

Technical Report No. 1/12, February 2012
JACKKNIFING THE RIDGE REGRESSION ESTIMATOR: A
REVISIT

Mansi Khurana, Yogendra P. Chaubey and Shalini Chandra

Jackknifing the Ridge Regression Estimator: A Revisit

MANSI KHURANA*

Banasthali University, Rajasthan, India

YOGENDRA P. CHAUBEY

Concordia University, Montréal, Canada

SHALINI CHANDRA

Indian Statistical Institute, North East Centre, Assam, India

Abstract

Singh *et al.* (1986) proposed an almost unbiased ridge estimator using Jackknife method that required transformation of the regression parameters. This article shows that the same method can be used to derive the Jackknifed ridge estimator of the original (untransformed) parameter without transformation. This method also leads in deriving easily the second order Jackknifed ridge that may reduce the bias further. We further investigate the performance of these estimators along with a recent method by Batah *et al.* (2008) called modified Jackknifed ridge theoretically as well as numerically.

Key Words: Multicollinearity, Ridge regression, Jackknife technique.

1 Introduction

In regression analysis, the presence of multicollinearity among independent variables is a common problem which exhibits serious undesirable effects on the analysis. Its one of the major consequence on ordinary least squares (OLS) is that the estimator produces huge sampling variances which may result in the exclusion of significant coefficients from the model (see Farrar and Glauber 1967). To deal with such instability, number of methods were developed, the most popular amongst them being the ridge regression (RR) estimator proposed by Hoerl and Kennard (1970) that has been the benchmark for almost all the estimators developed later in this context. Ridge estimation procedure was based on adding small positive quantities to the diagonal of $X'X$ which made RR estimator biased but ensured a

*This work was completed while the author was visiting Concordia University as part of Common Wealth visiting scholar programme.

smaller mean square error (MSE) than the OLS. Swindel (1976) illustrated a technique for combining the prior information with ridge regression. Sarkar (1992) proposed a restricted ridge regression estimator based on RR estimator and restricted least squares (RLS) estimator and derived the superiority of this estimator over the other two by the criteria of MSE when the restrictions are correct and also the conditions when they are not correct. A large number of research articles have appeared since then in this context, see Farebrother (1976), Firinguetti (1989), Crouse *et al.* (1995) and other references there in. These ridge based estimators received considerable attention due to their computational feasibility and some optimal properties but they may carry serious bias.

Singh *et al.* (1986) noted that the general ridge estimators may carry a substantial amount of bias which had been ignored in the literature. To circumvent this problem, they proposed an almost unbiased ridge estimator using the Jackknife procedure of bias reduction. They demonstrated that the Jackknifed estimator had smaller bias as well as MSE than the generalized ridge estimator (GRE) under some conditions. Thereafter, a number of papers have proposed and studied Jackknifed versions of ridge estimator and some other variants. Gruber (1991) compared the efficiency of Jackknifed and usual ridge-type estimators, Batah *et al.* (2008) studied the effect of Jackknifing on various ridge-type estimators. Interesting literature on the general Jackknife procedures may be found in Miller (1974) and Hinkley (1977).

In all these papers, Jackknife is performed on a linear set of parameters (to be explained later), using a transformed set of regressors so that the matrix of normal equations becomes diagonal. This introduces simplicity in computing the Jackknife, however, adds the complication of re-transformation to get the estimators of the original regression parameter vector. In the next section, we show that such a transformation is not really necessary and derive the Jackknifed RR estimator of the original regression parameter which is basically of the same form as obtained in Singh *et al.* (1986). This direct strategy of jackknifing also enables one to carry out the Jackknife procedure for other estimators of interest in this context.

Section 3 considers other estimators such as those studied by Batah *et al.* (2008) and the second-order Jackknifed RR estimator, that may have relevance for further reduction in bias. This section also presents a theoretical comparison of bias and MSE of different estimators. A comparative study of these estimators based on simulations is carried out in Section 4.

2 Jackknifed Ridge Estimator for the Multiple Linear Regression

Consider the following multiple linear regression model

$$y = X\beta + u \quad (2.1)$$

where y is an $(n \times 1)$ vector of observations on the variable to be explained, X is an $(n \times p)$ matrix of n observations on p explanatory variables assumed to be of full column rank, β is a $(p \times 1)$ vector of regression coefficients associated with them and u is an $(n \times 1)$ vector of disturbances, the elements of which are assumed to be independently and identically normally distributed with $E(u) = 0$ and $E(uu') = \sigma^2 I$.

Since the exposition of Gauss-Markov theorem, and especially during the past fifty years or so, practically all fields in the natural and social sciences have widely or sometimes blindly used the classical method of estimation, in particular, the OLS estimation method. In least squares approach to linear regression model, the unknown parameters are estimated by minimizing the sum of squared deviations between the data and the model. Applying the least squares method to (2.1) yields the OLS estimator of β which is given by

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

that is well known to be the *best linear unbiased*.

As referred to earlier, the OLS procedure may result in large sampling variances in the presence of multicollinearity and therefore may produce estimators which are not in tune with the researcher's prior belief. To combat this problem, Hoerl and Kennard (1970) proposed the generalized ridge estimator (GRE) by allowing some bias into the estimator which resulted in smaller MSE. The estimator is given by

$$\hat{\beta}_{GRE} = (X'X + D)^{-1}X'y \quad (2.3)$$

where D is a known matrix. Note that D may not necessarily be diagonal, though, in their original paper Hoerl and Kennard (1970) considered it to be so. This makes sense in the context of a transformed model where the matrix $X'X$ is diagonal (see later development). The GRE defined in (2.3) reduces to OLS for D being a $p \times p$ zero matrix. It was demonstrated that there always exists a $d_i > 0$ such that $\hat{\beta}_{GRE}$ performs better than $\hat{\beta}$ in terms of MSE. It may also be noted that ordinary ridge estimator (ORE) is a special case of (2.3) when all $d_i \equiv d, \forall i$.

