

Choosing ridge Parameters in the Linear regression Model with AR(1) Error: A Comparative Simulation Study

M. I. Alheety¹ and B. M. Golam Kibria²

¹Department of Mathematics
Alanbar University, Iraq
Email:alheety@yahoo.com

²Department of Mathematics and Statistics
Florida International University
Modesto A. Maidique Campus
Miami, FL 33199, USA
Email: kibriag@fiu.edu

ABSTRACT

The performance of shrinkage ridge estimators in the linear regression model has been studied to reduce the effect of multicollinearity on the estimation of the linear regression parameters. Trenkler (1984) proposed a ridge estimator in the linear regression model when the assumption of uncorrelatedness is not satisfied. Since there is no attempt to study the recent types of estimated ridge parameter when the assumption of uncorrelatedness is not satisfied, this paper tries to show the performance of some ridge estimators in the linear regression model with correlated error based on the minimum mean squared error (MSE) criterion. A simulation study and a numerical example have been made to evaluate the performance of these estimators of ridge parameter k . The simulation study suggests that some ridge estimators are promising and can be recommended for the practitioners.

Keywords: Linear Model; Multicollinearity; MSE; Ridge Regression; Simulation Study.
2000 Mathematics Subject Classification: 62J07; 62F10.

1 Introduction

Consider the following multiple linear regression model

$$Y = X\beta + \epsilon, \quad (1.1)$$

where Y is an $n \times 1$ vector of observations, X is an $n \times p$ matrix, β is an $p \times 1$ vector of unknown parameters, and ϵ is an $n \times 1$ vector of non observable errors with $E(\epsilon) = 0$ and $Cov(\epsilon) = \sigma^2 I_n$.

The most common method used for estimating the regression coefficients in (1.1) is the ordinary least squares (OLS) method which is defined as:

$$\hat{\beta} = (X'X)^{-1}X'Y. \quad (1.2)$$

Both OLS estimator and its covariance matrix heavily depend on the characteristics of $X'X$ matrix. If XX is ill-conditioned, i.e. the column vectors of X are linearly dependent, then the OLS estimators are sensitive to a number of errors. For example, some of the regression coefficients may be statistically insignificant or have the wrong sign, and they may result to wide confidence intervals for individual parameters. With ill-conditioned $X'X$ matrix, it is difficult to make valid statistical inferences about the regression parameters.

One of the most popular estimators dealing with multicollinearity is the ordinary ridge regression (ORR) estimator proposed by Hoerl and Kennard (1970a,b) and defined as:

$$\hat{\beta}_k = [X'X + kI_p]^{-1}X'Y = [I_p + k(X'X)^{-1}]^{-1}\hat{\beta}. \quad (1.3)$$

The constant $k > 0$ is known as shrinkage or biasing or ridge parameter. As k increase from zero upto infinity, the regression estimates go to zero. Though this estimator is biased, for certain value of k , they yield *minimum mean squares error* (MSE) compared to OLS (see Hoerl and Kennard (1970a,b)). However, the $MSE(\hat{\beta}(k))$ will depend on unknown parameters k , β and σ^2 , which can not be calculated in practice. But k , has to be estimated from the real data instead. Much of the discussions on ridge regression concern the problem of finding good empirical value of k . Many different techniques for estimating k have been proposed or suggested by different researchers. Hoerl and Kennard (1970a,b), Hoerl et al. (1975), McDonald and Galarneau (1975), Lawless (1978), Lawless and Wang (1976), Dempster et al. (1977), Gunst and Mason (1977), Hemmerle and Brantle (1978), Wichern and Churchill (1978), Golub et al. (1979), Gibbons (1981), Nordberg (1982), Saleh and Kibria (1993), Haq and Kibria (1996), Kibria (1996, 2003), Singh and Tracy (1999), Wencheko (2000), Khalaf and Shukur (2005), Alkhamisi et al. (2006), Alkhamisi and Shukur (2008), Alheety and Ramanathan (2009), Muniz and Kibria (2009) and very recently Mansson et al. (2010) to mention a few.

Time series data occur frequently in business, economics, and some fields of engineering and the assumption of uncorrelated or homoscedastic errors for time series data is often inappropriate. The presence of autocorrelation in the errors has several effects on OLS regression procedure. In this case, we note that OLS will no longer be efficient and that the usual estimator for the variance-covariance matrix will be biased. Having a biased variance-covariance matrix estimator means that confidence intervals and hypothesis tests are no longer soundly based procedures (Griffiths et al. (1993)). To overcome these effects, alternative methods of estimation were used. Weighted or generalized least squares (GLS) method could be used if there is sufficient knowledge of the autocorrelation structure. The generalized least squares estimator (GLS) is also unbiased and has lower variance than OLS. In model (1.1) we assumed that $Cov(\epsilon) = \sigma^2I$ which is called homoscedasticity i.e. $Var(\epsilon_i) = \sigma^2$, for $i = 1, \dots, n$ and uncorrelated i.e. $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. Now we make a broader assumption of unequal error variances, that is,

$$E(\epsilon) = 0, \quad Cov(\epsilon) = \sigma^2V.$$

Since σ^2V is the variance-covariance matrix of the errors, V must be positive definite (p.d.), so there exist an $n \times n$ symmetric matrix T , such that $T'T = V$ so that model (1.1) can be written as

$$T^{-1}Y = T^{-1}XB + T^{-1}\epsilon.$$

Let $Y_* = T^{-1}Y$, $X_* = T^{-1}X$ and $\epsilon_* = T^{-1}\epsilon$ then $E(\epsilon_*) = 0$ and $Cov(\epsilon_*) = \sigma^2I$. Therefore the transformed model

$$Y_* = X_*\beta + \epsilon_* \tag{1.4}$$

satisfies the assumption of error $\epsilon_* \sim N(0, \sigma^2I)$. So OLS estimator for model (1.3) is

$$\tilde{\beta} = (X_*'X_*)^{-1}X_*'Y_* = (X'V^{-1}X)^{-1}X'V^{-1}Y, \tag{1.5}$$

which is called the generalized least squares (GLS) estimator of β . The GLS estimator is the best linear unbiased estimator of β with $Cov(\tilde{\beta}) = \sigma^2(X'V^{-1}X)^{-1}$. Since the rank of X_* is equal to that of X , then the multicollinearity still also effects the GLS estimator.

