Hacettepe Journal of Mathematics and Statistics
Volume 47 (6) (2018), 1625 – 1633

New approaches for choosing the ridge parameters

J. Al-Jararha*

Abstract

Consider the standard multiple linear regression model $y = x\beta + \varepsilon$. If the correlation matrix $x^t x$ is ill-conditioned, the ordinary least squared estimate (ols) $\hat{\beta}$ of β is not the best choice. In this paper, multiple regularization parameters for different coefficients in ridge regression are proposed. The Mean Squared Error (MSE) of a ridge estimate based on the multiple regularization parameters is less than or equal to the MSE of the ridge estimate based on Hoerl and Kennard, 1970. The proposed approach, depending on the condition numbers, leave's zero for the largest eigenvalue of $x^t x$ and gives the largest value for the smallest eigenvalue of $x^t x$. Furthermore, if $x^t x$ is nearly a unit matrix, $x^t x$ is not an ill-conditioned one. The proposed approach gives approximately the same results as the ols estimates. The proposed approach can also be modified to give other new ridge parameters. The modified approach depends on the eigenvalues of $x^t x$ and differ from the ridge parameter proposed by Khalaf and Shukur by a factor. The body fat data set has severe multicollinearity and is used to compare different approaches.

Keywords: Ridge regression, multicollinearity, condition number, mean squared error, squared bias.

Mathematics Subject Classification (2010): Primary 62J02, Secondary 62J12

Received : 07.03.2016 *Accepted* : 17.11.2016 *Doi* : 10.15672/HJMS.2016.389

*Department of Statistics, Yarmouk University, Irbid, Jordan, Email: jjararha@gmail.com

1. Introduction

Consider the standard multiple linear regression model

$$y = x\beta + \varepsilon \tag{1.1}$$

where $y_{n\times 1}$ is a vector of responses, $x_{n\times p}$ is the design matrix of rank p, $\beta_{p\times 1}$ is a vector of unknown parameters, $\varepsilon \sim N_n\left(\mathbf{0}, \sigma^2 I_n\right)$, and I_n is the identity matrix of rank n. The ols estimator of β is

$$\hat{\beta} = r_{xx}^{-1} r_{xy}, \tag{1.2}$$

where $r_{xx} = x^t x$ is the correlation matrix of x variables and $r_{xy} = x^t y$ is the correlation vector between y and each x variables. It is known that $\hat{\beta}$ is an unbiased estimator for β and has minimum variance. The variance of $\hat{\beta}$ is given by

$$var\left(\hat{\beta}\right) = \sigma^2 r_{xx}^{-1}. \tag{1.3}$$

If r_{xx} is not nearly a unit matrix, multicollinearity may exists in the design matrix which in this case, r_{xx} matrix is an ill-conditioned one. From equation (1.2) and (1.3), the ols estimator $\hat{\beta}$ will give an inaccurate estimate for β and inflate its variance. Therefore, the ols estimator $\hat{\beta}$ of β is not the best choice. One of the solutions of the multicollinearity problem is the ridge regression method, which was proposed by [5]. Based on the ridge regression method, estimate β by

$$\tilde{\beta} = \left(r_{xx} + kI_p\right)^{-1} r_{xy}. \tag{1.4}$$

For some ridge parameter $k \geq 0$, from equation (1.4) simply add positive constant k to the main diagonal of the correlation matrix r_{xx}. However, the approach will be adopted by adding different constants to the main diagonal of r_{xx}. In this case, estimate β by

$$\tilde{\beta} = \left(r_{xx} + diag\left(k_1, k_2, \ldots, k_p\right)\right)^{-1} r_{xy}, \tag{1.5}$$

where $diag$ is the diagonal matrix with $k_1, k_2, \ldots, k_p$ on the main diagonal and $k_i \geq 0$, $i = 1, 2, \ldots, p$. It is clear that equation (1.2) and (1.4) are special cases from equation (1.5) with the choices $k_1 = k_2 = \ldots = k_p = 0$ and $k_1 = k_2 = \ldots = k_p = k$, respectively.

The main problem in ridge regression is the method of choosing the ridge parameter(s). [5] showed that such parameter exists and the MSE of the ridge parameter $\tilde{\beta}$ of β is less than the MSE of the ols $\hat{\beta}$ of β. [5] proposed a method for choosing the ridge parameter(s) which is described in detail in the next section. [8] proposed another approach for choosing the ridge parameter k. [1] modified the two different approaches proposed by [5] and [8]. [3] proposed methods for estimating the ridge parameter(s). [7] proposed different estimators of the ridge parameter k and compared, via simulation, with estimators proposed by [3], [8] and [9].