To alleviate the problem of bias in GRE, Singh *et al.* (1986) proposed an almost unbiased ridge estimator using the Jackknife technique that was introduced by Quenouille (1956) as

a general method for reducing the bias of an estimator. Later Tukey (1958) proposed that this technique may also offer a simple method to obtain the confidence intervals for the parameters of interest. A number of papers appeared dealing with large-sample properties and empirical validations in common applications like estimation of variances, correlations and ratios. See also the paper by Thorburn (1976) for an excellent exposition of the asymptotic properties of the Jackknife estimator. Jackknife procedure was initially applied to balanced situations i.e., equal sample sizes, equal variances, etc. A commendable review of the same may be found in Miller (1974a). Miller (1974b) gave the detailed account of jackknifing linear model estimates, and demonstrated that the Jackknife produces consistent results when the sample is large. Further, Hinkley (1977) examined the small sample properties of the standard Jackknife in the general linear model, and compared it to an alternative weighted Jackknife procedure.

The use of Jackknife procedure to reduce the bias of the ridge estimator and the properties of the Jackknifed ridge estimator were also studied by Singh and Chaubey (1987), Nomura (1988) and Gruber (1998). Singh *et al.* (1986), start with model (2.1) and for convenience decompose it as

$$y = Z\gamma + u \quad (2.4)$$

where, $Z = XG$ and $\gamma = G'\beta$. Here G is a $(p \times p)$ matrix whose columns are normalized eigen vectors of $X'X$. Also, $Z'Z = G'X'XG = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, λ_i being the i -th eigenvalue of $X'X$. The GRE of γ may be written as

$$\hat{\gamma}_{GRE} = (\Lambda + K)^{-1}Z'y = A^{-1}Z'y = A^{-1}\Lambda\hat{\gamma} = (I - A^{-1}K)\hat{\gamma} \quad (2.5)$$

where $K = \text{diag}(k_1, \dots, k_p)$, $k_i > 0$, $A = \Lambda + K$ and $\hat{\gamma}$ is the OLS estimator of γ . Since $\gamma = G'\beta$ and $GG' = I$, the GRE of β is

$$\hat{\beta}_{GRE} = G\hat{\gamma}_{GRE} = A_*^{-1}X'y, \quad (2.6)$$

where $A_* = X'X + K_*$ and $K_* = GKG'$.

Following Hinkley (1977), Singh *et al.* (1986) derived the Jackknifed form of $\hat{\gamma}_{GRE}$ as

$$\hat{\gamma}_{JRE} = [I - (A^{-1}K)^2]\hat{\gamma}. \quad (2.7)$$

It was shown that the bias of $\hat{\gamma}_{JRE}$ is smaller than that of $\hat{\gamma}_{GRE}$.

Here, it is important to note that Jackknife procedure is applied on a transformed set of regressors for inducing simplicity of computation which however leads complication of re-transformation to get the estimator of original regression parameters. We have shown that it is easy to get the estimator for original regression parameter explicitly. We know that

$$\hat{\beta}_{JRE} = G\hat{\gamma}_{JRE} \quad (2.8)$$

Using (2.7) in (2.8), we get

$$\begin{aligned}
\hat{\beta}_{JRE} &= G[I + (A^{-1}K)][I - (A^{-1}K)]\hat{\gamma} \\
&= G[I + (A^{-1}K)](A^{-1}\Lambda)\Lambda^{-1}Z'y \\
&= G[I + (A^{-1}K)]A^{-1}G'X'y \\
&= [GA^{-1}G' + GA^{-1}KA^{-1}G']X'y
\end{aligned} \tag{2.9}$$

Now writing $A = Z'Z + K = G'[X'X + GK'G]G$, so that, $GA^{-1}G' = (X'X + K_*)^{-1} = A_*^{-1}$ (say), and $GA^{-1}KA^{-1}G' = A_*^{-1}K_*A_*^{-1}$, where $K_* = GK'G'$.

Thus using the above values of $GA^{-1}G'$ and $GA^{-1}KA^{-1}G'$ in (2.9), we get

$$\begin{aligned}
\hat{\beta}_{JRE} &= [A_*^{-1} + A_*^{-1}K_*A_*^{-1}]X'y \\
&= [I + A_*^{-1}K_*]A_*^{-1}X'y \\
&= [I + A_*^{-1}K_*]A_*^{-1}(X'X)\hat{\beta} \\
&= [I + A_*^{-1}K_*][I - A_*^{-1}K_*]\hat{\beta} \\
&= [I - (A_*^{-1}K_*)^2]\hat{\beta}
\end{aligned} \tag{2.10}$$

where A_* and K_* are as described earlier. We see that $\hat{\beta}_{JRE}$ has exactly the same form as $\hat{\gamma}_{JRE}$ with $A = A_*$ and $K = K_*$.

We note that this form of $\hat{\beta}_{JRE}$ still requires knowledge of G . However, as shown below direct jackknifing $\hat{\beta}_{GRE}$ gives the same form of $\hat{\beta}_{JRE}$ and it does not require the knowledge of G .

Let us write $\hat{\beta}_{GRE}$, defined in (2.3) as

$$\hat{\beta}_{GRE} = (X'X + D)^{-1}X'y = B^{-1}X'y$$

For this, we first obtain $\hat{\beta}_{GRE(-i)}$, that is the GRE obtained from the data after deleting the i -th observation as given by,

$$\begin{aligned}
\hat{\beta}_{GRE(-i)} &= (X'X - x_i x_i' + D)^{-1}(X'y - x_i y_i) \\
&= (B - x_i x_i')^{-1}(X'y - x_i y_i)
\end{aligned}$$

which may be written as

$$\hat{\beta}_{GRE(-i)} = \left[B^{-1} + \frac{B^{-1}x_i x_i' B^{-1}}{1 - w_i} \right] (X'y - x_i y_i)$$

where

$$w_i = x_i' B^{-1} x_i.$$

This on simplification, gives

$$\hat{\beta}_{GRE(-i)} = \hat{\beta}_{GRE} - \frac{B^{-1}x_i}{1-w_i}(y_i - x_i'\hat{\beta}_{GRE})$$