Trenkler (1984) proposed the ridge estimator of β as:

$$\tilde{\beta}_k = (X'V^{-1}X + kI)^{-1}X'V^{-1}Y. \tag{1.6}$$

Because $\tilde{\beta}_k$ is a biased estimator, Özkale (2008) proposed a jackknife ridge estimator to reduce the bias of $\tilde{\beta}_k$.

Since the literature on the estimation of ridge regression parameters under correlated error is limited, the objective of the paper is to investigate the performance of some of the existing popular techniques for the estimator in (1.6) and to make a comparison among them based on mean square properties. The organization of the paper is as follows. We review some methods for estimating the ridge parameter k and consider a criterion for comparing the estimators in Section 2. Section 3 describes the Monte Carlo simulation An example has been considered in Section 4. Finally, some concluding remarks are presented in Section 5.

2 Some Ridge Regression Estimators

Hoerl and Kennard (1970a,b) showed that the estimated k which minimizes the MSE for the generalized ridge regression estimator $\hat{\beta}(k) = (X'X + K)^{-1}X'Y$ for model (1.1), where $K = diag(k_1, k_2, \dots, k_p)$, $k_i > 0$, where

$$MSE(\hat{\beta}_k) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^p \frac{k_i^2 \beta_i^2}{(\lambda_i + k_i)^2}, \tag{2.1}$$

when

$$k_i = \frac{\sigma^2}{\beta_i^2}. \tag{2.2}$$

For more information about that, we refer to Muniz and Kibria (2009). Therefore, in this section, we review some of these types of estimated ridge parameters according to model (1.3) as follows

1. The estimated k due to Hoerl and Kennard (1970a) (thereafter \hat{k}_{HK} or HK), is

$$\hat{k}_{HK} = \frac{\hat{\sigma}^2}{\hat{\beta}_{max}^2}, \tag{2.3}$$

where $\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-p}$ is the residual mean square estimate, which is an unbiased estimator of σ^2 , where σ^2 represents the error variance of model (1.3).

2. The estimated k due to Hoerl, Kennard and Baldwin (1975) (thereafter \hat{k}_{HKB} or HKB), by taking the harmonic mean of \hat{k}_i is

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \tilde{\beta}_i^2}. \quad (2.4)$$

3. From the Bayesian point of view, the estimated k due to Lawless and Wang (1976) (thereafter \hat{k}_{LW} or LW), is

$$\hat{k}_{LW} = \frac{p\hat{\sigma}^2}{\tilde{\beta}'\Lambda\tilde{\beta}}. \quad (2.5)$$

4. The estimated k due to Hocking, Speed and Lynn (1976) (thereafter \hat{k}_{HSL} or HSL), is

$$\hat{k}_{HSL} = \hat{\sigma}^2 \frac{\sum_{i=1}^p (\lambda_i \tilde{\beta}_i)^2}{(\sum_{i=1}^p \lambda_i \tilde{\beta}_i^2)^2}. \quad (2.6)$$

5. The estimated k due to Kibria (2003) (thereafter \hat{k}_{AM} or AM) by using the arithmetic mean of \hat{k}_i is

$$\hat{k}_{AM} = \frac{1}{p} \sum_{i=1}^p \frac{\hat{\sigma}^2}{\tilde{\beta}_i^2}. \quad (2.7)$$

6. The estimated k due to Kibria (2003) (thereafter \hat{k}_{GM} or GM) by using the geometric mean of \hat{k}_i is

$$\hat{k}_{GM} = \frac{\hat{\sigma}^2}{(\prod_{i=1}^p \tilde{\beta}_i^2)^{\frac{1}{p}}}. \quad (2.8)$$

7. The estimated k due to Kibria (2003) (thereafter \hat{k}_{MED} or MED) by using the median of \hat{k}_i is

$$\hat{k}_{MED} = Median \left\{ \frac{\hat{\sigma}^2}{\tilde{\beta}_i^2} \right\}, i = 1, 2, \dots, p. \quad (2.9)$$

8. The estimated k due to Khalaf and Shukur (2005) (thereafter \hat{k}_{KS} or KS) as a modification of \hat{k}_{HK} is

$$\hat{k}_{KS} = \frac{t_{max}\hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_{max}\tilde{\beta}_{max}^2}, \quad (2.10)$$

where t_{max} is the maximum eigenvalue of $X_*'X_*$ matrix

9. The estimated k due to Alkhamisi et al. (2006) by using both Khalaf and Shukur (2005) and Kibria (2003) methods are

$$\hat{k}_{KS}^{arith} = MKS = \frac{1}{p} \sum_{i=1}^p \left(\frac{t_i \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_i \tilde{\beta}_i^2} \right) \quad (2.11)$$

$$\hat{k}_{KS}^{max} = XMS = \max \left(\frac{t_i \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_i \tilde{\beta}_i^2} \right) \quad (2.12)$$

$$\hat{k}_{KS}^{med} = MEKS = \text{median} \left(\frac{t_i \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_i \tilde{\beta}_i^2} \right) \quad (2.13)$$

10. The estimated k due to Muniz and Kibria (2009) (thereafter \hat{k}_{KS}^{GM} or KM1), by using the geometric mean of $\left(\frac{t_i \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_i \tilde{\beta}_i^2} \right)$ is

$$\hat{k}_{KS}^{GM} = \left(\prod_i^p \frac{t_i \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_i \tilde{\beta}_i^2} \right)^{\frac{1}{p}} \quad (2.14)$$

11. The estimated k due to Muniz and Kibria (2009) the square root transformations are

$$\hat{k}_{KM2} = KM2 = \max \left(\frac{1}{m_i} \right), \quad (2.15)$$

$$\hat{k}_{KM3} = KM3 = \max(m_i), \quad (2.16)$$

$$\hat{k}_{KM4} = KM4 = \left(\prod_i^p \frac{1}{m_i} \right)^{\frac{1}{p}}, \quad (2.17)$$

$$\hat{k}_{KM5} = KM5 = \left(\prod_i^p m_i \right)^{\frac{1}{p}}, \quad (2.18)$$

$$\hat{k}_{KM6} = KM6 = \text{median} \left(\frac{1}{m_i} \right), \quad (2.19)$$

$$\hat{k}_{KM7} = KM7 = \text{median}(m_i), \quad (2.20)$$

where $m_i = \sqrt{\frac{\hat{\sigma}^2}{\tilde{\beta}_i^2}}$.