In case of the multinomial logit model, [11] considered several estimators for estimating the ridge parameter k. Based on the simulations, when the correlation between the independent variable increases the MSE increases. At the same time, increasing the sample size decreases the MSE even when the correlation between the independent variables is large. The non-Gaussian error terms and the highly collinear predictors are considered by [14] which compared the least squares ridge estimation and the Least Absolute Deviations (LAD) ridge estimation of the Seemingly Unrelated Regression Equations (SURE) models through the MSE. [6] considered linear regression having both heteroskedasticity and collinearity problems. The main result states thats the heteroskedasticity-robust variances can be improved and the resulting bias is minimized by using the matrix perturbation method. A strong consistency of the ridge estimates is established by [2]. The only requirement for the error term to be iid with absolute moment of order r $(0 < r \leq 1)$.

[10] proposed a quasi-stochastically constrained least squares estimator and provide the expectation of this estimator, demonstrate its consistency and asymptotic normality.

2. Ridge parameter

This section summarizes the approach proposed by [5]. Let

$$\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \lambda_{min} > 0,$$

be the eigenvalues of the matrix r_{xx} with corresponding eigenvectors $\nu_1, \nu_2, \ldots, \nu_p$, respectively, and $P = (\nu_1, \nu_2, \ldots, \nu_p)$. [5] showed that the MSE of $\tilde{\beta}$ is given by

$$\begin{aligned} MSE_{\tilde{\beta}}(k) &= E\left(\tilde{\beta} - \beta\right)^t \left(\tilde{\beta} - \beta\right) \\ &= \gamma_1(k) + \gamma_2(k) \end{aligned} \tag{2.1}$$

where $\gamma_1(k)$ and $\gamma_2(k)$ are the variance and squared bias of $\tilde{\beta}$, respectively, when the ridge parameter k is used. Further, $\gamma_1(k)$ and $\gamma_2(k)$ are defined by

$$\gamma_1(k) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k_i)^2}, \tag{2.2}$$

and

$$\gamma_2(k) = \sum_{i=1}^{p} \frac{\alpha_i^2 k_i^2}{(\lambda_i + k_i)^2} \tag{2.3}$$

where α_i is the i^{th} element of $\alpha = P\beta$.

It is known that the ols $\hat{\beta}$ is an unbiased estimator (with the choice $k_1 = k_2 = \ldots = k_p = 0$ in equation (1.5)) for β. Therefore, $\gamma_2(0) = 0$, and

$$\begin{aligned} MSE_{\tilde{\beta}}(0) &= \gamma_1(0) \\ &= \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i}. \end{aligned} \tag{2.4}$$

However, the ridge estimator $\tilde{\beta}$ of β is a biased estimator, and [5] showed that there always exists a $k > 0$ such that

$$MSE_{\tilde{\beta}}(k) < MSE_{\tilde{\beta}}(0), \tag{2.5}$$

if

$$k < \frac{\sigma^2}{\alpha_{max}^2}. \tag{2.6}$$

Therefore, [5] adopted

$$k_{hk_i} = \frac{\sigma^2}{\alpha_i^2}, \quad i = 1, 2, \ldots, p \tag{2.7}$$

as ridge parameters. [8] proposed the following ridge parameter

$$k_{gg} = \frac{\sigma^2}{\alpha_{max}^2 + \frac{n-p}{\lambda_{max}}\sigma^2}. \tag{2.8}$$

Other approaches are dealing with different methods of estimation the ridge parameter k. A new ridge parameter based on modification of $MSE_{\tilde{\beta}}(k)$ is proposed in the next section.

3. Proposed approaches

In this section, two approaches are proposed for choosing the ridge parameter $k = (k_1, k_2, \ldots, k_p)^t$ by refining the $MSE_{\tilde{\beta}}(k)$, given by equation (2.1).

