A Poisson Ridge Regression Estimator

By

Kristofer Månsson[1] and Ghazi Shukur[1,2]

[1] *Department of Economics and Statistics, Jönköping University, Sweden,*

[2] *Department of Economics and Statistics, Linnaeus University, Sweden.*

Abstract

The standard statistical method for analyzing count data is the Poisson regression model, which is usually estimated using maximum likelihood (ML). The ML method is very sensitive to multicollinearity. Therefore, we present a new Poisson ridge regression estimator (PRR) as a remedy to the problem of instability of the traditional ML method. To investigate the performance of the PRR and the traditional ML approaches for estimating the parameters of the Poisson regression model, we calculate the mean squared error (MSE) using Monte Carlo simulations. The result from the simulation study shows that the PRR method outperforms the traditional ML estimator in all of the different situations evaluated in this paper.

Keywords: Poisson regression; maximum likelihood; ridge regression; MSE; Monte Carlo simulations; Multicollinearity

Mathematics Subject Classification: Primary 62J07; Secondary 62J02

1. Introduction

In multiple regression analysis, it is usually impossible to interpret the estimates of the individual coefficients if the explanatory variables are highly inter-correlated. Such a problem is often referred to as the multicollinearity problem. In the literature there exist several ways to "solve" this problem. One such way is the ridge regression, about which a great number of studies are conducted. Most of the efforts in these studies concentrate on estimating the shrinkage ridge parameter (k) in different ways and compare it to the least squares estimator (LSE). This parameter was originally introduced by Hoerl and Kennard (1970a), who used ridge regression estimators to tackle the multicollinearity problem. They suggested a small positive number ($k \geq 0$) to be added to the diagonal elements of the $X'X$ matrix from the multiple regression, and the resulting estimators are obtained as

$$\hat{\beta}_{RR} = \left(X'X + kI_p \right)^{-1} X'y, \quad k \geq 0, \tag{1.2}$$

which is known as a ridge regression RR estimator. Where, X is an $n \times (p+1)$ observed matrix of the regressors, β is an $(p+1) \times 1$ vector of unknown parameters. For a positive value of *k*, this estimator provides a smaller mean squared error (MSE) compared to LSE.

Most of the later efforts in this area have concentrated on estimating the value of the ridge parameter k. Many different techniques for estimating k have been proposed or suggested by different researchers. To mention a few, Hoerl and Kennard (1970a,b), Hoerl et al. (1975), McDonald and Galarneau (1975), Lawless and Wang (1976), Dempster et al. (1977), Gibbons (1981), Kibria (2003), Khalaf and Shukur (2005), Alkhamisi et al. (2006), Alkhamisi and Shukur (2008) and Muniz and Kibria (2009). In these and other research, the performance of the ridge estimators was mainly compared based on simulation studies. Most of the researchers have generated data from normal or non-normal populations, with a given number of regressors and used MSE as a performance criterion.

In almost all situations where regression analysis is done the observations are assumed to be identically and independently distributed (iid). However, we also know that in the real-life context the iid assumption is too strong. As an example, the mean rate of occurrence of an event vary from case to case might depend on some variables. Count data regression is more proper than the OLS in studying the occurrence rate per unit of time conditional on some covariates. Examples of such situations include number of patents, takeover bids, bank failures, accident insurance, and criminal careers. Unless the mean of the counts is high (in which case the normal approximation and the OLS method may be satisfactory), using the OLS can lead to significant deficiencies. In such situations, the benchmark model for count data is the Poisson regression model.

The purpose of this study is to adopt and modify the new approaches mentioned in Kibria (2003), Khalaf and Shukur (2005), Alkhamisi et al. (2006), Alkhamisi and Shukur (2008) and very recently Muniz and Kibria (2009) to be applicable in Poisson regressions for count data, i.e. Poisson ridge regression (PRR). The performance of these parameters will then be compared with the traditional ML estimation method in term of MSE. This will mainly be done by means of Monte Carlo simulations under conditions where the sample size and the strength of correlations between the explanatory variables are varied.

The paper is organised as follows: in Section 2, we present the methodology of the different methods for estimating PRR. In Section 3, we illustrate the Monte Carlo design we use in this study. The results of the study are discussed in Section 4. In Section 5 we give a brief summary and conclusions.

2. Methodology

This section starts by defining the Poisson regression model and the traditional ML estimation method. Then the PRR estimator is derived using the same approach as in Hoerl and Kennard (1970a,b) and Schaeffer et al. (1984). Finally we generalize different methods of estimating the ridge parameter *k* that have been proposed in papers by Hoerl and Kennard (1970a,b), Kibria (2003), Alkhamisi et al. (2006) and Muniz and Kibria (2009).

2.1 Poisson regression

The standard statistical method for analyzing count data is the Poisson regression model. This model has found a widespread use in microeconometrics when the dependent variable, y_i, of the regression model is $Po(\mu_i)$ distributed where $\mu_i = \exp(\mathbf{x_i}\boldsymbol{\beta})$, $\mathbf{x_i}$ is the i*th* row of $\mathbf{X}$ which is an $n\times(p+1)$ data matrix with p explanatory variables and $\boldsymbol{\beta}$ is a $(p+1)\times 1$ vector of coefficients. The log likelihood of this model may be written as:

$$
\begin{aligned}
l(\boldsymbol{\mu};\mathbf{y}) = \sum_{i=1}^{n}\mu_i + \sum_{i=1}^{n} y_i \log(\mu_i) + \log\left(\prod_{i=1}^{n} y_i!\right) = \\
\sum_{i=1}^{n}\exp(\mathbf{x_i}\boldsymbol{\beta}) + \sum_{i=1}^{n} y_i \log\left(\exp(\mathbf{x_i}\boldsymbol{\beta})\right) + \log\left(\prod_{i=1}^{n} y_i!\right)
\end{aligned}
\qquad (2.1)
$$

The commonly applied estimation method for the Poisson regression model is ML. The parameters using this method are estimated by solving the following equation:

$$S(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\mu};\mathbf{y})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\left(y_i - \exp(\mathbf{x_i}\boldsymbol{\beta})\right)\mathbf{x_i} = 0. \qquad (2.2)$$

