

Improved Ridge Regression Estimators for Binary Choice Models: An Empirical Study

Kristofer Månsson^{1,2,#}, B.M. Golam Kibria^{3,*} and Ghazi Shukur^{1,4,#}

¹Department of Economics and Statistics, Jönköping University, Sweden

²Department of Economics and Statistics, Göteborg University, Sweden

³Department of Mathematics and Statistics, Florida International University, Miami, Florida, USA

⁴Centre for Labour Market Policy (CAFO), Department of Economics and Statistics, Linnaeus University, Sweden

Abstract: This paper suggests some new estimators of the ridge parameter for binary choice models that may be applied in the presence of a multicollinearity problem. These new ridge parameters are functions of other estimators of the ridge parameter that have shown to work well in the previous research. Using a simulation study we investigate the mean square error (MSE) properties of these new ridge parameters and compare them with the best performing estimators from the previous research. The results indicate that we may improve the MSE properties of the ridge regression estimator by applying the proposed estimators in this paper, especially when there is a high multicollinearity between the explanatory variables and when many explanatory variables are included in the regression model. The benefit of this paper is then shown by a health related data where the effect of some risk factors on the probability of receiving diabetes is investigated.

Keywords: Binary Choice Models, Estimation, MSE, Multicollinearity, Ridge Regression, Simulation.

1. INTRODUCTION

For binary choice models one often assumes that we want to model the relationship between some explanatory variables and a latent variable. This relationship may be described using the subsequent equation:

$$y_i^* = x_i' \beta + u_i, \quad (1.1)$$

where one may define x_i as the i th row of X , which is a $n \times (p+1)$ data matrix with p explanatory variables, β is a $(p+1) \times 1$ vector of coefficients and u_i is an error term. A latent variable is a random variable whose realized values are hidden. Instead, the property of the latent variable must be inferred indirectly using a statistical model such as the one presented in equation (1.1) that connects the latent (unobserved) variables to observed variables. Common examples of latent variables are different psychometric measures (see for example [1]) and in economics it may be the utility when purchasing a product (see [2]). In all of these cases we may not observe the latent variable directly. Instead we may investigate the following binary dependent variable:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1.2)$$

Depending on the distribution of the error term in equation (1.1) we may have two different regression models. The first is the probit model which should be used when the error term u_i is normally distributed. The second option is the logit model that may be applied if the error term u_i is logistic distributed. For both models the parameters are estimated using the maximum likelihood (ML) method by applying the following iterative weighted least square (IWLS) algorithm:

$$\hat{\beta}_{ML} = (X' \hat{W} X)^{-1} X' \hat{W} \hat{z} \quad (1.3)$$

where \hat{z} is a vector where the i th element equals

$$\hat{z}_i = \log(\hat{\pi}_i) + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1 - \hat{\pi}_i)}$$

and \hat{W} is a matrix where the non-diagonal elements equal to zero and the i th diagonal element equals $(\hat{\pi}_i)(1 - \hat{\pi}_i)$. However, when the explanatory variables are collinear then the estimated parameters become unstable with high variances. Hence, the consequences of multicollinearity are wider confidence intervals and an increase in the probability of conducting a type II error in any hypothesis testing regarding the estimated parameters. To solve this problem, caused by

*Address correspondence to this author at the Department of Mathematics and Statistics, Florida International University, Miami, Florida, USA; Tel: (305) 348 1419; Fax: (305) 348 6895; E-mail: kibriag@fiu.edu

#Co-Authors E-mail: Ghazi.Shukur@jibs.hj.se, kristofer.mansson@jibs.hj.se

AMS Subject classification: Primary 62J07, Secondary 62F10.

