Nizam a. (1998). Applied Regression Analysis and other multivariable methods. Duxbury press, Washington.
9. Lesaffre, E; Marx, B.D. (1993). Collinearity in Generalyzed Linear Regression. COMMUN. STATIST-THEORY METH; 22(7) 1933-1952.
10. Maddala, G.S. (1992). Introduction to Econometrics. Macmillan, New York.
11. Mardikyan, S; Cetin, E. (2008). Efficient choice of Biasing constant for Ridge Regression International Journal of Mathematical sciences, 3(11), 527-536
12. Nelder, J; Wedderburn, R.W.M.(1972). Generalyzed Linear Models. Journal of the Royal statistical society, A 135,370-384.
13. Ngo, S.H; Kemeny, S; Deak A. (2003). Performance of the ridge regression method as applied to complex linear and nonlinear models. Chemometrics and intelligent Laboratory systems, 67,69-78.
14 .Quenonille, M.H. (1956). Notes on bias in estimation. Biometrika, 43 353-360.
15.Singh,B., Chaubey, Y.P. and Divivedi, T. D. (1986). An almost Unbiased ridge estimator.Sankhya, B48 342-360.

Appendix

Table 1 : Parameter estimates

	OLS	GR	JR	J2R	J3R
Design Matrix 1	0 0.827 0.135 3.2	0 1.2156 0.9229 2.9316	0 0.8177 0.1349 3.1997	0 0.8178 0.1349 3.1997	0 0.8178 0.1349 3.1997
Design Matrix 2	0 -14.4799 -14.4799 5.7926	0 -14.2539 -14.2539 5.750	0 -14.2539 -14.2539 5.7501	0 -14.4764 -14.4764 5.7919	0 -14.4798 -14.4798 5.7925
Design Matrix 3	0 31.5286 -88.7767 -8.7051	0 31.2979 -71.8928 -9.8686	0 31.5023 -86.5238 -8.9002	0 31.5276 -88.7150 -8.7112	0 31.5282 -88.7712 -8.7064
Design Matrix 4	0 -14.2486 -14.2486 5.9836	0 -12.4487 -12.4487 5.6449	0 -10.8707 -10.8707 5.3486	0 -8.2765 -8.2765 4.8613	0 -6.2844 -6.2844 4.4870
Design Matrix 5	0 -151.2359 199.8510 -20.1993	0 -138.5480 153.9237 -15.4917	0 -149.1688 190.8510 -19.2119	0 -151.1533 199.4192 -20.1524	0 -151.2407 199.8561 -20.1961

Table 2: Variances of the OLS, GR, JR, J2R, J3R, J4R, J5R.

	OLS	GR	JR	J2R	J3R	J4R	J5R
Matrix 1	0 1.0959 0.0583 0.0175	0 1.06 0.0585 0.0174	0 1.0957 0.0583 0.0175	0 1.0942 0.0582 0.0174	0 1.0953 0.05872 0.0177	0 1.0953 0.0587 0.0177	0 1.0953 0.0587 0.0177
Matrix 2	0 0 9.0652 0.0189	0 0 6.9612 0.0182	0 0 8.7900 0.0189	0 0 9.0611 0.0189	0 0 9.0653 0.0189	0 0 9.0653 0.0189	0 0 9.0653 0.0189
Matrix 3	0 0 3.3638 0.0186	0 0 3.0319 0.0185	0 0 3.3466 0.0186	0 0 3.3638 0.0186	0 0 3.3638 0.0186	0 0 3.3638 0.0186	0 0 3.3638 0.0186
Matrix 4	0 0 9.0575 0.0189	0 0 6.9553 0.1876	0 0 8.7825 1.8804	0 0 9.0575 1.8804	0 0 9.0575 1.8804	0 0 9.0652 0.0189	0 0 9.0652 0.0189
Matrix 5	0 5.9621 0.4823 0.0044	0 3.7134 0.4196 0.00436	0 5.4438 0.4814 0.0044	0 5.9386 0.4823 0.0044	0 5.9611 0.4823 0.0044	0 5.9621 0.4823 0.0044	0 5.9621 0.4823 0.0044

Table 3 : Biases for OLS, GR, JR, J2R, J3R, J4R, J5R……

	Eigenvalues	GR	JR	J2R	J3R	J4R	J5R
Matrix 1	λ_2	0 0.0092 0.0015 0.0357	0 0.102×10^{-3} 0.016×10^{-3} 0.397×10^{-3}	0 1.263×10^{-7} 0.208×10^{-7} 4.943×10^{-7}	0 0.158×10^{-11} 0.025×10^{-11} 0.614×10^{-11}	0 0.197×10^{-15} 0.032×10^{-15} 0.763×10^{-15}	0 0.245×10^{-19} 0.040×10^{-19} 0.949×10^{-19}
Matrix 2	λ_2	0 0 -20.762 3.2813	0 0 -20.7626 3.2813	0 0 -20.7626 3.2813	0 0 -20.7626 3.2813	0 0 -20.7626 3.2813	0 0 -20.7626 3.2813
Matrix 3	λ_2	0 0 26.2634 3.2989	0 0 26.2634 3.2989	0 0 26.2634 3.2989	0 0 26.2634 3.2989	0 0 26.2634 3.2989	0 0 26.2634 3.2989
Matrix 4	λ_2	0 0 0.2125 102.387	0 0 20.7626 3.2813	0 0 20.7626 3.2813	0 0 20.7626 3.2813	0 0 20.7626 3.2813	0 0 20.7626 3.2813
Matrix 5	λ_2	0 47.9291 -22.980 0.6960	0 10.0611 -4.8442 0.1469	0 0.4469 -0.2125 0.0065	0 0.0198 -0.0096 0.0003	0 0.0008815 -0.000424 0.0000129	0 0.0003915 0.0001889 0.0000057