Since equation (2.2) is nonlinear in $\boldsymbol{\beta}$ the solution of $S(\boldsymbol{\beta})$ equalling zero is found using the following iterative weighted least square (IWLS) algorithm:

$$\boldsymbol{\beta}_{ML} = \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\hat{\mathbf{W}}\hat{\mathbf{z}}\right), \qquad (2.3)$$

where $\hat{\mathbf{W}} = diag\left[\hat{\mu}_i\right]$ and $\hat{\mathbf{z}}$ is a vector where the ith element equals $\hat{z}_i = \log(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$.
The ML estimator is asymptotically normally distributed with a covariance matrix that corresponds to the inverse of the matrix of the second derivatives:

$$Cov(\boldsymbol{\beta}_{ML}) = \left[E\left(-\frac{\partial^2 l(\mathbf{X};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)\right]^{-1} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}. \qquad (2.4)$$

Furthermore, the asymptotic MSE equals:

$$E\left(L_{ML}^2\right) = E(\boldsymbol{\beta}_{ML} - \boldsymbol{\beta})'(\boldsymbol{\beta}_{ML} - \boldsymbol{\beta}) = tr\left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1} = \sum_{j=1}^{J}\frac{1}{\lambda_j}, \qquad (2.5)$$

where λ_j is the jth eigenvalue of the $\mathbf{X}'\mathbf{WX}$ matrix. When the explanatory variables are highly correlated the weighted matrix of cross-products, $\mathbf{X'WX}$, is ill-conditioned which leads to instability and high variance of the ML estimator. In that situation, it is very hard to interpret the estimated parameters since the vector of estimated coefficients is on average too long.

2.2 The Poisson Ridge Regression Estimator

As a remedy to the problem caused by multicollinearity we propose the PRR method applied to count data. The derivation of this new method starts by noting that the ML method approximately minimizes the weighted sum of squared error (WSSE). Hence, $\boldsymbol{\beta}_{ML}$ can be seen as the optimal estimator in a WSSE sense. If we choose another estimator, $\hat{\mathbf{B}}$, of the parameter vector $\boldsymbol{\beta}$ we can write the WSSE of this estimator as

$$\phi = \left(\mathbf{y}-\hat{\mathbf{B}}\right)'\left(\mathbf{y}-\hat{\mathbf{B}}\right) = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_{ML})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_{ML}) + \left(\hat{\mathbf{B}}-\boldsymbol{\beta}_{ML}\right)'\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\left(\hat{\mathbf{B}}-\boldsymbol{\beta}_{ML}\right) = \phi_{\min} + \phi\left(\hat{\mathbf{B}}\right), \qquad (2.6)$$

where $\phi\left(\hat{\mathbf{B}}\right)$ is the increase of the WSSE when $\boldsymbol{\beta}_{ML}$ is replaced by $\hat{\mathbf{B}}$. To find the PRR estimator the length of $\hat{\mathbf{B}}'\hat{\mathbf{B}}$ should be minimized subject to the constraint $\phi\left(\hat{\mathbf{B}}\right) = \phi_0$. As a Lagrangian problem this may be stated as:

$$\text{Minimize } F = \hat{\mathbf{B}}'\hat{\mathbf{B}} + (1/k)\left(\hat{\mathbf{B}}-\boldsymbol{\beta}_{ML}\right)'\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\left(\hat{\mathbf{B}}-\boldsymbol{\beta}_{ML}-\phi_0\right), \qquad (2.8)$$

where $(1/k)$ is the Lagrange multiplier. Differentiating the above expression with respect to $\hat{\mathbf{B}}$ and setting the result equal to zero yields:

$$\frac{\delta F}{\delta \hat{\mathbf{B}}} = 2\hat{\mathbf{B}} + (1/k)\left(2\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\hat{\mathbf{B}} - 2\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\boldsymbol{\beta}_{ML}\right) = 0.$$

By solving the above equation with respect to $\hat{\mathbf{B}}$ we obtain the PRR estimator:

$$\boldsymbol{\beta}_{RR} = \left(\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X} + k\mathbf{I}\right)^{-1} \left(\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X} \boldsymbol{\beta}_{ML}\right) = \mathbf{Z}\boldsymbol{\beta}_{ML} . \tag{2.8}$$

The asymptotic MSE of this new estimator equals:

$$\begin{aligned} & E\left(L_{RR}^2\right) = E\left(\boldsymbol{\beta}_{RR} - \boldsymbol{\beta}\right)'\left(\boldsymbol{\beta}_{RR} - \boldsymbol{\beta}\right) = E\left[\left(\boldsymbol{\beta}_{ML} - \boldsymbol{\beta}\right)'\mathbf{Z}'\mathbf{Z}\left(\boldsymbol{\beta}_{ML} - \boldsymbol{\beta}\right)\right] + \left(\mathbf{Z}\boldsymbol{\beta} - \boldsymbol{\beta}\right)'\left(\mathbf{Z}\boldsymbol{\beta} - \boldsymbol{\beta}\right) = \\ & \sum_{j=1}^{J} \frac{\lambda_j}{\left(\lambda_j + k\right)^2} + \boldsymbol{\beta}'\left(\left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right) - \mathbf{I}\right)'\left(\left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X}\right) - \mathbf{I}\right)\boldsymbol{\beta} = \\ & \sum_{j=1}^{J} \frac{\lambda_j}{\left(\lambda_j + k\right)^2} + \boldsymbol{\beta}' k \left(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-2} \boldsymbol{\beta} = \gamma_1(k) + \gamma_2(k) \end{aligned} \quad , (2.9)$$

where $\gamma_1(k)$ is the asymptotic variance and $\gamma_2(k)$ is the squared bias. The PRR estimation method is attractive for two reasons. Firstly, it is a very simple method since it does not require any changes of the existing Poisson regression software. Secondly, it has a lower MSE than the ML estimate if we find a value of k such that the reduction in the variance term is greater than the increase of the squared bias.