Now defining pseudo-values as

$$\begin{aligned} Q_i &= \hat{\beta}_{GRE} + n(1-w_i)(\hat{\beta}_{GRE} - \hat{\beta}_{GRE(-i)}) \\ &= \hat{\beta}_{GRE} + nB^{-1}x_ie_i \end{aligned}$$

where $e_i = y_i - x_i'\hat{\beta}_{GRE}$ are the residuals using the ridge estimator. Using these pseudo-values the Jackknifed ridge estimator of β is given by

$$\begin{aligned} \hat{\beta}_{JRE} &= \bar{Q} = \frac{1}{n} \sum Q_i = \hat{\beta}_{GRE} + B^{-1}X'e \\ &= \hat{\beta}_{GRE} + B^{-1}X'(y - X\hat{\beta}_{GRE}) \\ &= \hat{\beta}_{GRE} + \hat{\beta}_{GRE} - B^{-1}X'X\hat{\beta}_{GRE} \\ &= \hat{\beta}_{GRE} + [I - B^{-1}X'X]\hat{\beta}_{GRE} \\ &= \hat{\beta}_{GRE} + (B^{-1}D)\hat{\beta}_{GRE} \\ &= [I + B^{-1}D]\hat{\beta}_{GRE} \end{aligned} \tag{2.11}$$

Also, $\hat{\beta}_{GRE} = B^{-1}(X'X)\hat{\beta} = [I - B^{-1}D]\hat{\beta}$. Using this in (2.11) gives us the Jackknife estimator $\hat{\beta}_{JRE}$ as

$$\hat{\beta}_{JRE} = [I - (B^{-1}D)^2]\hat{\beta} \tag{2.12}$$

We note that the JRE as obtained in (2.12) is of the same form as $\hat{\gamma}_{JRE}$ and this form makes it easier to study the properties of Jackknifed form of β . In the next section, we are going to introduce second-order Jackknife estimator and compare its performance with JRE and MJR (Batah *et al.* 2008). Since MJR was introduced for the transformed parameter γ , therefore for the remaining sections we will make use of the transformed model as defined in (2.4).

3 Modified Jackknifed Ridge and Second-Order Jackknifed Ridge

Batah *et al.* (2008) proposed a new estimator namely, Modified Jackknife ridge regression estimator (MJR) combining the ideas of GRE and JRE given by Singh *et al.* (1986). They obtained conditions for superiority of MJR over JRE and showed through a simulation study that the MJR estimator is generally superior to both GRE and JRE using MSE criterion. The MJR estimator is of the following form

$$\hat{\gamma}_{MJR} = [I - (A^{-1}K)^2][I - A^{-1}K]\hat{\gamma} = [I - (A^{-1}\Phi K)]\hat{\gamma} \tag{3.1}$$

where $\Phi = (I + A^{-1}K - A^{-1}K^*)$ and $K^* = KA^{-1}K$. This method is a bit adhoc because it just consists of replacing the OLS in (2.7) by ridge estimator. They have demonstrated that this estimator outperforms the JRE under some conditions which, however, can not be verified in practice. Further in this paper there is no discussion of bias of the resulting estimator. If further bias reduction of the ridge estimator was required, we could Jackknife $\hat{\beta}_{JRE}$ again. Noting that we can write $\hat{\gamma}_{JRE}$ as the ridge estimator in the form

$$\hat{\gamma}_{JRE} = [I - A^{-1}K^*]\hat{\gamma} \quad (3.2)$$

that is similar to the ridge estimator as given in (2.5) with K replaced by $K^* = KA^{-1}K$, hence its Jackknifed form that results in the second-order Jackknife (J2R) estimator can be readily written as

$$\hat{\gamma}_{J2R} = [I - (A^{-1}K^*)^2]\hat{\gamma}. \quad (3.3)$$

Similarly, the second-order Jackknifed ridge estimator for β is given by

$$\hat{\beta}_{J2R} = [I - (B^{-1}D^*)^2]\hat{\beta}, \quad (3.4)$$

where $D^* = DB^{-1}D$.

In the following subsections, we compare the performance of J2R with JRE and MJR in terms of bias and MSE. It is expected that J2R will reduce the bias of JRE and could provide a valid estimator of its variance-covariance matrix, using the general Jackknife methodology as indicated in Singh *et al.* (1986) for the Jackknifed ridge estimator. However, there is no guarantee that J2R will also be efficient in terms of MSE. This is the line of investigation in the remaining subsection.

3.1 Comparison between the bias of JRE, MJR and J2R

The following theorem compares the total squared bias of JRE and J2R, similar to that between JRE and GRE compared by Singh *et al.* (1986).

Theorem 1. Let K be a $(p \times p)$ diagonal matrix with non-negative entries. Then the second-order Jackknifed estimator $\hat{\gamma}_{J2R(i)}$ reduces the bias of the Jackknifed estimator $\hat{\gamma}_{JRE(i)}$, assuming $k_i > 0$. Further, the difference of total squared biases of the Jackknifed ridge (JRE) and second-order Jackknifed ridge (J2R) estimators of β as given by

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

is non-negative. It is strictly positive if at least one $k_i, i = 1, \dots, n$ is positive.