From equation (1.5), recall the definition of $\tilde{\beta}$,

$$\begin{aligned} \tilde{\beta} &= (r_{xx} + diag(k_1, k_2, \ldots, k_p))^{-1} r_{xy} \\ &= W r_{xy} \quad (3.1) \\ &= Z\hat{\beta} \quad (3.2) \end{aligned}$$

where

$W = (r_{xx} + diag(k_1, k_2, \ldots, k_p))^{-1}$ and $Z = \left(I_p + r_{xx}^{-1} diag(k_1, k_2, \ldots, k_p)\right)^{-1}$. Further,

$$Z = I_p - W\, diag(k_1, k_2, \ldots, k_p). \quad (3.3)$$

For $i = 1, 2, \ldots, p$. Let

$$\xi_i(W) = \frac{1}{\lambda_i + k_i} \quad (3.4)$$

$$\xi_i(Z) = \frac{\lambda_i}{\lambda_i + k_i} \quad (3.5)$$

be the eigenvalues of W and Z respectively.

Now, the bias of $\tilde{\beta}$ is given by

$$\begin{aligned} bias\left(\tilde{\beta}\right) &= E\left(\tilde{\beta} - \beta\right) \\ &= -(I_p - Z)\beta \\ &= -W\, diag(k_1, k_2, \ldots, k_p)\beta. \quad (3.6) \end{aligned}$$

The eigenvalues of $W\, diag(k_1, k_2, \ldots, k_p)$ are

$$\frac{k_i}{\lambda_i + k_i}, \text{ for } i = 1, \ldots, p. \quad (3.7)$$

Therefore, the matrix $W\, diag(k_1, k_2, \ldots, k_p)$ is a positive definite for $k_i > 0$ for $i = 1, \ldots, p$.

Define $\gamma_3(k)$ as the sum of the eigenvalues of $W\, diag(k_1, k_2, \ldots, k_p)$ i.e.

$$\gamma_3(k) = \sum_{i=1}^{p} \frac{k_i}{\lambda_i + k_i} \quad (3.8)$$

3.1. Remark. *$\gamma_3(k)$ is the trace of a positive definite matrix $W\, diag(k_1, k_2, \ldots, k_p)$ and can be used to control the bias of $\tilde{\beta}$ as we can see from equation (3.6). The key point is to balance between the bias and the variance of the ridge estimate of β, by keeping $\gamma_3(k)$ falls between $\gamma_3(0) = 0$ and $\gamma_3(k_{hk_i}) = \sigma^2 \sum_{i=1}^{p} 1/\left(\sigma^2 + \alpha_i^2 \lambda_i\right)$.*

$MSE_{\tilde{\beta}}(k)$ is minimized by [5]. Therefore, the proposed method chooses $k_1, k_2, \ldots, k_p$ such that the right hand side of equation (3.8) to be as small as possible. For $\varepsilon > 0$ define the following function

$$\begin{aligned} G(k) &= MSE_{\tilde{\beta}}(k) + \varepsilon\gamma_3(k) \\ &= \gamma_1(k) + \gamma_2(k) + \varepsilon\gamma_3(k), \quad (3.9) \end{aligned}$$

$G(k)$ is a direct modification of [5] by subtracting a controlled amount $\varepsilon\gamma_3(k)$ from $MSE_{\tilde{\beta}}(k)$, note the negative sign before the right hand side of equation (3.6), where $\gamma_1(k)$, $\gamma_2(k)$, and $\gamma_3(k)$ are defined by equation (2.2), (2.3), and (3.8) respectively.

3.2. Remark. *Minimizing $G(k)$ is expected to give better results than minimizing $MSE_{\tilde{\beta}}(k)$ since a small positive amount $\varepsilon\gamma_3(k)$ is subtracted from $MSE_{\tilde{\beta}}(k)$ by keeping $G(k) > 0$.*

In the same time, minimizing $G(k)$ is to choose k only and the comparisons will be based on the MSE.

Differentiate the right hand side of equation (3.9) with respect to k_i, equate to zero and solve for k_i we have

$$k_i = \frac{2\sigma^2 - \varepsilon\lambda_i}{2\alpha_i^2 + \varepsilon}, \quad i = 1, \ldots, p \tag{3.10}$$

To grantee $k_i \geq 0$ for $i = 1, \ldots, p$, equation (3.10) implies $\varepsilon \leq 2\sigma^2/\lambda_i$. Therefore, take $\varepsilon = 2\sigma^2/\lambda_{max}$. Substitute this value of ε in equation (3.10), we have

$$k_i = \frac{\left[\left(\frac{\lambda_{max}}{\lambda_i}\right) - 1\right]\sigma^2}{\left(\frac{\lambda_{max}}{\lambda_i}\right)\alpha_i^2 + \frac{1}{\lambda_i}\sigma^2}, \quad i = 1, \ldots, p \tag{3.11}$$