multicollinearity, [3] suggested using the following ridge regression estimator:

$$\hat{\beta}_{RR} = (X'WX + kI)^{-1} X'WX\hat{\beta}_{ML} \tag{1.4}$$

where $\hat{\beta}_{ML}$ is defined in (1.3). The shrinkage parameter k in equation (1.4) may take on values between zero and infinity, and when k equals zero, we have $\hat{\beta}_{RR} = \hat{\beta}_{ML}$. When k is greater than zero we have that $\|\hat{\beta}_{RR}\| \leq \|\hat{\beta}_{ML}\|$. Since $\hat{\beta}_{ML}$ is, on average, too long in the presence of multicollinearity, $\hat{\beta}_{RR}$ is expected to perform better and thus have a lower mean squared error (MSE) than $\hat{\beta}_{ML}$. This is well-established both theoretically and by means of Monte Carlo simulation for binary choice models by for example [3, 4]. The purpose of this paper is to introduce some new estimators of the ridge parameter in equation (1.4). These new ridge parameters are functions of other estimators of k that have shown to work well in the previous research. In a simulation study we investigate the MSE properties of these new ridge parameters and compare them with the best performing estimators of the ridge parameter from the existing research. The results indicate that the new ridge parameter may improve the MSE properties of the ridge estimator, especially when there is high multicollinearity between the explanatory variables and when many independent variables are included in the regression model. The advantage of using ridge regression techniques is finally shown by an empirical example where it is investigated how some risk factors affects the probability of receiving diabetes. This example is included mainly to show how the ridge regression may work when analyzing health related data and not to reach conclusive evidence of the effect of the risk factors on the probability of getting diabetes.

The organization of this paper is as follows: The different estimators of the ridge parameters is given in section 2. A simulation study has been conducted in section 3. To illustrate the findings of the paper a real life health related data are analysed in section 4. Finally some concluding remarks are given in section 5.

2. ESTIMATORS OF THE RIDGE PARAMETER K

Several methods of estimation for the ridge parameter k have been suggested by several researchers in different time. For the linear regression model and using ordinary least squares, several methods have been proposed by [5-15], and very recently [16-18] among others. For non-linear models some research has been conducted and some estimators of the ridge parameter have been proposed

by [3, 4, 19] among others. In this section we introduce the classical method suggested by [5, 6] along with some good ridge regression estimators according to the simulated result from [19]. Finally some new ridge regression estimators are introduced.

In order to explain these estimators we start by defining $X'WX = U\Lambda U'$ and $\hat{\alpha} = U'\hat{\beta}_{ML}$, where U is the matrix whose columns are the eigenvectors of $X'WX$ and Λ is the diagonal matrix of it's eigenvalues corresponding to the columns of U . The first estimator for the linear regression model was proposed by [5, 6] as follows:

$$k_1 = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2}, \tag{2.1}$$

where α_{\max} is the maximum element of vector, $\hat{\alpha}$ and $\hat{\sigma}^2$ is the residual variance of the raw residuals divided by the degrees of freedoms ($n-p-1$). Then, the following estimators that were found to be the optimal in [19] are considered:

$$k_2 = \max\left(\frac{1}{m_j}\right), k_3 = \prod_{j=1}^p \left(\frac{1}{m_j}\right)^{\frac{1}{p}}, k_4 = \text{median}\left(\frac{1}{m_j}\right), \tag{2.2}$$

$$k_5 = \max\left(\frac{1}{q_j}\right), k_6 = \prod_{j=1}^p \left(\frac{1}{q_j}\right)^{\frac{1}{p}}, k_7 = \text{median}\left(\frac{1}{q_j}\right)$$

where $m_j = \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}}$ and $q_j = \frac{\lambda_{\max}}{(n-p)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_j^2}$.

The purpose of this paper is to see whether a simpler way we can develop some estimators for the ridge parameter in equation (1.4). As such, using the above estimators, we propose the following ridge regression estimators:

$$k_8 = \frac{\sum_{l=2}^7 k_l}{(l-1)}, k_9 = \text{median}(k_1, k_2, \dots, k_7), \text{ and } k_{10} = \prod_{l=2}^7 (k_l)^{1/(l-1)} \tag{2.3}$$

where $l = 2, \dots, 7$. Hence, these new ridge parameters are functions of other estimators of k that have shown to work well in the previous research.