2.3 The MSE properties of the PRR Estimator

Hoerl and Kennard (1970a,b) showed that there exists a k greater than zero such that the MSE is always lower for RR than OLS. Here, it will be shown that there also exists such a k for which the MSE of the PRR is lower than the MSE of ML. In order to show this we first note that $\gamma_1(k)$ is a monotonic decreasing function of k. Then it has to be shown that $\gamma_2(k)$ is a monotonic increasing function of k which may be easily seen in the following equation:

$$\gamma_2(k) = k^2 \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{W}\mathbf{X} + k\mathbf{I}\right)^{-2} \boldsymbol{\beta} = k^2 \sum_{i=1}^{J} \frac{\alpha_i^2}{\left(\lambda_i + k\right)^2} , \tag{2.9}$$

where α_i^2 equals $\boldsymbol{\gamma}\boldsymbol{\beta}_{ML}$ and $\boldsymbol{\gamma}$ is the eigenvector of $\mathbf{X}'\mathbf{W}\mathbf{X}$. Since, by definition, $k \geq 0$ and $\lambda_i > 0$ for all i we may conclude that $\gamma_2(k)$ is a monotonic increasing function of k. Now in order to show that $E\left(L_{RR}^2\right) < E\left(L_{ML}^2\right)$ we have to take the first derivative of the $E\left(L_{RR}^2\right)$ with respect to k:

$$\frac{\delta E\left(L_{RR}^2\right)}{\delta k} = \frac{\delta \gamma_1(k)}{\delta k} + \frac{\delta \gamma_2(k)}{\delta k} = -2\sum_{i=1}^{l} \frac{\lambda_i}{\left(\lambda_i + k\right)^3} + 2k\sum_{i=1}^{l} \frac{\lambda_i \alpha_i^2}{\left(\lambda_i + k\right)^3}. \qquad (2.10)$$

It has already been shown that $\gamma_1(k)$ and $\gamma_2(k)$ are monotonically increasing and decreasing functions of *k,* respectively. Furthermore, it has also been shown that their first derivatives are always non-positive and non-negative, respectively. Hence, it is only necessary to show that there always exists a *k* greater than zero such that $\frac{\delta E\left(L_{RR}^2\right)}{\delta k} < 0$ to show that $E\left(L_{RR}^2\right) < E\left(L_{ML}^2\right)$. The condition for this is shown in (2.10) to be:

$$k < \frac{1}{\alpha_{\max}^2}, \qquad (2.11)$$

where $\alpha_{\max}^2$ is defined as the maximum element of α_i^2.

2.4 Proposed Ridge Parameter Estimators

To estimate the ridge parameter *k* we apply several different methods. The most classical RR is the following:

$$K1 = \hat{k}_{HK1} = \frac{s^2}{\hat{\alpha}_{\max}^2},$$

proposed by Hoerl and Kennard (1970a,b), where $s^2 = \frac{\sum_{i=1}^{n}\left(\mathbf{y}_i - \hat{\mu}_i\right)^2}{n-p-1}$. A modified version of this estimator is proposed as:

$$K2 = \hat{k}_{HKM} = \frac{1}{\hat{\alpha}_{\max}^2},$$

since the corresponding version of the existence theorem for linear regression in Hoerl and Kennard (1970a) for Poisson regression shows that the optimal value of *k* equals $\frac{1}{\alpha_{\max}^2}$ instead of $\frac{\sigma^2}{\alpha_{\max}^2}$. However, since these estimators have been shown in many studies (e.g.

Schaeffer (1986), Kibria (2003) and Alkhamisi and Shukur (2008)) to underestimate the optimal value of *k*, we include the following two estimators from Kibria (2003):

$$K3 = \hat{k}_{GM} = \frac{s^2}{\left(\prod_{i=1}^{p} \hat{\alpha}_i^2\right)^{\frac{1}{p}}} \text{ and } K4 = \hat{k}_{MED} = Median\{m_i^2\},$$

where $m_i = \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}$. Another estimator of *k* is the following:

$$K5 = \hat{k}_{\max}^{KS} = \max(s_i),$$

first proposed by Alkhamisi et al. (2006) where $s_i = \frac{t_i \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_i \hat{\alpha}_i^2}$ and t_i is the eigenvalues of the $\mathbf{X'X}$ matrix. The ridge parameter estimators proposed by Hoerl and Kennard (1970a,b) and Kibria (2003) share the same characteristic that their estimated values decrease as the degree of correlation (ρ) increases since the estimated vector of coefficients becomes on average longer. This characteristic is unattractive because larger values of *k* are needed to solve the problem of near singularity of the $\mathbf{X'WX}$ matrix as ρ increases. Based on the previous estimators and the idea of square root transformations taken from Alkhamisi and Shukur (2008) the following estimators were suggested by Muniz and Kibria (2009):

$$K6 = \hat{k}_{KM2} = \max\left(\frac{1}{m_i}\right), K7 = \hat{k}_{KM4} = \left(\prod_{i=1}^{p} \frac{1}{m_i}\right)^{\frac{1}{p}}, \; K8 = \hat{k}_{KM6} = median\left(\frac{1}{m_i}\right).$$

These estimators are based on the inverse of m_i so they actually increase as ρ becomes larger. As a result, these estimators are assumed to be the most robust to multicollinearity.

3. The Monte Carlo simulation

This section consists of a brief description of how the data is generated together with a discussion about the different factors varied in the simulation study. Then the criteria for judging the performance of the different estimation methods are presented.

3.1 The Design of the Experiment

The dependent variable of the Poisson regression model is generated using pseudo-random numbers from the $Po(\mu_i)$ where

$$\mu_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right), \qquad i = 1,2,...n,\ j = 1,2,...p. \tag{3.1}$$