3.3. Remark. *From equation (3.11), rewrite k_i as*

$$\begin{aligned} k_i \quad &= \quad \begin{cases} 0, & i = 1 \\ \dfrac{\sigma^2}{\frac{\left(\frac{\lambda_{max}}{\lambda_i}\right)}{\left[\left(\frac{\lambda_{max}}{\lambda_i}\right)-1\right]}\alpha_i^2 + \frac{\frac{1}{\lambda_i}}{\left[\left(\frac{\lambda_{max}}{\lambda_i}\right)-1\right]}\sigma^2}, & i = 2, 3, \ldots, p \end{cases} \\ &< \quad \frac{\sigma^2}{\alpha_{max}^2}. \end{aligned} \tag{3.12}$$

Therefore, the condition defined by equation (2.6) holds for this choice of k_i.

3.4. Remark. *The ridge parameters given by equation (3.11) are functions of the Condition Numbers $= \frac{\lambda_{max}}{\lambda_i}$. Condition numbers can be used as indication of multicollinearity, $100 \leq \frac{\lambda_{max}}{\lambda_i} \leq 1000$ indicates a mild multicollinearity, and $\frac{\lambda_{max}}{\lambda_i} > 1000$ indicates severe multicollinearity problem* [12]. *It is clear from equation (3.11), k_i is approximately zero if r_{xx} is approximately a unit matrix, then the multicollinearity problem disappears. In this case, the ols $\hat{\beta}$ and the ridge estimator $\tilde{\beta}$ are approximately the same.*

3.5. Remark. *From equation (3.11) it is clear that we are touching the exact problem by giving zero for the largest eigenvalue of $x^t x$ and the largest value is given for the smallest eigenvalue of $x^t x$. This will be considered one of the main advantages of using the proposed multiple regularization parameters.*

From equation (3.11), rewrite k_i as

$$\begin{aligned} k_i \quad &= \quad \frac{\lambda_{max}\sigma^2}{\lambda_{max}\alpha_i^2 + \sigma^2} - \frac{\sigma^2}{\left(\frac{\lambda_{max}}{\lambda_i}\right)\alpha_i^2 + \frac{1}{\lambda_i}\sigma^2} \\ &= \quad \frac{\lambda_{max}\sigma^2}{\lambda_{max}\alpha_i^2 + \sigma^2} - \frac{\sigma^2\lambda_i}{\lambda_{max}\alpha_i^2 + \sigma^2}, \qquad i = 1, \ldots, p \end{aligned} \tag{3.13}$$

In case of severe multicollinearity, λ_{min} is positive and very close to zero. Therefore, we can ignore the term $\sigma^2\lambda_i/\left(\lambda_{max}\alpha_i^2 + \sigma^2\right)$. In this case, equation (3.13), reduces to

$$k_{0_i} = \frac{\sigma^2}{\alpha_i^2 + \frac{1}{\lambda_{max}}\sigma^2}, \quad i = 1, \ldots, p \tag{3.14}$$

3.6. Remark. *The ridge parameters given by equation (3.14) are similar to the ridge parameter proposed by* [8],

$$k_{gg} = \frac{\sigma^2}{\alpha^2_{max} + \frac{n-p}{\lambda_{max}}\sigma^2}. \tag{3.15}$$

For the purpose of comparisons, rewrite k_{gg} *as*

$$k_{gg_i} = \frac{\sigma^2}{\alpha^2_i + \frac{n-p}{\lambda_{max}}\sigma^2}. \tag{3.16}$$

The difference between k_{gg_i} *and* k_{0_i} *is* $n-p$*, the coefficient of* σ^2/λ_{max} *appears in the denominator.*

Further, from equation (3.14), it is clear that for

$$k_{0_i} < \frac{\sigma^2}{\alpha^2_{max}}$$

the condition defined by equation(2.6) is satisfied for k_{0_i}.

4. Empirical study

In ridge regression, the body fat data set is considered by [13], [?] and others. To standardize the model the following terms are defined.