3. THE MONTE CARLO SIMULATION

Since a theoretical comparison among the estimators is not possible, a simulation study has been conducted in this section. It contains two components

(i) the design of the simulation study and (ii) results discussion.

3.1. The Design of the Simulation Study

The main focus of this paper is to compare the MSE properties of the ML and ridge regression estimator when the explanatory variables are collinear. Therefore, the main factor varied in the design of the simulation study is the degree of correlation (ρ^2) between the explanatory variables. The subsequent equation defining the data generating process for the explanatory variables is used since it enables us to vary the strength of the correlation (see [10, 11] among others):

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

In equation (3.1) z_{ij} are pseudo-random numbers from the standard normal distribution. The parameter ρ determines the strength of the correlation and we consider three different values that corresponds to 0.85, 0.95 and 0.99. We simulate n observations for the dependent variable using equation (1.1) where u_i are random numbers generated using the standard normal distribution and the logistic distribution for the probit and logit models, respectively. Then the latent variable is made binary by applying equation (1.2). The parameter values are chosen so that $\beta' \beta = 1$ and all individual parameter values are equal to each other. We consider four, six and eight explanatory variables in the simulation. We also consider the number of observations that corresponds to 100, 200 and 400 observations for the model consisting of four explanatory variables, 150, 200 and 400 observations for the model consisting of six explanatory variables and 200, 400 and 600 observations for the regression model consisting of eight explanatory variables. The reason we change the number of observations

depending on the number of explanatory variables is to ensure convergence of the maximum likelihood estimator. The experiment is replicated (R) 2000 times by generating new pseudo-random numbers, while the MSE is calculated as follows:

$$MSE = \frac{\sum_{i=1}^R (\hat{\beta} - \beta)'_i (\hat{\beta} - \beta)_i}{R}, \quad (3.2)$$

where $\hat{\beta}_j$ is the estimator of β at the i th repetition.

3.2. Result Discussion

The estimated MSEs of the estimators for the logit model and for $p=4$, $p=6$ and $p=8$ are presented in Tables 1, 2 and 3 respectively and for probit model for $p=4$, $p=6$ and $p=8$ are presented in Tables 4, 5 and 6 respectively. From these tables we see that the degree of correlation in general has a negative impact on the MSE of the different estimators. The effect of multicollinearity on the MSE is much more severe for the ML estimator than the ridge regression estimator. One may observe that the estimated MSE for the ridge regression estimator is much lower than the corresponding ML estimator when applying the ridge parameters from [19]. However, the new suggested ridge parameters (in particular k_8 and k_9) outperform the older ones, especially in the presence of a high degree of collinearity among the explanatory variables. When looking at the number of explanatory variables we may see that increasing p leads to a higher estimated MSE for all estimators. Once again the benefit of applying ridge regression increases with the number of explanatory variables and also the advantage of using the new ridge parameters suggested in this paper. The final factor affecting the performance of the different estimation methods is the number of observations. Increasing this factor leads to a decrease

Table 1: Estimated MSE for the Logit Model when $p=4$

ρ	n	ML	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
0.85	100	1.34	0.93	0.48	0.75	0.68	0.29	0.28	0.28	0.19	0.19	0.18
	200	0.55	0.43	0.36	0.45	0.43	0.13	0.13	0.13	0.14	0.14	0.17
	400	0.27	0.23	0.22	0.24	0.24	0.10	0.10	0.10	0.12	0.12	0.15
0.95	100	4.10	2.49	0.47	0.96	0.80	0.32	0.31	0.31	0.23	0.23	0.23
	200	1.78	1.17	0.55	0.90	0.80	0.18	0.17	0.17	0.18	0.18	0.20
	400	0.79	0.58	0.44	0.58	0.55	0.14	0.15	0.15	0.18	0.18	0.22
0.99	100	21.0	11.58	0.28	0.52	0.40	0.39	0.35	0.35	0.28	0.27	0.31
	200	9.16	5.22	0.38	0.85	0.65	0.23	0.21	0.21	0.20	0.20	0.21
	400	4.23	2.46	0.52	1.00	0.82	0.19	0.19	0.19	0.21	0.21	0.23