The parameter values in equation (3.1) are chosen so that $\sum_{j=1}^{p} \beta_j^2 = 1$ and $\beta_1 = \cdots = \beta_p$, which are common restrictions in many simulation studies (e.g. Kibria (2003)). The first factor we choose to vary in the design of the experiment is ρ, which is the main interest of this paper. In the design of the experiment four different values of ρ corresponding to 0.85, 0.90, 0.95 and 0.99 are considered. To be able to generate data with different degrees of correlation we use the following formula:

$$x_{ij} = \left(1-\rho^2\right)^{(1/2)} z_{ij} + \rho z_{ip}, \qquad i = 1,2,...n,\ j = 1,2,...p \tag{3.2}$$

where z_{ij} are pseudo-random numbers generated using the standard normal distribution. To reduce eventual start-up value effects we discard the first 200 observations. Another factor we choose to vary is the sample size since previous studies (e.g. Muniz and Kibria (2009) and Månsson and Shukur (2010)) indicate that the gain of applying PRR is larger when n is small. However, the asymptotic MSE shows that there may be a substantial gain of using PRR even in large samples. We therefore evaluate the performance of the different estimation methods in both small and large sample sizes. However, since we investigate models with different numbers of explanatory variables we choose to fix the number of degrees of freedoms ($\Delta = n - p - 1$) instead of the number of observations. The value of the intercept (β_0) is also a factor we choose to vary. Decreasing this factor leads to a lower average value of μ_i which leads to less variation (i.e. more values equal to zero) in the sample. When decreasing this factor, we need larger sample sizes since otherwise the sample will often consists of only zeros which leads to a non-convergence of the Iterative WLS (IWLS) algorithm. In Table 1, the different combinations of values of the intercept and the sample size can be found. A final factor we consider is the number of explanatory variables (p) since it is of interest to find which ridge parameter is best for different number of p. We chose to generate models with 2 and 4 explanatory variables.

Table 1: The different combinations of intercept and sample sizes

	Degree of freedom							
Intercept	10	15	20	30	50	75	100	150
1	*	*	*	*	*			
0		*	*	*	*	*		
-1				*	*	*	*	*

3.2 Judging the performance of the estimators

To investigate the performance of the PRR and ML method, we calculate the MSE using the following equation:

$$\text{MSE} = \frac{\sum_{i=1}^{R} SE_i}{R} = \frac{\sum_{i=1}^{R} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)_i ' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)_i}{R}, \qquad (3.3)$$

where $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ obtained from ML or PRR, and R equals 2000 which corresponds to the number of replicates used in the Monte Carlo simulation. When the MSE is calculated we only use the slope parameters. Furthermore, the proportion of replication (out of 2000) for which the slope parameters of the ML estimator has a smaller squared error (SE) than the other estimators is also calculated.

4. Results

In this section the results from our Monte Carlo study are presented. The MSE for the different estimation methods can be found in Tables 2-4, and in what follows the effects of varying different factors on the performance of both the ML and the PRR methods are discussed.

4.1 The performance as a function of ρ

Our main interest of this paper is to investigate the effect of increasing the degree of correlation on the performance of the different estimation methods. From the results of the simulation study it becomes clear that increasing ρ has a negative impact on both ML and PRR. However, even though the MSE increases for both estimation methods there is still a substantial difference of how much it increases among the different estimation methods. At this stage it is important to mention that the least robust method of estimating the parameters of the Poisson regression model is ML. This can be seen both by looking at the immense increase of the MSE and the decrease of the proportion for which ML outperforms the different ridge estimators. Among the different estimators of *k* there is also a big variation in the robustness even though all of them are better than ML. In the methodology section we noted that calculating the ridge parameter using the inverse of m_i should be considered as the most robust method of estimating *k*. The results from our simulation study confirm this observation; especially the K6 which estimates *k* using the maximum value of the inverse of m_i has shown to be very robust. This estimator is the best option when we have high degree of correlation, i.e. when ρ equals 0.95 or 0.99. When ρ is less than 0.95, and when we only have two explanatory variables, the differences are not huge between the different ridge estimators. When we have four explanatory variables and a low degree of correlation, we find the K3, K4, K6, K7 and K8 to be the best options. Hence, the K6 estimator is the best option or very close to the best for all the evaluated situations.

4.2 The performance as a function of β_0

Decreasing the value of the intercept leads to less variation (i.e. more values that equal zero) in the sample. For given ρ**,** p and Δ this leads to an immense increase in the MSE of the ML estimator and the K1 to K5 estimators, while the estimators based on the inverse of m_i are basically unaffected. The proportion of times the ML has a lower SE than the PRR decreases for all of the different ridge estimators as the intercept decreases. Hence, we may conclude that the gain of using PRR increases as the value of the intercept decreases both by looking at the MSE and the proportion of times ML outperforms PRR.

4.3 The performance as a function of Δ and *p*

A desirable property of any statistical estimator is the convergence to the true value of the parameter as the sample size increases. This property holds for the ML estimator and most of the different estimators of the ridge parameter *k* since the MSE decreases with the sample size. When looking at the proportion of replication for which ML produces a smaller MSE than PRR, we can see that the proportion either increases or stay the same as *n* becomes larger. Hence, the benefit of using the PRR method is greater when the sample size is small. The effect of increasing the number of explanatory variables for a given ρ , Δ and β_0 leads to an increase of the MSE. Furthermore, we can also see that the proportion for which the ML outperforms the PRR decreases. Thus, we may conclude that there is a greater gain of using PRR instead of the ML when we have many explanatory variables.

5. Conclusions

In this paper, a PRR estimator is proposed. By means of Monte Carlo simulations we evaluate the traditional ML estimator and this new method using different estimators of the ridge parameter *k*. The results from the simulation study show that the sample size, the value of the intercept, the number of independent variables and the correlation between the independent variables are important factors for the performance of the different estimation methods. In most of the cases, the MSE decreases when the first two factors increases and becomes higher as the other factors increases. The result also shows that the proposed PRR method, regardless which ridge estimator used, has a lower MSE than the ML method for all different situations that has been evaluated. Hence, the main conclusion from this paper is that ML should not be used when the data is collinear since the vector of estimated parameters becomes too long. The PRR should always be preferred. The estimator introduced by Hoerl and Kennard (1970a,b) offers some reduction of the MSE but it underestimates the optimal *k*. The best option is to use the K6 estimator since it reduces the MSE substantially in all of the different situations investigated in this paper.