The standardized j^{th} observation of the i^{th} predictor variable is

$$x_{ij} = \frac{X_{ij} - \bar{X}_i}{\sqrt{n-1}S_{X_i}}, \tag{4.1}$$

and the standardized j^{th} observation of the response variable is

$$y_j = \frac{Y_j - \bar{Y}}{\sqrt{n-1}S_Y}, \tag{4.2}$$

where

$$S_{X_i} = \sqrt{\sum_{j=1}^{n} \left(X_{ij} - \bar{X}_i\right)^2 / (n-1)},$$

$$S_Y = \sqrt{\sum_{j=1}^{n} \left(Y_j - \bar{Y}\right)^2 / (n-1)},$$

for $i = 1, \ldots, p$ and $j = 1, \ldots, n$. The standardized model in matrix form is given by

$$y = x\beta + \varepsilon \tag{4.3}$$

with the usual assumptions. Where

$$x = \begin{pmatrix} x_{11} & \ldots & x_{p1} \\ \vdots & \ldots & \vdots \\ x_{1n} & \ldots & x_{pn} \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \text{and} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}.$$

Recall from equation (1.5), the ridge estimate of β is

$$\tilde{\beta} = \left(r_{xx} + diag\left(k_1, k_2, \ldots, k_p\right)\right)^{-1} r_{xy} \tag{4.4}$$

and the ridge estimate of y is

$$\tilde{y} = x\tilde{\beta}. \tag{4.5}$$

The error sum of squares for ridge regression is

$$SSE_{ridge} = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 ; \tag{4.6}$$

The body fat data set consists of ($n = 20$) observations and ($p = 3$) predictor variables. The predictor variables are X_1 : triceps skinfold thickness, X_2 : thigh circumference, and X_3 : midarm circumference. The response variable Y : the body fat. The correlation between the variables X_1 and X_2 is 0.92384 and the p_{value} for testing $H_0 : \rho_{x_1 x_2} = 0$ is less than 0.0001. Further, $\lambda_1 = 2.0664727$, $\lambda_2 = 0.9328007$, $\lambda_3 = 0.0007266$, $\lambda_1/\lambda_2 = 2.21534$, and $\lambda_1/\lambda_3 = 2844.030691$. Therefore, this data set has severe multicollinearity.

For the body fat data set, compute $\gamma_1(k)$, $\gamma_2(k)$, and $MSE_{\tilde{\beta}}(k)$ for $k = 0$, k_{hk_i}, k_{gg_i}, k_i and k_{0_i}. Where $\gamma_1(k)$, $\gamma_2(k)$, and $MSE_{\tilde{\beta}}(k)$, are variance, squared bias , and the MSE of $\tilde{\beta}$ and defined by equation (2.2), (2.3), and (2.1) respectively.

Under simple random sampling without replacement design, simulate m random samples of sizes $n = 5, 6, \ldots, 10, 12, \ldots, 18$. For $i = 1, \ldots, m$, compute $\gamma_1^{(i)}(k)$, $\gamma_2^{(i)}(k)$, $MSE_{\tilde{\beta}}^{(i)}(k)$, for the same choices of k.

Simulate $m = 5000$ random samples for $n = 5, 6, \ldots, 10, 12, 14$; $m = 4845$ for $n = 16$; and $m = 190$ for $n = 18$.

Based on the ridge parameter k estimate σ^2 by

$$\tilde{\sigma}^2_{ridge}(k) \quad = \quad SSE_{ridge}/(n-p), \tag{4.7}$$

be the ridge estimator of σ^2. Let $\gamma_1(k)$, $\gamma_2(k)$, and $MSE_{\tilde{\beta}}(k)$ be the averages of $\gamma_1^{(i)}(k)$, $\gamma_2^{(i)}(k)$, $MSE_{\tilde{\beta}}^{(i)}(k)$, computed from m simulated random samples. The results are summarized in Table (1).

4.1. Results and Conclusions. From Table (1) we have the following results:

(1) The results from the body fat data set are summarized by:
 (a) $MSE_{\tilde{\beta}}(k_i) < MSE_{\tilde{\beta}}(k_{hk_i}) < MSE_{\tilde{\beta}}(k_{0_i}) < MSE_{\tilde{\beta}}(k_{gg_i}) < MSE_{\tilde{\beta}}(0)$.
 (b) $\gamma_1(k_i) < \gamma_1(k_{hk_i}) < \gamma_1(k_{0_i}) < \gamma_1(k_{gg_i}) < \gamma_1(0)$.
 (c) $\gamma_2(0) < \gamma_2(k_{gg_i}) < \gamma_2(k_i) < \gamma_2(k_{0_i}) < \gamma_2(k_{hk_i})$.