Proof: Recall from (2.5), we have

$$\hat{\gamma}_{JRE} = [I - (A^{-1}K)^2]\hat{\gamma} = [I - A^{-1}K^*]\hat{\gamma}$$

where $K^* = KA^{-1}K$, thus

$$Bias(\hat{\gamma}_{JRE}) = E(\hat{\gamma}_{JRE}) - \gamma = -A^{-1}K^*\gamma \quad (3.5)$$

Similarly using the expression for $\hat{\gamma}_{J2R}$ from (3.3)

$$\hat{\gamma}_{J2R} = [I - (A^{-1}K^*)^2]\hat{\gamma}$$

we find

$$Bias(\hat{\gamma}_{J2R}) = E(\hat{\gamma}_{J2R}) - \gamma = -(A^{-1}K^*)^2\gamma. \quad (3.6)$$

Comparing $|Bias(\hat{\gamma}_{JRE})|_i$ with $|Bias(\hat{\gamma}_{J2R})|_i$ where $|\cdot|_i$ denotes the absolute value of the i -th component, we have

$$\begin{aligned} |Bias(\hat{\gamma}_{JRE})|_i - |Bias(\hat{\gamma}_{J2R})|_i &= \frac{k_i^*}{\lambda_i + k_i}|\gamma_i| - \frac{(k_i^*)^2}{(\lambda_i + k_i)^2}|\gamma_i| \\ &= \frac{(\lambda_i k_i)^2 + 2\lambda_i k_i^3}{(\lambda_i + k_i)^4}|\gamma_i| \end{aligned} \quad (3.7)$$

that is clearly positive assuming $k_i > 0$, which proves that the second-order Jackknife reduces the bias of the Jackknifed ridge estimator.

Also, if we look at the expression of total squared bias in terms of original parameter vector β , we find that

$$\begin{aligned} D_1 &= \sum \{|Bias(\hat{\beta}_{JRE})|_i^2 - |Bias(\hat{\beta}_{J2R})|_i^2\} \\ &= \beta' \{G[(A^{-1}K^*)^2 - (A^{-1}K^*)^4]G'\} \beta \end{aligned} \quad (3.8)$$

where use has been made of the fact that $Bias(\hat{\beta}) = GBias(\hat{\gamma})$. Clearly, the matrix in braces in R.H.S of (3.8) is positive definite due to the non-negativity of the expression in (3.7), hence $D_1 > 0$. This proves the theorem. \square

Next we compare the total squared bias of MJR with GRE of the original parameter vector β .

Theorem 2. Let K be a $(p \times p)$ diagonal matrix with non-negative entries, then the difference of total squared biases of the modified Jackknifed ridge (MJR) and generalized ridge estimators (GRE) of β as given by

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

is positive.

Proof: Using the expression for MJR as given in (3.1)

$$\hat{\gamma}_{MJR} = [I - (A^{-1}K)^2][I - A^{-1}K]\hat{\gamma} = [I - (A^{-1}\Phi K)]$$

where $\Phi = (I + A^{-1}K - A^{-1}K^*)$ and $K^* = KA^{-1}K$, we have

$$Bias(\hat{\gamma}_{MJR}) = -A^{-1}\Phi K\gamma. \quad (3.9)$$

Also, using the expression for GRE of γ from (2.5), we have

$$Bias(\hat{\gamma}_{GRE}) = -A^{-1}K\gamma. \quad (3.10)$$

Comparing the expression for bias of MJR as given in (3.9) with that in (3.10) component wise, we have

$$\begin{aligned} |Bias(\hat{\gamma}_{MJR})|_i - |Bias(\hat{\gamma}_{GRE})|_i &= \frac{k_i\phi_i}{\lambda_i + k_i}|\gamma_i| - \frac{k_i}{\lambda_i + k_i}|\gamma_i| \\ &= \frac{k_i[1 + \frac{k_i}{\lambda_i + k_i} - \frac{k_i^2}{(\lambda_i + k_i)^2}]}{\lambda_i + k_i}|\gamma_i| - \frac{k_i}{\lambda_i + k_i}|\gamma_i| \\ &= \frac{\lambda_i k_i^2}{(\lambda_i + k_i)^3}|\gamma_i| \end{aligned}$$

which is a positive quantity.

Now comparing the expressions for the total bias as in Theorem 1, we have

$$\begin{aligned} D_2 &= \sum \{ |Bias(\hat{\beta}_{MJR})|_i^2 - |Bias(\hat{\beta}_{GRE})|_i^2 \} \\ &= \beta' \{ G[(A^{-1}\Phi K)^2 - (A^{-1}K)^2]G' \} \beta \\ &= \beta' \{ G[(A^{-1}\Phi KA^{-1}\Phi K - A^{-1}KA^{-1}K)]G' \} \beta \\ &= \beta' \{ GA^{-1}[(\Phi K)^2 - K^2]A^{-1}G' \} \beta. \end{aligned} \quad (3.11)$$

It is easy to see that the matrix in braces in the above equation is positive definite that proves the result. \square

Remark 1. Using the result in Singh *et al.* (1986) that proves that JRE dominates the GRE in terms of total squared bias, and combining the results of theorems 1 and 2, we have the following ordering,

$$SB(\hat{\beta}_{J2R}) < SB(\hat{\beta}_{JRE}) < SB(\hat{\beta}_{GRE}) < SB(\hat{\beta}_{MJR}), \quad (3.12)$$

where SB denotes, the total squared bias. This automatically shows that J2R dominates MJR in terms of total squared bias.

Remark 2. It was shown using simulation studies by Batah *et al.* (2008) that MJR has better performance as compared to JRE in terms of MSE, however no bias comparisons were made. The following section provides some theoretical comparisons of MSE of JRE, MJR and J2R on the similar lines as in Batah *et al.* (2008).

3.2 Comparison between MSE's of JRE, MJR and J2R

Batah *et al.* (2008) stated conditions under which MJR dominates the JRE in terms of MSE matrix. The following theorem provides a similar comparison between $\hat{\gamma}_{J2R}$ and $\hat{\gamma}_{JRE}$. First we state a lemma that has been used by many researchers working in this context.

Lemma 1. [Farebrother (1976)] Let A be a positive definite matrix, γ be a $p \times 1$ vector. Then $A - \gamma\gamma'$ is a nonnegative definite matrix if and only if $\gamma'A^{-1}\gamma \leq 1$ is satisfied.