Table 2: Estimated MSE when $\beta_0 = 1$

	Estimated MSE with *p=2.*									Estimated MSE with *p=4.*								
	ML	K1	K2	K3	K4	K5	K6	K7	K8	ML	K1	K2	K3	K4	K5	K6	K7	K8
ρ=0.85																		
10	0.192	0.122	0.089	0.082	0.136	0.185	0.140	0.157	0.151	0.675	0.438	0.480	0.204	0.221	0.606	0.261	0.367	0.342
		(20)	(23)	(31)	(38)	(19)	(19)	(19)	(19)		(8.1)	(7.9)	(9.3)	(9.2)	(7.4)	(8.1)	(7.9)	(7.8)
15	0.103	0.077	0.057	0.049	0.090	0.101	0.088	0.095	0.092	0.291	0.227	0.242	0.127	0.131	0.279	0.189	0.231	0.225
		(24)	(28)	(36)	(43)	(24)	(24)	(24)	(24)		(11)	(10)	(13)	(13)	9.9	(11)	(10)	(10)
20	0.063	0.055	0.042	0.039	0.070	0.062	0.059	0.061	0.060	0.158	0.134	0.141	0.084	0.086	0.154	0.125	0.142	0.140
		(28)	(31)	(37)	(43)	(27)	(28)	(28)	(28)		(15)	(14)	(17)	(17)	(14)	(15)	(14)	(14)
30	0.034	0.031	0.026	0.026	0.059	0.034	0.033	0.033	0.033	0.085	0.078	0.081	0.058	0.058	0.084	0.077	0.081	0.081
		(30)	(33)	(43)	(49)	(29)	(30)	(30)	(30)		(21)	(21)	(25)	(25)	(20)	(21)	(21)	(21)
50	0.049	0.038	0.030	0.030	0.040	0.048	0.047	0.047	0.047	0.038	0.037	0.037	0.031	0.032	0.038	0.036	0.037	0.037
		(27)	(32)	(38)	(41)	(25)	(26)	(25)	(25)		(26)	(26)	(30)	(31)	(26)	(26)	(26)	(26)
ρ=0.90																		
10	0.266	0.167	0.111	0.081	0.125	0.254	0.176	0.204	0.194	0.922	0.551	0.621	0.239	0.253	0.804	0.294	0.441	0.409
		(17)	(19)	(25)	(32)	(15)	(16)	(15)	(15)		(3.9)	(3.7)	(5.1)	(5.3)	(3.5)	(4.0)	(3.8)	(3.8)
15	0.149	0.113	0.078	0.061	0.106	0.145	0.121	0.132	0.129	0.409	0.293	0.322	0.138	0.144	0.385	0.230	0.303	0.292
		(18)	(21)	(29)	(36)	(17)	(18)	(18)	(18)		(5.8)	(5.8)	(7.7)	(8.3)	(5.5)	(6.0)	(5.7)	(5.9)
20	0.091	0.075	0.052	0.042	0.079	0.089	0.081	0.086	0.084	0.238	0.192	0.204	0.106	0.107	0.229	0.169	0.202	0.197
		(21)	(24)	(30)	(37)	(20)	(21)	(21)	(21)		(8.5)	(8.3)	(10)	(11)	(8.1)	(8.5)	(8.3)	(8.2)
30	0.050	0.045	0.034	0.028	0.060	0.049	0.047	0.048	0.048	0.119	0.107	0.111	0.070	0.071	0.117	0.103	0.112	0.111
		(26)	(29)	(36)	(43)	(25)	(26)	(25)	(26)		(12)	(12)	(14)	(14)	(12)	(12)	(12)	(12)
50	0.027	0.025	0.021	0.018	0.043	0.026	0.026	0.026	0.026	0.056	0.053	0.054	0.041	0.041	0.055	0.052	0.054	0.054
		(31)	(34)	(41)	(48)	(31)	(31)	(31)	(31)		(18)	(18)	(22)	(22)	(18)	(19)	(18)	(18)
ρ=0.95																		
10	0.622	0.344	0.216	0.125	0.166	0.560	0.263	0.340	0.308	1.729	0.934	1.064	0.323	0.348	1.386	0.324	0.566	0.516
		(9.2)	(11)	(15)	(22)	(8.3)	(8.7)	(8.5)	(8.6)		(1.1)	(1.1)	(1.5)	(1.7)	(1.0)	(1.1)	(1.0)	(1.1)
15	0.284	0.173	0.097	0.057	0.105	0.267	0.194	0.228	0.215	0.825	0.512	0.584	0.203	0.221	0.735	0.301	0.450	0.421
		(11)	(13)	(17)	(26)	(10)	(11)	(10)	(10)		(1.8)	(1.8)	(2.2)	(2.5)	(1.5)	(1.8)	(1.7)	(1.7)
20	0.179	0.120	0.069	0.042	0.084	0.171	0.143	0.159	0.153	0.539	0.372	0.417	0.163	0.169	0.492	0.265	0.364	0.348
		(12)	(14)	(20)	(28)	(12)	(12)	(12)	(12)		(2.5)	(2.5)	(3.1)	(3.1)	(2.3)	(2.6)	(2.4)	(2.4)
30	0.108	0.084	0.050	0.029	0.059	0.104	0.096	0.102	0.100	0.245	0.200	0.213	0.103	0.103	0.234	0.177	0.212	0.208
		(15)	(18)	(23)	(30)	(15)	(15)	(15)	(15)		(4.0)	(3.8)	(5.5)	(5.7)	(3.7)	(4.0)	(3.7)	(3.8)
50	0.053	0.048	0.034	0.021	0.043	0.052	0.050	0.052	0.051	0.112	0.101	0.104	0.063	0.063	0.109	0.097	0.106	0.105
		(20)	(22)	(29)	(35)	(19)	(20)	(19)	(20)		(6.6)	(6.6)	(8.8)	(9.5)	(6.4)	(6.7)	(6.5)	(6.6)
ρ=0.99																		
10	3.103	1.159	0.521	0.214	0.211	2.112	0.299	0.517	0.415	9.648	4.171	5.055	1.263	1.461	5.213	0.267	0.658	0.552
		(2.4)	(2.9)	(5.4)	(10.1)	(1.9)	(2.2)	(2.1)	(2.1)		(0.1)	(0.1)	(0.1)	(0.1)	(0.0)	(0.1)	(0.1)	(0.1)
15	1.499	0.617	0.244	0.082	0.108	1.146	0.334	0.519	0.442	4.285	2.036	2.468	0.632	0.780	2.866	0.332	0.722	0.630
		(3.2)	(4.1)	(6.3)	(12.4)	(2.4)	(3.0)	(2.5)	(2.8)		(0.1)	(0.1)	(0.2)	(0.2)	(0.1)	(0.1)	(0.1)	(0.1)
20	1.015	0.450	0.161	0.058	0.104	0.805	0.353	0.511	0.446	2.641	1.330	1.632	0.418	0.474	1.903	0.368	0.733	0.663
		(3.4)	(4.2)	(7.3)	(14.4)	(2.9)	(3.2)	(2.9)	(3.1)		(0.1)	(0.1)	(0.0)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
30	0.570	0.296	0.098	0.030	0.077	0.480	0.313	0.398	0.367	1.397	0.801	0.959	0.295	0.325	1.107	0.367	0.615	0.569
		(4.7)	(5.8)	(8.1)	(16.0)	(4.3)	(4.6)	(4.3)	(4.4)		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
50	0.295	0.196	0.071	0.019	0.054	0.259	0.216	0.250	0.238	0.625	0.429	0.490	0.179	0.177	0.543	0.314	0.434	0.420
		(5.2)	(6.7)	(9.4)	(15.3)	(4.9)	(5.0)	(4.9)	(4.9)		(0.1)	(0.1)	(0.2)	(0.2)	(0.1)	(0.1)	(0.1)	(0.1)