To make a comparison among the estimators, a criterion for measuring “goodness” of an estimator is needed. Therefore, the mean squares error (MSE) criterion is used throughout our study to measure the goodness of an estimator. Since we are interested to see the performance of the estimators under the assumption of autocorrelated error, the proposed V matrix in equation (1.4) would be as follows

$$V = \frac{1}{1 - \rho^2} = \begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix}. \quad (2.21)$$

3 The Monte Carlo Simulation

3.1 Simulation Technique

Based on Golsing and Puterman (1985) and then followed by Firinguetti (1989), this section conducts a simulation study to compare the performance of the estimated k given from equations (2.3) to (2.20) when the error terms are correlated. To achieve different degrees of collinearity, following McDonald and Galarneau (1975), Gibbons (1981) and Kibria (2003), the explanatory variables are generated by using the following equation.

$$x_{ij} = (1 - \gamma^2)^{1/2} z_{ij} + \gamma z_{ip}, i = 1, \dots, n, j = 1, 2, \dots, p, \quad (3.1)$$

where z_{ij} are independent standard normal pseudo-random numbers, p is the number of the explanatory variables, n is the sample size and γ is specified so that the correlation between any two explanatory variables is given by γ^2 . In this simulation we consider $p = 5$ and $n = 10, 20, 30, 50$ and 100 . Three different sets of correlation are considered according to the value of $\gamma = 0.85, 0.95$ and 0.99 . Also the explanatory variables are standardized so that $X'X$ will be in a correlation form.

The V matrix given in (2.21) is used in this simulation study where two values of ρ are given as $0.1, 0.9$. According to Kibria (2003), Alheety and Gore (2008) and Muniz and Kibria (2009), we consider the coefficient vector corresponding to the largest eigenvalue of $X'V^{-1}X$ matrix. The n observations for the dependent variable are determined by the following equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip} + e_i, \quad (3.2)$$

where $e_i = \rho e_{i-1} + u_i$ is the terms of the error vector from the AR(1) process (see Ozkale, 2009). The term u_i is generated from each of the following symmetric distributions: $N(0, 3^2)$ and $N(0, 10^2)$ distributions. In this study, β_0 is taken to be zero. The experiment is repeated 2000 times by generating new error terms. The *MSEs* for the estimators are calculated as follows

$$MSE(\hat{\beta}^*) = \frac{1}{2000} \sum_{r=1}^{2000} (\hat{\beta}_{(r)}^* - \beta)' (\hat{\beta}_{(r)}^* - \beta), \quad (3.3)$$

where $\hat{\beta}^*$ is the ridge estimator for different estimated value of k considered for comparison reasons.

3.2 Results Discussion

The simulated results for different methods that used to choose the ridge parameter k are presented in Tables 3.1 to 3.4. From these tables, we observed that as sample size increase the MSE decrease for all estimators. All estimators are sensitive to number of observations (n), correlation between regressors (γ) and ρ . The performance of the estimators do not vary greatly when $n = 100$ or more specifically, when the sample sizes are large. However, for $n < 100$, all proposed ridge regression estimators are performing better than the GLS estimator for all γ

and ρ . We also noticed that, the estimators LW , GM , XKS , $KM2$, $KM3$, $KM4$, $KM6$ and $KM7$ are performing better than the rest when the sample sizes are small. There is significant effect of n and γ on the performance of the estimators, GLS , HKB , HLS , AM , $MEKS$ and GKS . We also notice that on average $KM2$ and $KM3$ performed the best compared to the rest for small sample sizes.

4 Numerical example

We illustrate the results by considering the dataset discussed by Bayhan and Bayhan (1998). Tables 4.1 and 4.3 give 75 weekly observations of sales. Sixty observations in Table 4.1 are taken as historical data and fifteen observations in Table 4.2 are taken from the last 15 weeks as a fresh data. In these two Tables y_i and y_j denote weekly quantities of shampoos sold, while x_{i1}, x_{i2} and x_{j1}, x_{j2} denote the weekly list prices (averages from selected supermarkets) of the firm's shampoos and of a certain brand of soap, substituted from shampoos, respectively.

Using the data in Table 2, the matrix $X'X$ has eigenvalues $\lambda_1 = 41392.6$, $\lambda_2 = 0.5$. The condition number (CN) is $CN = 91873.2$ and it is very large. That is, this data has strong multicollinearity. The Durbin-Watson statistics $d = \frac{\sum_{i=2}^n (\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^n \hat{e}_i^2}$ can be used to detect the presence of autocorrelation, where \hat{e}_i denote the residuals of a linear regression model. If we use the data in Table 4.2 for computing d , we find that $d = 0.38$. For a significance level of 0.05 and for $n = 15$, the critical values of the Durbin-Watson statistic are $d_L = 0.95$ and $d_U = 1.54$. Since $d < d_L$, it is concluded that autocorrelation is present in this data. First we give the model for the rescaled variables:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t, \quad E(\epsilon_t) = 0, \quad E(\epsilon_t \epsilon_{t+p}) = \sigma^2 \rho_i, \quad (4.1)$$

where the term ρ_j is the j -lagged autocorrelation of the error terms and expressed as $\rho_j = \frac{C_j}{C_0}$ and $C_j = E(\epsilon_i \epsilon_{i+j})$, $j = 0, 1, \dots, 14$. To compute $\tilde{\beta}$, we need the matrix V , which is defined as in (2.21). An estimate of V requires the knowledge of the correlation structure of the residuals. Therefore, we followed Bayhan and Bayhans (1998) approach, which uses historical data to estimate V , and used $\hat{\rho}_j$,

$$\hat{\rho}_j = \frac{\hat{C}_j}{\hat{C}_0}, \quad \hat{C}_j = \frac{1}{n} \sum_{i=1}^{n-j} (\hat{e}_i - \bar{e})(\hat{e}_{i+j} - \bar{e}), \quad j = 0, \dots, 14, \quad n = 60, \quad (4.2)$$

where $\hat{e} = Y - X\hat{\beta}$ is the OLS residuals of historical data.