(2) The simulation results from the body fat data set are summarized by:
 (a) For $n = 5, \ldots, 10$:

$$MSE_{\tilde{\beta}}(k_i) < MSE_{\tilde{\beta}}(k_{gg_i}) < MSE_{\tilde{\beta}}(k_{0_i}) < MSE_{\tilde{\beta}}(k_{hk_i}) < MSE_{\tilde{\beta}}(0).$$

 For $n = 12,\ 14$:

$$MSE_{\tilde{\beta}}(k_i) < MSE_{\tilde{\beta}}(k_{0_i}) < MSE_{\tilde{\beta}}(k_{hk_i}) < MSE_{\tilde{\beta}}(k_{gg_i}) < MSE_{\tilde{\beta}}(0),$$

 and $MSE_{\tilde{\beta}}(k)$ follows the same pattern as the data set for $n = 16,\ 18$.
 (b) For $n = 5,\ 6,\ 7$:

$$\gamma_1(k_i) < \gamma_1(k_{gg_i}) < \gamma_1(k_{0_i}) < \gamma_1(k_{hk_i}) < \gamma_1(0),$$

 for $n = 8,\ 9,\ 10$:

$$\gamma_1(k_i) < \gamma_1(k_{0_i}) < \gamma_1(k_{gg_i}) < \gamma_1(k_{hk_i}) < \gamma_1(0),$$

 and $\gamma_1(k)$ follows the same pattern as the original data set for $n = 12,\ 14,\ 16,\ 18$.
 (c) For $n = 5, \ldots, 10,\ 12, \ldots, 18$, $\gamma_2(k)$ follows the same pattern as the original data set.
 Based on the computations and simulations from the body fat data set, we can conclude that our choice for the ridge parameter $k = k_i$ has minimum variance and minimum MSE among all other choices for the ridge parameters.

Table 1. The computations are rounded into two digits after decimals. $\gamma_1(k)$, $\gamma_2(k)$, and $MSE_{\tilde{\beta}}(k)$ are the means of m random samples, and for different sample sizes $n = 5, \ldots, 18$. When $n = 20$: $\gamma_1(k)$, $\gamma_2(k)$, and $MSE_{\tilde{\beta}}(k)$ are computed from the data set. σ^2 is estimated by $\tilde{\sigma}^2_{ridge}(k)$.

n		$k=0$	$k=k_{hk_i}$	$k=k_{gg_i}$	$k=k_i$	$k=k_{0_i}$
	$\gamma_1(k)$	340.87	36.96	36.91	30.93	36.93
5	$\gamma_2(k)$	000.00	23.35	23.35	23.35	23.35
	$MSE_{\tilde{\beta}}(k)$	340.87	60.31	60.26	54.28	60.28
	$\gamma_1(k)$	113.32	10.31	10.27	8.77	10.28
6	$\gamma_2(k)$	0.00	9.35	9.34	9.34	9.35
	$MSE_{\tilde{\beta}}(k)$	113.32	19.65	19.61	18.11	19.63
	$\gamma_1(k)$	76.22	6.31	6.30	5.55	6.30
7	$\gamma_2(k)$	0.00	6.39	6.38	6.39	6.39
	$MSE_{\tilde{\beta}}(k)$	76.22	12.70	12.68	11.93	12.69
	$\gamma_1(k)$	57.74	4.31	4.31	3.90	4.31
8	$\gamma_2(k)$	0.00	5.03	5.02	5.03	5.03
	$MSE_{\tilde{\beta}}(k)$	57.74	9.34	9.33	8.93	9.34
	$\gamma_1(k)$	46.34	3.03	3.04	2.79	3.03
9	$\gamma_2(k)$	0.00	3.96	3.95	3.95	3.96
	$MSE_{\tilde{\beta}}(k)$	46.34	6.99	6.99	6.74	6.99
	$\gamma_1(k)$	39.14	2.27	2.27	2.12	2.27
10	$\gamma_2(k)$	0.00	3.39	3.38	3.39	3.39
	$MSE_{\tilde{\beta}}(k)$	39.14	5.66	5.66	5.51	5.66
	$\gamma_1(k)$	30.47	1.41	1.42	1.34	1.41
12	$\gamma_2(k)$	0.00	2.60	2.59	2.60	2.60
	$MSE_{\tilde{\beta}}(k)$	30.47	4.01	4.01	3.94	4.01
	$\gamma_1(k)$	24.73	0.99	1.00	0.96	0.99
14	$\gamma_2(k)$	0.00	2.20	2.19	2.20	2.20
	$MSE_{\tilde{\beta}}(k)$	24.73	3.19	3.19	3.15	3.19
	$\gamma_1(k)$	20.88	0.69	0.71	0.67	0.69
16	$\gamma_2(k)$	0.00	2.00	1.99	2.00	2.00
	$MSE_{\tilde{\beta}}(k)$	20.88	2.70	2.70	2.68	2.70
	$\gamma_1(k)$	18.29	0.50	0.52	0.49	0.51
18	$\gamma_2(k)$	0.00	1.87	1.86	1.87	1.87
	$MSE_{\tilde{\beta}}(k)$	18.29	2.38	2.38	2.36	2.38
	$\gamma_1(k)$	16.10	0.31	0.33	0.31	0.32
20	$\gamma_2(k)$	0.00	1.85	1.83	1.85	1.85
	$MSE_{\tilde{\beta}}(k)$	16.10	2.17	2.17	2.16	2.17