Table 3: Estimated MSE when β_0 =0

	Estimated MSE with $p=2$.									Estimated MSE with $p=4$.								
	ML	K1	K2	K3	K4	K5	K6	K7	K8	ML	K1	K2	K3	K4	K5	K6	K7	K8
ρ=0.85																		
15	0.288	0.137	0.153	0.137	0.184	0.268	0.171	0.192	0.185	0.791	0.497	0.452	0.188	0.203	0.690	0.298	0.428	0.397
		(24)	(21)	(33)	(39)	(17)	(18)	(18)	(18)		(2.9)	(3.3)	(6.0)	(6.0)	(2.2)	(2.6)	(2.5)	(2.5)
20	0.187	0.091	0.107	0.096	0.135	0.178	0.136	0.147	0.144	0.470	0.333	0.302	0.139	0.150	0.433	0.257	0.333	0.316
		(28)	(23)	(35)	(42)	(19)	(20)	(20)	(20)		(4.5)	(4.7)	(7.1)	(7.4)	(3.7)	(4.0)	(3.9)	(3.9)
30	0.094	0.050	0.061	0.054	0.078	0.091	0.082	0.086	0.085	0.228	0.183	0.169	0.090	0.096	0.218	0.172	0.200	0.195
		(27)	(23)	(35)	(40)	(20)	(20)	(20)	(20)		(7.5)	(8.0)	(11.5)	(11.8)	(6.8)	(7.0)	(7.0)	(7.0)
50	0.051	0.031	0.039	0.032	0.044	0.049	0.048	0.049	0.049	0.100	0.089	0.084	0.056	0.058	0.097	0.091	0.096	0.095
		(35)	(29)	(39)	(42)	(27)	(27)	(27)	(27)		(11.3)	(11.8)	(16.1)	(16.1)	(10.5)	(10.8)	(10.5)	(10.6)
80	0.029	0.020	0.025	0.019	0.023	0.029	0.029	0.029	0.029	0.055	0.052	0.050	0.038	0.040	0.054	0.053	0.054	0.054
		(37)	(33)	(40)	(42)	(32)	(32)	(32)	(32)		(17.8)	(18.3)	(22.3)	(23.1)	(17.5)	(17.5)	(17.4)	(17.4)
ρ=0.90																		
15	0.439	0.181	0.201	0.169	0.207	0.397	0.214	0.246	0.236	1.243	0.742	0.678	0.256	0.287	1.025	0.334	0.517	0.466
		(17)	(14)	(25)	(30)	(10)	(11)	(11)	(11)		(2.0)	(2.2)	(3.4)	(3.6)	(1.4)	(1.8)	(1.6)	(1.6)
20	0.276	0.118	0.142	0.116	0.150	0.256	0.177	0.196	0.190	0.698	0.417	0.466	0.169	0.185	0.621	0.310	0.431	0.403
		(20)	(16)	(27)	(32)	(13)	(14)	(14)	(14)		(3.3)	(3.1)	(4.7)	(4.9)	(2.7)	(2.8)	(2.7)	(2.7)
30	0.133	0.062	0.079	0.064	0.084	0.127	0.113	0.118	0.117	0.342	0.257	0.232	0.112	0.120	0.319	0.226	0.275	0.265
		(23)	(20)	(30)	(33)	(17)	(17)	(17)	(17)		(3.7)	(3.9)	(6.0)	(5.7)	(3.4)	(3.6)	(3.5)	(3.5)
50	0.069	0.037	0.048	0.036	0.044	0.067	0.063	0.065	0.064	0.152	0.129	0.120	0.068	0.073	0.146	0.131	0.142	0.140
		(25)	(21)	(30)	(32)	(19)	(19)	(19)	(19)		(7.0)	(7.2)	(9.3)	(9.5)	(6.7)	(6.9)	(6.7)	(6.8)
80	0.043	0.025	0.034	0.023	0.026	0.042	0.042	0.042	0.042	0.085	0.077	0.073	0.050	0.052	0.083	0.079	0.083	0.082
		(30)	(26)	(34)	(36)	(25)	(25)	(25)	(25)		(10.6)	(11.1)	(13.7)	(14.1)	(10.2)	(10.3)	(10.3)	(10.3)
ρ=0.95																		
15	0.813	0.250	0.293	0.214	0.244	0.684	0.282	0.339	0.318	2.577	1.433	1.296	0.424	0.478	1.857	0.367	0.649	0.565
		(13)	(10)	(18)	(24)	(7.1)	(7.6)	(7.3)	(7.4)		(0.3)	(0.3)	(0.5)	(0.7)	(0.2)	(0.3)	(0.2)	(0.3)
20	0.560	0.171	0.228	0.159	0.195	0.485	0.263	0.306	0.291	1.400	0.835	0.734	0.246	0.276	1.139	0.375	0.610	0.541
		(13)	(10)	(19)	(23)	(6.8)	(7.7)	(7.4)	(7.4)		(0.2)	(0.2)	(0.5)	(0.7)	(0.1)	(0.2)	(0.2)	(0.2)
30	0.304	0.093	0.134	0.097	0.128	0.276	0.216	0.232	0.228	0.712	0.418	0.479	0.159	0.173	0.627	0.338	0.468	0.439
		(15)	(12)	(20)	(24)	(10)	(10)	(10)	(10)			(0.8)	(0.8)	(1.1)	(1.3)	(0.7)	(0.8)	(0.7)
50	0.145	0.050	0.074	0.052	0.069	0.135	0.127	0.131	0.130	0.311	0.239	0.215	0.096	0.105	0.288	0.225	0.266	0.257
		(18)	(15)	(22)	(25)	(13)	(13)	(13)	(13)		(1.9)	(2.1)	(2.5)	(2.7)	(1.8)	(1.9)	(1.8)	(1.8)
80	0.094	0.039	0.059	0.035	0.040	0.089	0.088	0.089	0.089	0.175	0.147	0.135	0.069	0.073	0.166	0.149	0.163	0.160
		(20)	(17)	(22)	(24)	(15)	(15)	(15)	(15)		(3.4)	(3.5)	(4.2)	(4.4)	(3.2)	(3.2)	(3.2)	(3.2)
ρ=0.99																		
15	5.083	1.317	1.653	0.807	0.738	2.916	0.224	0.351	0.290	14.359	6.996	6.136	1.835	1.964	5.856	0.239	0.587	0.452
		(3.1)	(2.3)	(6.3)	(9.0)	(1.2)	(1.5)	(1.5)	(1.5)		(0.0)	(0.0)	(0.1)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)
20	3.000	0.689	1.010	0.411	0.400	1.881	0.293	0.428	0.367	7.701	3.919	3.341	0.959	1.121	3.982	0.284	0.664	0.521
		(4.7)	(3.4)	(7.4)	(10.3)	(1.7)	(2.2)	(2.0)	(2.0)		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
30	1.658	0.331	0.546	0.234	0.257	1.108	0.355	0.469	0.422	3.684	2.004	1.643	0.478	0.524	2.228	0.385	0.778	0.644
		(5.1)	(3.8)	(7.0)	(10.2)	(2.4)	(2.8)	(2.6)	(2.6)		(0.0)	(0.0)	(0.0)	(0.2)	(0.0)	(0.0)	(0.0)	(0.0)
50	0.849	0.150	0.294	0.136	0.160	0.613	0.387	0.446	0.427	1.743	1.053	0.883	0.287	0.311	1.248	0.439	0.738	0.653
		(4.7)	(3.5)	(6.0)	(8.6)	(2.6)	(2.6)	(2.7)	(2.7)		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
80	0.518	0.095	0.184	0.096	0.116	0.394	0.331	0.358	0.351	0.968	0.634	0.540	0.188	0.203	0.756	0.413	0.599	0.553
		(5.5)	(4.1)	(6.7)	(8.3)	(3.4)	(3.4)	(3.4)	(3.4)		(0.1)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)	(0.1)	(0.1)