The estimated off diagonal elements of V matrix is 0.72284, 0.42003, 0.28663, 0.15967, 0.10987, 0.16687, 0.20766, 0.20862, 0.17573, 0.17168, 0.12975, -0.02658, -0.13842 and -0.12330. GLS of the regression coefficients are:

$$\tilde{\beta} = \begin{pmatrix} 0.8944 \\ -0.0087 \end{pmatrix}.$$

Therefore, the estimator of σ^2 is obtained as

$$\hat{\sigma}^2 = \frac{(Y - X\tilde{\beta})'(Y - X\tilde{\beta})}{n - p} = 0.28.$$

To evaluate the shrinkage parameter k , the values of the estimated ridge parameter methods given in equations (2.3) to (2.20) with their MSE are given in Table 4.3. From the table we observed that all estimators performed better than GLS in the sense of smaller MSE except the estimators KS, MKS, MEKS and XKS. However, XKS and MKS have almost the doubled MSE than the KS or MEKS. The estimators HKB, LW, GM, MED, KM3, KM7, performed equivalently well. The estimators of Muniz and Kibria (2009) performed better than others. Therefore, According to this study, if we need to fit the linear regression model that suffers for Multicollinearity and correlated error, we recommend to use in practice the estimators of ridge parameter due to Muniz and Kibria (2009).

5 Summary and Concluding Remarks

This paper considered some recent estimators for estimating the ridge parameter k when the assumption of uncorrelatedness is not satisfied. We have considered several estimators based on the work of Kibria (2003), Alkhamisi et al. (2006) and Muniz and Kibria (2009). Most of their work were based on the assumption that the error variables are uncorrelated. However, we have considered these estimators when the error of the regression model follow AR(1) process. To see the performance of the estimators, a simulation study has been made. Based on the simulation study, some of the estimated k due to Muniz and Kibria (2009) have smaller MSE than the generalized least squared estimator and some other estimators. Also in some cases the performance of these types (Muniz and Kibria (2009)) is changed when the value of ρ is changed from low to high. Estimators $KM2$ and $KM3$ performed the best compared to the rest when sample size is small. To illustrate the findings of the paper, a real life data has been analyzed.

Acknowledgments

This paper was written while the second author was on sabbatical leave (2010-2011). He is grateful to Florida International University for awarding him the sabbatical leave which, gave him excellent research facilities.

References

- Alheety M. I. and Gore, S. D. (2008). A new estimator in multiple linear regression , *Model Assisted Statistics and Applications*, 3(3), 187-200.
- Alheety, M. I. and Ramanathan, T. V. (2009). Confidence Interval for Shrinkage Parameters in Ridge Regression, *Communications in Statistics - Theory and Methods*, 38(19), 3489-3497.
- Alkhamisi, M. and Shukur, G. (2008). Developing ridge parameters for SUR model. *Communications in Statistics - Theory and Methods*, 38(19), 3489-3497.
- Alkhamisi, M., Khalaf, G. and Shukur, G.(2006). Some modifications for choosing ridge parameters. *Communications in Statistics-Theory and Methods*, 35, 2005-2020.
- Bayhan, G.M. and Bayhan M. (1998). Forecasting using autocorrelated errors and multicollinear predictor variables. *Computers and Industrial Engineering*, 34, 413-421.

- Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*, Second Edition, Wiley, New York.
- Dempster, A. P., Schatzoff, M. and Wermuth, N.(1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association*, 72, 77-91.
- Firinguetti, L. (1989). A simulation study of ridge regression estimators with autocorrelated errors. *Communications in Statistics - simulation* , 18(2), 673-702.
- Gibbons, D. G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association*, 76, 131-139.
- Gosling, B.J. and M.L. Puterman, "Ridge Estimation in Regression Problems with Auto-correlated Errors: A Monte Carlo Study. *Communications in Statistics-Simulation and Computation*, 14 (3), 577-614.
- Golub, G. H., Health, M., Wahaba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21, 215-223.
- Gunst, R. F.; Mason R. L. (1977). Biased estimation in regression: An evaluation using mean squared error. *Journal of the American Statistical Association*, 72,616-628.
- Haq, M. S.; Kibria, B. M. G.(1996). A shrinkage estimator for the restricted linear regression model: Ridge regression approach. *Journal of Applied Statistical Science*, 3(4), 301-316.
- Hemmerle, W. J.; Brantle, T. F.(1978). Explicit and Constraint generalized ridge estimator. *Technometrics*, 2, 109-120.
- Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: biased estimation for non- orthogonal Problems. *Technometrics*, 12, 55-67.
- Hoerl, A. E. and Kennard, R. W. (1970b). Ridge Regression: Application to Non-Orthogonal Problems. *Technometrics*, 12, 69-82.
- Hoerl, A. E., Kennard, R. W. and Baldwin, K. F. (1975). Ridge regression: Some simulation. *Communications in Statistics*, 4, 105-123.
- Hocking,R. R., Speed, F. M. and Lynn, M. J. (1976). A class of biased estimators in linear regression. *Technometrics*, 18, 425-438.
- Khalaf, G., Shukur, G. (2005). Choosing ridge parameters for regression problems. *Communications in Statistics - Theory and Methods* A34:1177-1182.
- Kibria, B. M. G. (1996). On preliminary test ridge regression estimators for linear restriction in a regression model with non-normal disturbances. *Communications in Statistics*, A 25, 2349-2369.
- Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, 32, 419-435.