Theorem 3. Let K be a $(p \times p)$ diagonal matrix with non-negative entries. Then the difference of the MSE matrix of the second-order Jackknife estimator and the Jackknifed ridge estimator,

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

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

$$\gamma'\{L^{-1}[\sigma^2 M + (A^{-1}K^*)^2 \gamma\gamma'(K^*A^{-1})^2]L^{-1}\}^{-1}\gamma \leq 1$$

where $L = A^{-1}K^*$, $K^* = KA^{-1}K$ and $M = [I - A^{-1}K^*]\Lambda^{-1}[I - A^{-1}K^*]'\{[I + A^{-1}K^* + I]A^{-1}K^*\}$.

Proof: Using the expression of the variance-covariance matrix of the least squares estimator, i.e. $Var(\hat{\gamma}) = \sigma^2(Z'Z)^{-1} = \sigma^2\Lambda^{-1}$, and the expression for JRE as given in (3.2), we can write

$$\begin{aligned} Var(\hat{\gamma}_{JRE}) &= E[(\hat{\gamma}_{JRE} - E(\hat{\gamma}_{JRE}))(\hat{\gamma}_{JRE} - E(\hat{\gamma}_{JRE}))'] \\ &= [I - A^{-1}K^*]Var(\hat{\gamma})[I - A^{-1}K^*]' \\ &= [I - A^{-1}K^*]\sigma^2\Lambda^{-1}[I - A^{-1}K^*]'. \end{aligned}$$

Further using the expression for $Bias(\hat{\gamma}_{JRE})$ from (3.5), we can write

$$MSE(\hat{\gamma}_{JRE}) = [I - A^{-1}K^*]\sigma^2\Lambda^{-1}[I - A^{-1}K^*]' + A^{-1}K^*\gamma\gamma'K^*A^{-1}. \quad (3.13)$$

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

$$Var(\hat{\gamma}_{J2R}) = [I - (A^{-1}K^*)^2]\sigma^2\Lambda^{-1}[I - (A^{-1}K^*)^2]'$$

and the expression for $Bias(\hat{\gamma}_{J2R})$ as given in (3.6) we have

$$MSE(\hat{\gamma}_{J2R}) = [I - (A^{-1}K^*)^2]\sigma^2\Lambda^{-1}[I - (A^{-1}K^*)^2]' + (A^{-1}K^*)^2\gamma\gamma'(K^*A^{-1})^2. \quad (3.14)$$

And therefore from (3.13) and (3.14), Δ_1 becomes

$$\Delta_1 = \sigma^2M + (A^{-1}K^*)^2\gamma\gamma'(K^*A^{-1})^2 - A^{-1}K^*\gamma\gamma'K^*A^{-1} \quad (3.15)$$

where

$$\begin{aligned} M &= [I - (A^{-1}K^*)^2]\Lambda^{-1}[I - (A^{-1}K^*)^2]' - [I - A^{-1}K^*]\Lambda^{-1}[I - A^{-1}K^*]' \\ &= [I - A^{-1}K^*]\Lambda^{-1}[I - A^{-1}K^*]'\{[I + A^{-1}K^*][I + A^{-1}K^*]' - I\} \\ &= [I - A^{-1}K^*]\Lambda^{-1}[I - A^{-1}K^*]'\{[I + A^{-1}K^* + I][I + A^{-1}K^* - I]\} \\ &= [I - A^{-1}K^*]\Lambda^{-1}[I - A^{-1}K^*]'\{[I + A^{-1}K^* + I]A^{-1}K^*\} \end{aligned} \quad (3.16)$$

We know that $[I - A^{-1}K^*]$ is a diagonal matrix with positive entries and the matrix in braces of (3.16) is also positive definite. Therefore we conclude that M is also a positive definite matrix and hence the difference Δ_1 is positive definite if and only if $L^{-1}\Delta_1L^{-1}$ is positive definite. We see that

$$L^{-1}\Delta_1L^{-1} = L^{-1}[\sigma^2M + (A^{-1}K^*)^2\gamma\gamma'(K^*A^{-1})^2]L^{-1} - \gamma\gamma' \quad (3.17)$$

The matrix $[\sigma^2M + (A^{-1}K^*)^2\gamma\gamma'(K^*A^{-1})^2]$ in the above equation is symmetric positive definite. Therefore using Lemma 1, we conclude that $L^{-1}\Delta_1L^{-1}$ is positive definite if and only if following inequality holds

$$\gamma'\{L^{-1}[\sigma^2M + (A^{-1}K^*)^2\gamma\gamma'(K^*A^{-1})^2]L^{-1}\}^{-1}\gamma \leq 1.$$

□

Remark 3. We see that $\Delta_1 = 0$ if and only if $L^{-1}\Delta_1L^{-1} = 0$, which gives

$$L^{-1}[\sigma^2 M + (A^{-1}K^*)^2 \gamma \gamma' (K^* A^{-1})^2] L^{-1} = \gamma \gamma'$$

which is not true because the rank of left hand matrix is p and the rank of right hand matrix is either 0 or 1. So Δ_1 can not be zero whenever $p > 1$.

Remark 4. Noting that the matrix $(A^{-1}K^*)^2 \gamma \gamma' (K^* A^{-1})^2$ in (3.15) is positive definite, a simple sufficient condition for Δ_1 to be positive definite using Lemma 1 is given by

$$\sigma^{-2} \gamma' L M^{-1} L \gamma \leq 1.$$

The following theorem compares the MSE matrix of J2R and MJR, similar to the analysis above.

Theorem 4. Let K be a $(p \times p)$ diagonal matrix with non-negative entries. Then the difference

$$\Delta_2 = MSE(\hat{\gamma}_{J2R}) - MSE(\hat{\gamma}_{MJR})$$

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

$$\gamma' \{N^{-1}[\sigma^2 H + (A^{-1}K^*)^2 \gamma \gamma' (K^* A^{-1})^2] N^{-1}\}^{-1} \gamma \leq 1$$

where $H = [I - A^{-1}K^*] \Lambda^{-1} [I - A^{-1}K^*]' \{[I + A^{-1}K^*][I + A^{-1}K^*]' - [I - A^{-1}K][I - A^{-1}K]'\}$, $N = A^{-1}\Phi K$, Φ and K^* are as defined earlier.