Table 4: Estimated MSE when β_0 =-1

	Estimated MSE with *p=2.*									Estimated MSE with *p=4.*								
	ML	K1	K2	K3	K4	K5	K6	K7	K8	ML	K1	K2	K3	K4	K5	K6	K7	K8
ρ=0.85																		
30	0.280	0.204	0.193	0.180	0.258	0.248	0.124	0.151	0.138	0.707	0.533	0.532	0.240	0.252	0.603	0.238	0.375	0.351
		(23)	(23)	(39)	(50)	(20)	(26)	(22)	(24)		(6.1)	(5.8)	(11)	(12)	(5.5)	(6.8)	(5.9)	(5.9)
50	0.134	0.110	0.109	0.116	0.216	0.124	0.084	0.101	0.094	0.286	0.242	0.247	0.144	0.151	0.266	0.178	0.229	0.220
		(25)	(25)	(45)	(60)	(23)	(28)	(24)	(26)		(10)	(9.9)	(17)	(18)	(9.4)	(11)	(9.7)	(9.8)
80	0.078	0.069	0.069	0.081	0.189	0.074	0.060	0.068	0.065	0.159	0.143	0.146	0.100	0.103	0.151	0.122	0.142	0.138
		(29)	(29)	(49)	(64)	(28)	(30)	(28)	(29)		(14)	(14)	(22)	(23)	(13)	(15)	(14)	(14)
100	0.054	0.049	0.050	0.065	0.189	0.052	0.045	0.050	0.048	0.108	0.100	0.101	0.076	0.079	0.103	0.091	0.100	0.098
		(31)	(31)	(53)	(71)	(29)	(32)	(30)	(30)		(15)	(15)	(22)	(25)	(15)	(16)	(15)	(15)
150	0.034	0.032	0.032	0.047	0.177	0.033	0.030	0.032	0.032	0.062	0.059	0.060	0.051	0.053	0.060	0.057	0.060	0.059
		(34)	(33)	(58)	(76)	(32)	(34)	(33)	(33)		(22)	(22)	(29)	(32)	(22)	(22)	(22)	(22)
ρ=0.90																		
30	0.399	0.267	0.251	0.216	0.284	0.347	0.145	0.187	0.169	1.011	0.718	0.723	0.291	0.311	0.839	0.278	0.481	0.444
		(17)	(17)	(33)	(43)	(15)	(20)	(16)	(17)		(3.6)	(3.5)	(6.5)	(6.9)	(3.4)	(4.0)	(3.6)	(3.6)
50	0.195	0.148	0.149	0.137	0.228	0.177	0.108	0.134	0.124	0.431	0.348	0.356	0.178	0.179	0.387	0.220	0.310	0.297
		(21)	(21)	(39)	(53)	(19)	(23)	(19)	(20)		(5.9)	(5.8)	(9.8)	(11)	(5.5)	(6.3)	(5.7)	(5.7)
80	0.114	0.095	0.096	0.094	0.191	0.106	0.080	0.094	0.089	0.234	0.204	0.208	0.119	0.122	0.219	0.165	0.202	0.197
		(22)	(22)	(40)	(57)	(20)	(23)	(20)	(21)		(8.0)	(7.9)	(13)	(15)	(7.5)	(8.7)	(7.9)	(8.0)
100	0.079	0.069	0.070	0.076	0.183	0.075	0.063	0.070	0.068	0.156	0.141	0.144	0.093	0.094	0.148	0.124	0.142	0.139
		(25)	(24)	(46)	(62)	(23)	(26)	(23)	(24)		(10)	(10)	(14)	(17)	(9.9)	(10)	(10)	(10)
150	0.052	0.048	0.048	0.057	0.171	0.050	0.045	0.049	0.048	0.092	0.086	0.087	0.065	0.067	0.088	0.081	0.087	0.086
		(26)	(26)	(47)	(66)	(25)	(27)	(25)	(26)		(12)	(12)	(17)	(19)	(12)	(12)	(12)	(12)
ρ=0.95																		
30	0.847	0.491	0.434	0.335	0.358	0.673	0.162	0.228	0.199	2.034	1.339	1.328	0.438	0.463	1.489	0.292	0.615	0.565
		(10)	(11)	(22)	(31)	(8.5)	(12)	(9.2)	(10.4)		(1.2)	(1.2)	(2.2)	(2.2)	(1.1)	(1.4)	(1.2)	(1.2)
50	0.413	0.261	0.258	0.195	0.259	0.345	0.146	0.197	0.177	0.844	0.617	0.632	0.230	0.238	0.699	0.283	0.489	0.460
		(12)	(12)	(27)	(40)	(10)	(13)	(11)	(12)		(2.2)	(2.2)	(4.0)	(4.7)	(2.1)	(2.3)	(2.3)	(2.2)
80	0.254	0.183	0.187	0.134	0.198	0.221	0.134	0.171	0.159	0.446	0.358	0.367	0.157	0.159	0.392	0.234	0.337	0.326
		(14)	(13)	(27)	(41)	(13)	(14)	(13)	(13)		(2.8)	(2.7)	(4.7)	(5.7)	(2.5)	(2.8)	(2.5)	(2.7)
100	0.167	0.126	0.130	0.098	0.180	0.148	0.108	0.130	0.123	0.320	0.270	0.276	0.134	0.133	0.290	0.206	0.267	0.260
		(14)	(15)	(30)	(46)	(13)	(16)	(13)	(14)		(3.1)	(3.1)	(4.3)	(5.5)	(2.9)	(3.2)	(3.0)	(2.9)