- Lawless, J. F.(1978). Ridge and related estimation procedure. *Communications in Statistics*, A 7, 139-164.
- Lawless, J. F. and Wang, P.(1976). A simulation study of ridge and other regression estimators. *Communications in Statistics*, A 5, 307-323.
- Mansson, K., Shukur, G and Kibria, B. M. G. (2010). A Simulation Study of some Ridge Regression Estimators under Different Distributional Assumptions. *Communications in Statistics-Simulation and Computation*, 39(8), 1639-1670.
- 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.
- Muniz, G. and Kibria, B. M. Golam (2009). On some ridge regression estimators: An empirical comparisons, *Communications in Statistics-Simulation and Computation*, 38, 621-630.
- Özkale, M. R. (2008). A jackknifed ridge estimator in the linear regression model with heteroscedastic or correlated errors. *Statistics and Probability Letters*, 78(18), 3159-3169.
- Saleh, A. K. Md. and Kibria, B. M. G. (1993). Performances of some new preliminary test ridge regression estimators and their properties. *Communications in Statistics-Theory and Methods*, 22, 2747-2764.
- Singh, S. and Tracy, D. S. (1999). Ridge-regression using scrambled responses. *Metrika*, 147-157.
- Trenkler, G. (1984). On the performance of biased estimators in the linear regression model with correlated or heteroscedastic errors. *Journal of Econometrics*, 25, 179-190.
- Wichern, D. W.; Churchill, G. A. (1978). A comparison of ridge estimators. *Technometrics*, 20, 301-311.
- Wencheke, E. (2000). Estimation of the signal-to-noise in the linear regression model, *Statistical Papers*, 41, 327-343.

Table 4.1: Historical data for weekly sales of shampoos and prices

Obs.	y_i1	x_i1	x_i2	Obs.	y_i1	x_i1	x_i2
1	28.445	49	12.5	31	31.446	84.4	21.2
2	28.547	49	12.5	32	31.549	85	21.2
3	28.644	51.2	13	33	31.641	85	21.2
4	28.746	51.2	13	34	31.743	78	20.1
5	28.849	40.3	13	35	31.848	78	20.1
6	28.94	52	13	36	31.94	81.3	20.1
7	29.045	52.3	13.8	37	32.043	83.1	21
8	29.142	58	14.4	38	32.146	83	21
9	29.248	58	14.4	39	32.25	88.5	22.3
10	29.25	58	14.4	40	32.344	88.5	22.3
11	29.443	62	16	41	32.441	88.5	22.3
12	29.545	62	16	42	32.545	68.7	22.9
13	29.644	62	16	43	32.643	68.7	22.9
14	29.747	52	17.1	44	32.748	91.3	22.9
15	29.841	67.2	17.1	45	32.842	91.3	22.9
16	29.045	67.2	17.1	46	32.95	91.3	23
17	30.046	67.2	18	47	33.039	92.8	23
18	30.142	67.2	18	48	33.144	92.8	23
19	30.245	72.4	18	49	33.249	76	25.4
20	30.348	72.4	18	50	33.347	76	26
21	30.441	72.4	18	51	33.442	93.4	24.1
22	30.549	72.4	21	52	33.543	93.4	24.1
23	30.641	80	21	53	33.647	93.4	24.1
24	30.739	72	18.3	54	33.746	96.3	24.1
25	30.849	72	18.3	55	33.849	96.3	24.3
26	30.949	55	19	56	33.94	97.2	24.3
27	31.051	48	19	57	34.041	97.2	24.3
28	31.148	80.1	19.4	58	34.143	75.2	25.1
29	31.245	80.1	21.2	59	34.248	100	25.1
30	31.342	84.4	21.2	60	34.345	101.5	25.4

Table 4.2: Fresh data for weekly sales of shampoos and prices

Obs.	y_{j1}	x_{j1}	x_{j2}
1	34.481	101.3	25.3
2	34.369	102	25.5
3	34.268	102.7	25.7
4	34.16	102.35	25.9
5	34.215	104.2	26.1
6	34.308	104.9	26.2
7	34.402	105.6	26.4
8	34.479	106.9	26.6
9	34.58	107	26.8
10	34.682	107.7	27
11	34.78	108.5	27.1
12	34.875	109.1	27.3
13	34.963	109.9	27.5
14	35.054	110.6	27.7
15	35.173	111.3	27.8

Table 4.3: The values of estimated ridge parameter with MSE of ridge estimator in (5) for each value

EST.	\hat{k}	MSE
GLS	0	11.48
HK	0.3500	2.24
HKB	0.7000	0.81
LW	0.7060	0.8
HSL	0.3500	2.24
AM	0.1750	7.27
GM	35.9838	0.76
MED	0.0018	0.79
KS	0.0554	43.5
MKS	0.0287	97.14
MEKS	0.0554	43.5
XKS	0.0287	97.14
GKS	0.010	2.32
KM2	1.6903	0.49
KM3	60.8219	0.77
KM4	0.1667	7.89
KM5	5.9986	0.62
KM6	0.8533	0.65
KM7	30.7	0.75

Table 3.1: Estimated MSE for different methods for estimating the ridge parameter with $N(0, 3^2)$ and $\rho = 0.1$