Proof: Writing $\hat{\gamma}_{MJR}$ as

$$\hat{\gamma}_{MJR} = [I - (A^{-1}\Phi K)] \hat{\gamma}$$

where $\Phi = (I + A^{-1}K - A^{-1}K^*)$ and $K^* = KA^{-1}K$, we have

$$Var(\hat{\gamma}_{MJR}) = [I - (A^{-1}\Phi K)] \sigma^2 \Lambda^{-1} [I - (A^{-1}\Phi K)]'$$

and

$$Bias(\hat{\gamma}_{MJR}) = -A^{-1}\Phi K \gamma$$

Thus

$$MSE(\hat{\gamma}_{MJR}) = [I - (A^{-1}\Phi K)] \sigma^2 \Lambda^{-1} [I - (A^{-1}\Phi K)]' + (A^{-1}\Phi K) \gamma \gamma' (K \Phi A^{-1}) \quad (3.18)$$

Now using (3.14) and (3.18), we get

$$\Delta_2 = \sigma^2 H + (A^{-1}K^*)^2 \gamma \gamma' (K^* A^{-1})^2 - (A^{-1}\Phi K) \gamma \gamma' (K \Phi A^{-1}) \quad (3.19)$$

where

$$\begin{aligned}
H &= [I - (A^{-1}K^*)^2]\Lambda^{-1}[I - (A^{-1}K^*)^2]' - [I - A^{-1}\Phi K]\Lambda^{-1}[I - A^{-1}\Phi K]' \\
&= [I - (A^{-1}K^*)^2]\Lambda^{-1}[I - (A^{-1}K^*)^2]' \\
&\quad - [I - A^{-1}K^*][I - A^{-1}K]\Lambda^{-1}[I - A^{-1}K]'[I - A^{-1}K^*]' \\
&= [I - A^{-1}K^*]\Lambda^{-1}[I - A^{-1}K^*]'\{[I + A^{-1}K^*][I + A^{-1}K^*]' \\
&\quad - [I - A^{-1}K][I - A^{-1}K]'\} \tag{3.20}
\end{aligned}$$

We see that the matrix in braces of (3.20) is a positive definite and that $[I - A^{-1}K^*]$ is also positive. This implies that H is also a positive definite matrix. Now, proceeding as in the proof of Theorem 3, we get the desired result. \square

Remark 5. An analysis similar to that in Remark 4, noting that the matrix $(A^{-1}K^*)^2\gamma\gamma'(K^*A^{-1})^2$ in (3.19) is positive definite, a simple sufficient condition for the matrix Δ_2 to be positive definite is given by

$$\sigma^{-2}\gamma'NH^{-1}N\gamma \leq 1.$$

The conditions in Theorems 3 and 4 as well as those in Remarks 3 and 4 are not verifiable in practice, hence it is important to investigate the relative performance of the new estimator proposed here under a variety of conditions. In the following section we provide a simulation study for this purpose that is similar to the one undertaken in Batah *et al.* (2008).

4 A Simulation Study

In the present section, our aim is to compare the performance of OLS, ORE, JRE, MJR, and J2R estimators on the basis of associated bias and mean square error with the help of Monte Carlo experiments. The true model is $y = X\beta + u$ where $u \sim N(0, 1)$. Here β is taken as the normalized eigen vector corresponding to the largest eigen value of $X'X$. The explanatory variables are generated from the following equation

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}}w_{ij} + \rho w_{ip}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p. \tag{4.1}$$

where w_{ij} are independent standard normal pseudo-random numbers and ρ^2 is the correlation between x_{ij} and $x_{ij'}$ for $j, j' < p$ and $j \neq j'$. When j or $j' = p$, the correlation will be ρ . We have taken $\rho = 0.8, 0.95$ and 0.99 to investigate the effects of different degrees of collinearity with sample sizes $n = 15, 50$ and 100 . The feasible value of k is obtained by the optimal formula $k = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$ as given by Hoerl *et al.* (1975), so that

$$\hat{k} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

where

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n - p}.$$

This simulation study is patterned on that of McDonald and Galarnaeu (1975). For computations, the original model is first decomposed into a canonical form to get the estimator of γ and then transformed back to the estimator of β . For these different choices of ρ and n , the experiment is replicated 5000 times. The average absolute bias and average mean square error are computed as follows:

$$Bias(\hat{\beta}_i) = \frac{1}{5000} \sum |\hat{\beta}_{ij} - \beta_i| \quad (4.2)$$

$$MSE(\hat{\beta}_i) = \frac{1}{5000} \sum (\hat{\beta}_{ij} - \beta_i)^2 \quad (4.3)$$

Here, $\hat{\beta}_{ij}$ denotes the estimate of i -th parameter in j -th replication and β_1, β_2 and β_3 are the true parameter values. Results of the simulation study are given in Table 1 and 2. Our conclusions based on this study are summarized below:

1. From table 1, we note that bias of OLS is almost close to zero and bias of J2R is not very far from OLS as expected. This shows that if we further increase the order of Jackknife estimator, its bias will tend towards the bias of OLS.
2. JRE has also served its purpose by reducing the bias of ORE in all the cases.
3. We also notice that the bias of MJR is higher among all the estimators in almost all the cases.
4. From table 2, we see that mean square error of all the estimators is increasing with the increasing value of ρ and decreasing with the increasing value of n .
5. As expected, variance of OLS is highest amongst all and here also mean square error of J2R is closer to that of OLS. We find that J2R does not improve upon JRE and MJR in terms of MSE which we have also shown in theorems 3 and 4.