5. Final conclusions

In this paper a new approach for choosing the ridge parameter k has been proposed. In case of severe multicollinearity the proposed approach is also modified for choosing the ridge parameter k.

In case where the correlation matrix $x^t x$ is ill-conditioned based on computations and simulations from the real data set, the proposed approach for the ridge parameter k has a minimum MSE and a minimum variance of $\tilde{\beta}$ among other approaches discussed in this paper. Furthermore the proposed approach can be adopted in the case where no multicollinearity problem exist since the proposed approach and the ols method give approximately the same results.

Acknowledgments The author is thankful to the anonymous referees for their valuable suggestions and comments which led to an improvement of this article.

References

[1] Alkhamisi, M., Khalaf, G., Shukur, G., Some Modifications for Choosing Ridge Parameters, Communications in Statistics - Theory and Methods, 35 (11), 2005-2020, 2006.

[2] Da Silva, J. L., Mexia, J. T., Ramos, L. P., On the Strong Consistency of Ridge Estimates, Communications in Statistics - Theory and Methods, 44 (3), 617-626, 2015.

[3] Gisela, M., Kibria, B. M. G., On Some Ridge Regression Estimators: An Emirical Comparisons, Communications in Statistics - Simulation and Computation, 38 (3), 621–630, 2009.

[4] Hocking, R. R., Methods and Applications of Linear Models, Wiley Series in Probability and Statistics, 2003.

[5] Hoerl, A. E., Kennard, R. W., Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12, 55-67, 1970.

[6] Huang, C.-C. L., Jou, Y.-J., Cho, H.-J., VIF-based adaptive matrix perturbation method for heteroskedasticity-robust covariance estimators in the presence of multicollinearity, Communications in Statistics - Theory and Methods, http://dx.doi.org/10.1080/03610926.2015.1060340.

[7] Khalaf, G., Mansson, K., Shukur, G., Modified Ridge Regression Estimators, Communications in Statistics - Theory and Methods, 42, 1476–1487, 2013.

[8] Khalaf, G., Shukur, G., Choosing Ridge Parameter for Regression Problems, Communications in Statistics - Theory and Methods, 34 (5), 1177–1182, 2005.

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

[10] Li, S., Sheng, Y., Li, Y., The quasi-stochastically constrained least squares method for ill linear regression, Communications in Statistics - Theory and Methods, 45 (2), 217-225, 2016.

[11] Månsson, K., Shukur, G., Kibria, B. M. G., Performance of Some Ridge Regression Estimators for the Multinomial Logit Model, Communications in Statistics - Theory and Methods, http://dx.doi.org/10.1080/03610926.2013.784996.

[12] NCSS Statistical Software, Chapter 335: Ridge Regression, NCSS, LLC. All Rights reserved.

[13] Neter, J., Kutner, M. H., Nachtsheim, C., J., Wasserman, W., Applied Linear Statistical Models, 4th Edition, 1996.

[14] Zeebari, Z., Shukur, G., On the Least Absolute Deviations Method for Ridge Estimation of SURE Models, Communications in Statistics - Theory and Methods, http://dx.doi.org/10.1080/03610926.2012.755203