GLS	HK	HKB	LW	HSL	AM	GM	MED	KS	MKS	XKS	MEKS	GKS	KM2	KM3	KM4	KM5	KM6	KM7
6.1048	2.7451	2.505	1.1108	2.1614	5.3638	1.1502	$N(0, 3^2)$, n=10	2.7993	1.3629	$\gamma = 0.85$ 1.0078	2.1569	2.15	0.931	0.925	0.9263	0.9875	0.9246	1.0084
6.2444	3.2299	2.8475	1.0614	2.6863	5.6012	1.3953	$N(0, 3^2)$, n=10	3.2803	1.5434	$\gamma = 0.95$ 0.9031	4.1703	3.486	0.8909	0.8567	0.8476	0.9477	0.85	1.0193
26.07	12.22	10.61	1.033	6.004	23.10	3.136	$N(0, 3^2)$, n=10	6.0124	1.3466	$\gamma = 0.99$ 0.8885	15.841	11.508	0.9459	0.8778	0.8978	1.0325	0.9045	1.2693
1.9426	1.2166	1.1383	0.8934	1.2145	1.7808	0.9123	$N(0, 3^2)$, n=20	1.3108	1.361	$\gamma = 0.85$ 0.9919	1.7083	1.6291	0.8819	0.8798	0.8548	0.8664	0.8558	0.8774
4.2511	2.0958	1.7771	0.859	1.4655	3.7443	1.0228	$N(0, 3^2)$, n=20	2.1749	1.5541	$\gamma = 0.95$ 0.9102	3.4973	3.1531	0.8995	0.8542	0.847	0.8578	0.851	0.8892
32.506	13.390	10.231	0.8011	2.6569	28.082	1.9779	$N(0, 3^2)$, n=20	13.456	1.1623	$\gamma = 0.99$ 0.7944	25.833	18.728	0.9416	0.8167	0.8673	0.8226	0.88	0.9006
1.2694	0.9907	0.9568	0.8843	1.0295	1.2026	0.8916	$N(0, 3^2)$, n=30	1.0787	1.1486	$\gamma = 0.85$ 1.0052	1.2123	1.199	0.8858	0.9018	0.8689	0.8724	0.8686	0.8753
2.3672	1.3975	1.26	0.8718	1.1884	2.1448	0.9386	$N(0, 3^2)$, n=30	1.4931	1.5016	$\gamma = 0.95$ 0.994	2.163	2.0516	0.9029	0.8808	0.8619	0.868	0.8648	0.8822
13.654	5.7867	4.5239	0.8132	1.4338	11.868	1.3374	$N(0, 3^2)$, n=30	5.8681	1.4799	$\gamma = 0.99$ 0.8356	11.9815	10.1602	0.9354	0.8307	0.869	0.825	0.8804	0.8667
1.1123	0.9305	0.9031	0.853	0.9311	1.0687	0.867	$N(0, 3^2)$, n=50	1.0041	1.0587	$\gamma = 0.85$ 0.9647	1.0954	1.087	0.8746	0.8856	0.8454	0.8495	0.8482	0.8505
1.788	1.1948	1.0952	0.8526	1.0003	1.6509	0.8993	$N(0, 3^2)$, n=50	1.2911	1.3977	$\gamma = 0.95$ 1.0017	1.7277	1.6861	0.9003	0.8718	0.8568	0.8494	0.8604	0.8572
6.2136	2.8226	2.3219	0.8077	1.0163	5.444	1.0917	$N(0, 3^2)$, n=50	2.921	1.7206	$\gamma = 0.99$ 0.8741	5.9203	5.4718	0.9293	0.8289	0.8669	0.8114	0.8752	0.8313
0.9306	0.8722	0.8627	0.8495	0.872	0.9147	0.864	$N(0, 3^2)$, n=100	0.9094	0.9227	$\gamma = 0.85$ 0.9053	0.9276	0.9264	0.881	0.8966	0.8515	0.8542	0.8557	0.8513
1.2861	0.9609	0.9196	0.8067	0.8326	1.2109	0.8374	$N(0, 3^2)$, n=100	1.056	1.1649	$\gamma = 0.95$ 0.9585	1.2737	1.2616	0.8845	0.8394	0.8347	0.8065	0.8387	0.8079
2.6516	1.4651	1.3136	0.8305	0.8855	2.3887	0.9292	$N(0, 3^2)$, n=100	1.5777	1.6237	$\gamma = 0.99$ 0.9732	2.599	2.5311	0.9303	0.8524	0.8772	0.8309	0.8818	0.8371

Table 3.2: Estimated MSE for different methods for estimating the ridge parameter with $N(0, 10^2)$ and $\rho = 0.1$

GLS	HK	HKB	DW	HSL	AM	GM	MED	KS	MKS	XKS	MEKS	GKS	KM2	KM3	KM4	KM5	KM6	KM7
6.3488	3.1757	2.9319	1.2997	2.9725	5.6747	1.4719	$N(0, 10^2)$, 1.765	n=10 3.24	1.6948	1.1338	3.6186	3.159	1.005	1.0206	1.0401	1.1052	1.038	1.2387
										$\gamma = 0.85$								
9.402	4.4542	4.166	1.2866	3.7858	8.398	1.7692	$N(0, 10^2)$, 2.316	n=10 4.5116	1.7038	1.0933	5.2541	4.5101	0.9954	1.0089	1.0146	1.1792	1.0116	1.2911
										$\gamma = 0.95$								
166.24	71.025	60.911	1.0949	13.823	144.96	7.1894	$N(0, 10^2)$, 20.769	n=10 71.064	1.1565	1.0047	80.352	37.475	0.9957	0.9952	0.9922	1.1778	0.9933	1.5574
										$\gamma = 0.99$								
2.1205	1.4201	1.3239	1.0517	1.516	1.9664	1.069	$N(0, 10^2)$, 1.1485	n=20 1.5281	1.5933	1.1729	1.962	1.8735	0.991	0.99	1.016	1.0294	1.0111	1.0537
										$\gamma = 0.85$								
4.0072	2.0906	1.9015	1.0393	1.9397	3.5981	1.1494	$N(0, 10^2)$, 1.3818	n=20 2.1863	1.7474	1.1022	3.5438	3.1525	0.9924	0.9917	1.0007	1.0369	0.9981	1.0793
										$\gamma = 0.95$								
8.7362	3.9438	3.3449	1.0147	2.7328	7.6799	1.3362	$N(0, 10^2)$, 2.0656	n=20 4.03	1.7303	1.051	7.6329	6.4435	0.9945	0.9928	0.9938	1.0365	0.9937	1.0974
										$\gamma = 0.99$								
1.7119	1.2652	1.1948	1.0278	1.3062	1.6078	1.0347	$N(0, 10^2)$, 1.0795	n=30 1.3742	1.4625	1.191	1.6177	1.5961	0.9915	0.9924	1.0071	1.0136	1.0025	1.0278
										$\gamma = 0.85$								
2.7015	1.6484	1.4822	1.0224	1.5509	2.4574	1.0841	$N(0, 10^2)$, 1.2115	n=30 1.7489	1.7048	1.1495	2.5267	2.3833	0.9933	0.9962	0.9992	1.0224	0.9981	1.0462
										$\gamma = 0.95$								
6.9512	3.2055	2.7575	0.9994	2.1988	6.1191	1.2403	$N(0, 10^2)$, 1.8484	n=30 3.3026	1.7939	1.0621	6.3579	5.6368	0.9938	0.988	0.9916	1.0173	0.9921	1.0641
										$\gamma = 0.99$								
1.3418	1.1168	1.0869	1.009	1.1611	1.2929	1.0094	$N(0, 10^2)$, 1.0316	n=50 1.2154	1.2782	1.1643	1.3243	1.3133	0.9886	0.9911	0.999	0.9998	0.9971	1.007
										$\gamma = 0.85$								
1.8708	1.3175	1.2254	0.9995	1.2718	1.7406	1.0306	$N(0, 10^2)$, 1.101	n=50 1.4268	1.5422	1.1715	1.8284	1.7873	0.9919	0.9917	0.992	1.0015	0.9915	1.0146
										$\gamma = 0.95$								
7.6885	3.628	2.9095	0.9817	2.0709	6.7435	1.1751	$N(0, 10^2)$, 1.8947	n=50 3.7288	1.8196	1.0493	7.3906	6.7182	0.9916	0.9812	0.9836	0.9911	0.9847	1.0228
										$\gamma = 0.99$								
1.1063	1.0302	1.02	0.999	1.0495	1.0888	0.9968	$N(0, 10^2)$, 1.0026	n=100 1.085	1.0983	1.0808	1.1032	1.1019	0.9922	0.9924	0.9982	0.9949	0.9968	0.9972
										$\gamma = 0.85$								
1.4427	1.1568	1.1125	0.9934	1.1396	1.3769	1.0102	$N(0, 10^2)$, 1.0486	n=100 1.2601	1.3495	1.1751	1.4326	1.4214	0.9915	0.9901	0.9896	0.9939	0.9893	1.0011
										$\gamma = 0.95$								
2.7729	1.6893	1.471	0.9864	1.3094	2.5146	1.044	$N(0, 10^2)$, 1.2358	n=100 1.7996	1.7886	1.1374	2.735	2.6658	0.9924	0.9871	0.9878	0.9899	0.9893	1.0011
										$\gamma = 0.99$								