5 Other Concluding Remarks

In the present study, we derived the Jackknifed form of generalized ridge estimator $\hat{\beta}_{GRE}$ that avoid using any transformation, as had been the practice in the literature so far. We also proposed a new estimator based on the second-order Jackknife (J2R) which ensures a

smaller bias than those of the Jackknifed ridge estimator (JRE) and modified Jackknifed ridge estimator (MJR). It is seen that J2R may not be as efficient as JRE or MJR, however MJR may carry a much larger bias as became evident through the simulation studies. It may be useful to combine these estimators along the lines of Singh and Chaubey (1987) to improve upon MJR, JRE and J2R in terms of both bias and mean squared error. This will be investigated into another paper.

Acknowledgments

Y.P. Chaubey would like to acknowledge Natural Sciences and Engineering Research Council of Canada for the partial support of this research through a Discovery Research Grant. Mansi Khurana would like to thank Banasthali University in granting her leave to facilitate the access of Common Wealth visiting scholarship at Concordia University.

Table 1: Comparison of Bias of Different Estimators

n	ρ	OLS	ORE	JRE	MJR	J2R
15	0.8	0.000796607	0.019932830	0.0023494510	0.021127870	0.000949268
		0.003872392	0.019707720	0.0022376270	0.021139810	0.003831971
		0.005912432	0.037625430	0.0071932280	0.039017160	0.005819491
	0.95	0.001530710	0.017651900	0.0042949410	0.018535000	0.003007977
		0.007440947	0.011722090	0.0044564130	0.013290280	0.006600416
		0.009056449	0.023449020	0.0064332980	0.023187410	0.007344464
	0.99	0.003388197	0.001115321	0.0089504600	0.001018375	0.008610536
		0.016470394	0.012823031	0.0092888780	0.010947171	0.012506116
		0.016593310	0.004931625	0.0049867870	0.007036277	0.008663070
50	0.8	0.002104420	0.009019226	0.0021982780	0.009114149	0.002103926
		0.002389848	0.009439763	0.0024850460	0.009536641	0.002389148
		0.001263187	0.010585202	0.0010410690	0.010799617	0.001262333
	0.95	0.004043723	0.009288766	0.0039053630	0.009184668	0.004047090
		0.004592184	0.010052440	0.0045481220	0.010019484	0.004596323
		0.005052215	0.004450479	0.0046752720	0.004789447	0.005057859
	0.99	0.008950703	0.009257118	0.0075440090	0.008329471	0.008886674
		0.010164711	0.012574605	0.0107358930	0.012275974	0.010921653
		0.015116081	0.006059302	0.0142715980	0.004874950	0.015720642
100	0.8	0.001884865	0.005252948	0.0018929420	0.005262753	0.001884693
		0.002053578	0.005243242	0.0020486800	0.005241511	0.002053270
		0.001758128	0.004140953	0.0016827520	0.004212672	0.001757826
	0.95	0.003621840	0.005855833	0.0033506180	0.005664673	0.003601681
		0.003946028	0.005736671	0.0035095580	0.005429302	0.003913498
		0.005149614	0.000981643	0.0045466250	0.001424118	0.005109286
	0.99	0.008016873	0.006672324	0.0063549840	0.005749086	0.007724603
		0.008734457	0.005524346	0.0059544290	0.003997305	0.008243769
		0.014052604	0.003385189	0.0101242140	0.001202657	0.013366352

Table 2: Comparison of MSE of Different Estimators

n	ρ	OLS	ORE	JRE	MJR	J2R
15	0.8	0.19795727	0.10500900	0.16402858	0.09073415	0.19283316
		0.10953435	0.07250247	0.09972949	0.06729506	0.10869350
		0.09816408	0.06161525	0.08509099	0.05604219	0.09628443
	0.95	0.73091920	0.26792930	0.48280620	0.20443050	0.65104220
		0.40443450	0.19131410	0.31218970	0.15965990	0.38375670
		0.52999470	0.19938030	0.35234710	0.15404340	0.47237980
	0.99	3.58113700	1.10385200	2.06807900	0.80832700	2.93104800
		1.98152600	0.77428360	1.33932100	0.61421620	1.75621100
		3.51268700	1.04339250	1.97785700	0.75332770	2.83721400
50	0.8	0.04483244	0.03454313	0.04325834	0.03342195	0.04479028
		0.03966295	0.03128027	0.03846737	0.03041172	0.03963371
		0.03588494	0.02700028	0.03432280	0.02592576	0.03583691
	0.95	0.16553520	0.08715341	0.13748780	0.07508182	0.16151240
		0.14644780	0.08038150	0.12407930	0.07030963	0.14347200
		0.19049110	0.08648805	0.14790440	0.06989500	0.18342920
	0.99	0.81103920	0.29366640	0.53235850	0.22351450	0.71833940
		0.71752070	0.27043880	0.48445320	0.20843790	0.64413030
		1.23029320	0.36188230	0.71283780	0.25044860	1.03510770
100	0.8	0.02089901	0.01815047	0.02066509	0.01795757	0.02089680
		0.03012368	0.02488273	0.02955958	0.02443416	0.03011663
		0.02127735	0.01734774	0.02082090	0.01698969	0.02127124
	0.95	0.07716557	0.05076290	0.07070845	0.04713234	0.07667902
		0.11122591	0.06527134	0.09764992	0.05832660	0.10996109
		0.11620775	0.06169573	0.09843280	0.05304472	0.11440034
	0.99	0.37807250	0.16463220	0.28044560	0.13331800	0.35378040
		0.54495100	0.20685190	0.37151590	0.15896210	0.49365910
		0.74904350	0.23927770	0.46307450	0.16958470	0.65514070