150	0.105	0.088	0.090	0.071	0.153	0.096	0.081	0.092	0.088	0.177	0.158	0.161	0.096	0.096	0.165	0.139	0.160	0.158
		(16)	(16)	(32)	(52)	(15)	(16)	(15)	(15)		(5.1)	(5.0)	(7.6)	(9.0)	(4.9)	(5.2)	(4.9)	(5.1)
ρ=0.99																		
30	5.008	2.165	1.694	1.171	1.021	2.611	0.113	0.220	0.147	11.235	5.681	5.866	1.661	1.905	4.878	0.173	0.547	0.454
		(2.2)	(2.3)	(7.4)	(12.5)	(1.4)	(3.0)	(2.0)	(2.1)		(0.1)	(0.0)	(0.1)	(0.1)	(0.0)	(0.1)	(0.0)	(0.0)
50	2.524	1.083	0.970	0.575	0.521	1.411	0.131	0.277	0.196	4.749	2.712	2.823	0.735	0.832	2.533	0.258	0.731	0.638
		(2.8)	(2.9)	(8.6)	(15.9)	(2.0)	(3.5)	(2.5)	(2.8)		(0.0)	(0.0)	(0.1)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)
80	1.384	0.598	0.586	0.322	0.308	0.818	0.163	0.293	0.235	2.609	1.596	1.671	0.474	0.510	1.573	0.328	0.790	0.715
		(3.8)	(3.9)	(9.9)	(17.8)	(3.0)	(4.1)	(3.1)	(3.3)		(0.1)	(0.1)	(0.1)	(0.4)	(0.1)	(0.1)	(0.1)	(0.1)
100	1.012	0.483	0.484	0.253	0.240	0.627	0.178	0.295	0.247	1.753	1.156	1.211	0.355	0.377	1.156	0.346	0.732	0.677
		(3.7)	(3.7)	(9.7)	(20.5)	(3.3)	(4.1)	(3.4)	(3.7)		(0.0)	(0.0)	(0.1)	(0.3)	(0.0)	(0.0)	(0.0)	(0.0)
150	0.641	0.330	0.353	0.184	0.186	0.424	0.192	0.276	0.247	1.021	0.742	0.764	0.242	0.236	0.745	0.337	0.616	0.589
		(3.5)	(3.5)	(10.1)	(22.0)	(3.2)	(3.8)	(3.1)	(3.4)		(0.1)	(0.1)	(0.3)	(0.6)	(0.0)	(0.1)	(0.0)	(0.0)

References

Alkhamisi, M. A., Khalaf, G. and Shukur, G. (2006). Some modifications for Choosing Ridge Parameter. *Communications in Statistics- Theory and Methods*, 35: 1-16.

Alkhamisi, M. and Shukur, G. (2008). Developing ridge parameters for SUR model. *Communications in Statistics- Theory and Methods*, 37: 544-564.

Dempster, A.P., Schatzoff, M., Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association*, 72: 77-91.

Gibbons, D.G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association*, 76: 131-139.

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- Theory and Methods*, 4: 105–123.

Khalaf, G. and Shukur, G. (2005). Choosing ridge parameters for regression problems. *Communications in Statistics- Theory and Methods* , 34: 1177-1182.

Kibria, B.M.G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics- Theory and Methods*, 32: 419-435.

Lawless, J.F. and Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics- Theory and Methods*, 5: 307-323.

McDonald, G.C., 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. G. (2009). On some ridge regression estimators: An Empirical Comparisons. *Communications in Statistics-Simulation and Computation*, 38: 621- 630.

Månsson, K. and Shukur, B.M.G. (2010). On ridge parameters in logistic regression. To appear in *Communications in Statistics- Theory and Methods*.

Schaefer, R.L., Roi, L. D. and Wolfe, R. A. (1984). A ridge logistic estimator. *Communications in Statistics- Theory and Methods*, 13: 99-113.

Schaefer, R.L. (1986). Alternative Estimators in logistic regression when the data is collinear. *Journal of Statistical Computation and Simulation*, 25: 75-91.