Table 3.3: Estimated MSE for different methods for estimating the ridge parameter with $N(0, 3^2)$ and $\rho = 0.9$

GLS	HK	HKB	DW	HSL	AM	GM	MED	KS	MKS	XKS	MEKS	GKS	KM2	KM3	KM4	KM5	KM6	KM7	
							$N(0, 3^2)$,	$n=10$		$\gamma = 0.85$									
2.7305	1.9755	1.9791	2.0507	1.8572	2.6298	1.3588	1.4225	1.9889	1.3205	1.0041	1.7877	1.8243	0.9132	0.9521	1.1923	1.2543	1.1738	1.3269	1.3269
							$N(0, 3^2)$,	$n=10$		$\gamma = 0.95$									
5.3461	3.5462	3.5152	2.9288	2.9786	5.1098	1.7938	2.0324	3.5664	1.5455	1.039	3.4259	3.0109	0.9172	0.979	1.2316	1.4561	1.2089	1.0119	1.0119
							$N(0, 3^2)$,	$n=10$		$\gamma = 0.99$									
50.147	27.471	29.399	5.8652	6.7689	17.434	7.0376	10.931	16.480	1.4749	0.9129	17.815	13.636	0.8582	0.8909	0.8799	2.3291	0.8741	3.5371	3.5371
							$N(0, 3^2)$,	$n=20$		$\gamma = 0.85$									
1.5454	1.1816	1.2085	1.2275	1.1449	1.4982	0.9778	0.9797	1.2024	1.1003	0.934	1.2946	1.2833	0.8826	0.8938	0.9858	0.96	0.9881	0.9834	0.9834
							$N(0, 3^2)$,	$n=20$		$\gamma = 0.95$									
1.9702	1.4684	1.4411	1.388	1.2051	1.8972	1.0612	1.1226	1.4806	1.1282	0.9235	1.6707	1.5704	0.8995	0.9065	1.0235	1.022	1.0098	1.0654	1.0654
							$N(0, 3^2)$,	$n=20$		$\gamma = 0.99$									
4.2255	2.7276	2.6697	1.664	1.6482	4.0098	1.4216	1.7989	2.7451	1.2471	0.9125	3.6624	3.1808	0.8837	0.8948	0.9266	1.0998	0.9169	1.3612	1.3612
							$N(0, 3^2)$,	$n=30$		$\gamma = 0.85$									
1.1085	0.9837	0.9901	1.0232	0.9916	1.0907	0.9253	0.9315	1.0125	1.0132	0.9368	1.0743	1.0647	0.8983	0.903	0.9559	0.9233	0.9527	0.9282	0.9282
							$N(0, 3^2)$,	$n=30$		$\gamma = 0.95$									
1.3713	1.1405	1.1298	1.1328	1.0656	1.3367	0.9806	1.0078	1.1573	1.0995	0.9457	1.295	1.2762	0.9116	0.9172	0.9831	0.9667	0.977	0.9543	0.9543
							$N(0, 3^2)$,	$n=30$		$\gamma = 0.99$									
2.8895	1.97	1.9173	1.3446	1.1374	2.7515	1.205	1.4254	1.9886	1.229	0.8962	2.7048	2.4877	0.8659	0.8734	0.8775	0.9862	0.875	1.0863	1.0863
							$N(0, 3^2)$,	$n=50$		$\gamma = 0.85$									
0.9579	0.9249	0.9195	0.9362	0.9183	0.9511	0.9104	0.9093	0.9331	0.9412	0.9183	0.9519	0.9499	0.9039	0.912	0.9229	0.9069	0.9218	0.9113	0.9113
							$N(0, 3^2)$,	$n=50$		$\gamma = 0.95$									
1.0493	0.9776	0.9638	0.9832	0.9491	1.0357	0.9275	0.9368	0.9919	1.0046	0.9407	1.0405	1.0352	0.9045	0.9114	0.9347	0.9232	0.931	0.9267	0.9267
							$N(0, 3^2)$,	$n=50$		$\gamma = 0.99$									
2.2254	1.5691	1.4729	1.1051	0.8936	2.1103	1.0526	1.1902	1.5805	1.067	0.8803	2.157	2.0375	0.8713	0.8725	0.8682	0.917	0.8692	0.9423	0.9423
							$N(0, 3^2)$,	$n=100$		$\gamma = 0.85$									
0.9149	0.9091	0.909	0.9118	0.9081	0.9133	0.9143	0.9116	0.9116	0.9136	0.9106	0.9145	0.9143	0.905	0.9189	0.9088	0.9066	0.9083	0.9065	0.9065
							$N(0, 3^2)$,	$n=100$		$\gamma = 0.95$									
0.9541	0.927	0.9229	0.9336	0.9159	0.9483	0.9166	0.9167	0.9361	0.9451	0.9275	0.9525	0.9515	0.9091	0.9176	0.9185	0.9135	0.9179	0.914	0.914
							$N(0, 3^2)$,	$n=100$		$\gamma = 0.99$									
1.1959	1.0302	1.0035	0.9634	0.9039	1.1623	0.9313	0.9523	1.0432	1.0267	0.9196	1.1883	1.1767	0.9005	0.9038	0.8997	0.9102	0.9002	0.9159	0.9159
							$N(0, 3^2)$,	$n=100$		$\gamma = 0.99$									