References

- [1] Batah, F.S.M., Ramanathan, T.V. and Gore, S.D. (2008). The efficiency of modified Jack-knife and ridge type regression estimators: A comparison. *Surveys in Mathematics and its Applications*, **3** 111-122.
- [2] Crouse, R.H., Jin, C. and Hanumara, R.C. (1995). Unbiased ridge estimation with prior information and ridge trace. *Communications in Statistics–Theory and Methods*, **24(9)** 2341-2354.
- [3] Farrar, D.E. and Glauber, R.R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, **49(1)** 92-107.
- [4] Farebrother, R.W. (1976). Further results on the mean square error of ridge regression. *Journal of Royal Statistical Society*, **B38** 248-250.
- [5] Firinguetti, L. (1989). A simulation study of ridge regression estimators with auto-correlated errors. *Communications in Statistics–Simulation and Computation*, **18(2)** 673-702.
- [6] Gruber, M.H.J. (1991). The efficiency of Jack-knife and usual ridge type estimators: A comparison. *Statistics and Probability Letters*, **11** 49-51.
- [7] Gruber, M.H.J. (1998). *Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators*. New York: Marcell Dekker.
- [8] Hinkley, D.V. (1977). Jack-knifing in unbalanced situations. *Technometrics*, **19(3)** 285-292.
- [9] Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, **20** 69-82.
- [10] Hoerl, A.E., Kennard, R.W. and Baldwin, K. (1975). Ridge regression: Some simulations. *Communications in Statistics–Theory and Methods*, **4** 105-123.
- [11] McDonald, G.C. and Galarneau, D.I. (1975). A Monte-Carlo evaluation of some Ridge-type estimators. *Journal of the American Statistical Association*, **70** 407-416.
- [12] Miller, R.G. (1974a). The Jack-knife: A Review. *Biometrika*, **61** 1-15.
- [13] Miller, R.G. (1974b). An unbalanced Jack-knife. *Annals of Statistics*, **2** 880-891.
- [14] Nomura, M. (1988). On the almost unbiased ridge regression estimator. *Communications in Statistics–Simulation and Computation*, **17** 729-743.

- [15] Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika*, **43** 353-360.
- [16] Sarkar, N. (1992). A new estimator combining the ridge regression and the restricted least squares methods of estimation. *Communications in Statistics–Theory and Methods*, **21** 1987-2000.
- [17] Singh, B., Chaubey, Y.P. and Dwivedi, T.D. (1986). An almost unbiased ridge estimator. *Sankhya*, **B48** 342-36.
- [18] Singh, B. and Chaubey, Y.P. (1987). On some improved ridge estimators. *Statistical Papers*, **28** 53-67.
- [19] Swindel, B.F. (1976). Good ridge estimators based on prior information. *Communications in Statistics-Theory and Methods*, **5** 1065-1075.
- [20] Thorburn, D. (1976). Some asymptotic properties of Jackknife statistics. *Biometrika*, **63** 305-313.
- [21] Tukey, J.W. (1958). Bias and confidence in not quite large samples (Abstract). *Annals of Mathematical Statistics*, **29** p. 614.

List of Recent Technical Reports

83. Ze-Chun Hu, Zhi-Ming Ma and Wei Sun, *Formulae of Beurling–Deny and Lejan for Non-Symmetric Dirichlet Forms*, February 2006
84. Ze-Chun Hu and Wei Sun, *A Note on Exponential Stability of the Non-Linear Filter for Denumerable Markov Chains*, February 2006
85. H. Brito–Santana, R. Rodríguez–Ramos, R. Guinovart–Díaz, J. Bravo–Castillero and F.J. Sabina, *Variational Bounds for Multiphase Transversely Isotropic Composites*, August 2006
86. José Garrido and Jun Zhou, *Credibility Theory for Generalized Linear and Mixed Models*, December 2006
87. Daniel Dufresne, José Garrido and Manuel Morales, *Fourier Inversion Formulas in Option Pricing and Insurance*, December 2006
88. Xiaowen Zhou, *A Superprocess Involving Both Branching and Coalescing*, December 2006
89. Yogendra P. Chaubey, Arusharka Sen and Pranab K. Sen, *A New Smooth Density Estimator for Non–Negative Random Variables*, January 2007
90. Md. Sharif Mozumder and José Garrido, *On the Relation between the Lévy Measure and the Jump Function of a Lévy Process*, October 2007
91. Arusharka Sen and Winfried Stute, *A Bi-Variate Kaplan–Meier Estimator via an Integral Equation*, October 2007
92. C. Sangüesa, *Uniform Error Bounds in Continuous Approximations of Nonnegative Random Variables Using Laplace Transforms*, January 2008
93. Yogendra P. Chaubey, Naâmane Laïb and Arusharka Sen, *A Smooth Estimator of Regression Function for Non-negative Dependent Random Variables*, March 2008
94. Alejandro Balbás, Beatriz Balbás and Antonio Heras, *Optimal Reinsurance with General Risk Functions*, March 2008

95. Alejandro Balbás, Raquel Balbás and José Garrido, *Extending Pricing Rules with General Risk Functions*, April 2008
96. Yogendra P. Chaubey and Pranab K. Sen, *On the Selection of the Smoothing Parameter in Poisson Smoothing of Histogram Estimator: Computational Aspects*, December 2008
97. Runhuan Feng and José Garrido, *Actuarial Applications of Epidemiological Models*, December 2008
98. Alejandro Balbás, Beatriz Balbás and Raquel Balbás, *CAPM and APT Like Models with Risk Measures*, June 2009
99. Jun Zhou and José Garrido, *A Loss Reserving Method Based on Generalized Linear Models*, June 2009
100. Yogendra P. Chaubey and Isha Dewan, *Smooth Estimation of Survival and Density Functions for a Stationary Associated Process Using Poisson Weights*, September 2009
101. Krishna K. Saha and Debaraj Sen, *Improved Confidence Interval for the Dispersion Parameter in Count Data Using Profile Likelihood*, September 2009
102. Araceli Reyes, *Difficulties Teaching Mathematics with the ACE Cycle*, September 2009
103. Mansi Khurana, Yogendra P. Chaubey and Shalini Chandra, *Jackknifing the Ridge Regression Estimator: A Revisit*, February 2012

Copies of technical reports can be requested from:

Dr. Wei Sun
Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd. West,
Montreal, QC, H3G 1M8 CANADA