Table 3.4: Estimated MSE for different methods for estimating the ridge parameter with $N(0, 10^2)$ and $\rho = 0.9$

GLS	HK	HKB	LW	HSL	AM	GM	MED	KS	MKS	XKS	MEKS	GKS	KM2	KM3	KM4	KM5	KM6	KM7
5.3564	3.222	3.5215	2.9155	3.2418	5.111	1.7746	$N(0, 10^2)$, n=10	1.9775	1.594	1.1664	2.7742	2.728	1.0848	1.1146	1.4548	1.5683	1.4174	1.6593
							$N(0, 10^2)$, n=10											
9.6328	4.7548	5.5765	2.9782	3.9361	9.073	1.7678	1.9306	4.7888	1.5594	1.1622	2.6352	2.7786	1.0539	1.1002	1.3564	1.5052	1.3372	1.5695
							$N(0, 10^2)$, n=10											
30.178	17.185	19.301	5.9407	9.3223	28.768	5.6475	8.5981	17.200	1.5245	1.0618	13.389	11.266	1.0052	1.0548	1.1068	2.2968	1.075	3.0292
							$N(0, 10^2)$, n=20											
1.48	1.2633	1.2656	1.3219	1.2947	1.4521	1.0877	1.1299	1.2824	1.2048	1.0487	1.3981	1.36	1.0492	1.0103	1.1979	1.105	1.1743	1.129
							$N(0, 10^2)$, n=20											
4.007	2.5731	2.5838	1.711	2.0964	3.8115	1.3867	1.6532	2.5988	1.4038	1.0478	3.1141	2.8067	0.9971	1.0053	1.0897	1.2063	1.0634	1.3101
							$N(0, 10^2)$, n=20											
9.6949	5.5075	5.5035	1.7851	2.7414	9.115	1.7941	2.6705	5.5229	1.1909	1.0035	6.8458	5.2488	0.9914	0.9954	1.0286	1.2521	1.0209	1.4686
							$N(0, 10^2)$, n=30											
1.0969	1.0501	1.0468	1.0748	1.072	1.0895	1.0185	1.0261	1.073	1.08	1.0565	1.0871	1.0854	1.0379	1.0074	1.0684	1.0311	1.0646	1.0356
							$N(0, 10^2)$, n=30											
1.5264	1.283	1.2639	1.2302	1.2266	1.4893	1.0736	1.1442	1.2969	1.1666	1.019	1.4705	1.4289	1.0069	0.9959	1.0983	1.0652	1.072	1.1011
							$N(0, 10^2)$, n=30											
3.8026	2.448	2.3696	1.3101	1.5543	3.5928	1.2538	1.5879	2.4603	1.15	1.0037	3.313	2.9245	0.9933	0.995	1.0255	1.0968	1.015	1.1895
							$N(0, 10^2)$, n=50											
1.0501	1.0193	1.0162	1.0307	1.0267	1.0447	0.9995	1.0041	1.0318	1.0389	1.0198	1.0469	1.045	1.0113	0.9951	1.0318	1.0056	1.0289	1.0086
							$N(0, 10^2)$, n=50											
1.1578	1.0687	1.0632	1.0649	1.0676	1.1429	1.0109	1.0304	1.0839	1.0918	1.0227	1.1482	1.141	1.0023	0.9931	1.0453	1.0145	1.0341	1.0253
							$N(0, 10^2)$, n=50											
2.2353	1.6006	1.5553	1.1054	1.2867	2.1315	1.0932	1.2816	1.6189	1.2033	1.0123	2.1705	2.0589	0.992	0.9925	1.0064	1.0311	1.0006	1.0795
							$N(0, 10^2)$, n=100											
1.0156	1.0007	0.9986	1.0043	1.0035	1.0127	0.992	0.9944	1.009	1.0128	1.0063	1.0149	1.0145	0.9962	0.9907	1.0061	0.9936	1.0043	0.9953
							$N(0, 10^2)$, n=100											
1.0649	1.0241	1.0191	1.0179	1.0222	1.0572	0.9976	1.0052	1.0379	1.0491	1.0202	1.0631	1.0613	0.9954	0.9921	1.0142	0.9992	1.0097	1.0033
							$N(0, 10^2)$, n=100											
1.3603	1.1584	1.1326	1.0196	1.0748	1.3213	1.013	1.0673	1.1722	1.123	1.0114	1.3506	1.3354	0.9914	0.9917	0.9985	1.0021	0.9953	1.0175
							$N(0, 10^2)$, n=100											