Table 2: Estimated MSE for the Logit Model when $\rho=6$

ρ	n	ML	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
0.85	100	1.70	1.31	0.63	1.08	1.00	0.29	0.28	0.28	0.22	0.21	0.31
	200	1.18	0.94	0.59	0.88	0.83	0.21	0.21	0.20	0.19	0.19	0.33
	400	0.52	0.45	0.39	0.47	0.46	0.14	0.14	0.14	0.19	0.19	0.30
0.95	100	5.36	3.72	0.62	1.49	1.28	0.34	0.32	0.32	0.25	0.25	0.29
	200	3.65	2.58	0.69	1.45	1.29	0.26	0.25	0.24	0.22	0.22	0.34
	400	1.62	1.22	0.66	1.07	1.00	0.18	0.18	0.19	0.23	0.23	0.40
0.99	100	28.5	18.3	0.31	0.78	0.63	0.44	0.37	0.36	0.30	0.29	0.26
	200	19.6	12.7	0.36	1.04	0.83	0.33	0.29	0.28	0.25	0.25	0.26
	400	8.89	5.85	0.57	1.54	1.29	0.23	0.22	0.22	0.24	0.24	0.33

Table 3: Estimated MSE for the Logit Model when $\rho=8$

ρ	n	ML	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
0.85	100	2.29	1.88	0.73	1.37	1.27	0.56	0.53	0.52	0.38	0.38	0.50
	200	0.96	0.84	0.58	0.80	0.77	0.27	0.26	0.26	0.26	0.26	0.46
	400	0.22	0.20	0.16	0.19	0.19	0.07	0.07	0.07	0.08	0.08	0.13
0.95	100	7.59	5.66	0.74	1.91	1.64	0.67	0.61	0.60	0.48	0.47	0.54
	200	2.92	2.30	0.85	1.59	1.46	0.38	0.36	0.36	0.35	0.35	0.64
	400	1.92	1.56	0.76	1.29	1.21	0.31	0.31	0.30	0.34	0.34	0.61
0.99	100	40.01	27.68	0.69	1.03	0.84	0.86	0.69	0.67	0.59	0.55	0.46
	200	16.87	11.88	0.61	1.76	1.45	0.50	0.44	0.43	0.42	0.41	0.53
	400	10.73	7.63	0.74	2.05	1.75	0.42	0.39	0.39	0.40	0.40	0.62

Table 4: Estimated MSE for the Probit Model when $\rho=4$

ρ	n	ML	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
0.85	100	0.82	0.63	0.33	0.49	0.46	0.19	0.18	0.18	0.16	0.16	0.17
	200	0.30	0.26	0.22	0.26	0.25	0.10	0.11	0.11	0.14	0.14	0.15
	400	0.13	0.12	0.12	0.13	0.12	0.08	0.08	0.08	0.10	0.10	0.10
0.95	100	2.18	1.46	0.40	0.75	0.64	0.23	0.22	0.22	0.19	0.19	0.21
	200	0.89	0.66	0.41	0.58	0.54	0.16	0.16	0.16	0.21	0.21	0.23
	400	0.39	0.32	0.27	0.33	0.32	0.14	0.15	0.15	0.19	0.19	0.20
0.99	100	11.11	6.51	0.31	0.64	0.52	0.30	0.26	0.26	0.24	0.23	0.26
	200	4.50	2.72	0.42	0.86	0.71	0.20	0.20	0.20	0.22	0.23	0.23
	400	2.05	1.31	0.50	0.85	0.75	0.21	0.22	0.22	0.28	0.28	0.29

of the estimated MSE. Here one may see that the ridge regression outperforms the ML for all different sample sizes. However, the benefit of ridge regression estimators is higher in small sample sizes. From the simulation study we can conclude that the ridge

regression estimators outperform the ML estimator. Furthermore, we show that one may improve the MSE properties of the ridge regression estimator by applying the new ridge parameters proposed in this paper, especially when there is a high multicollinearity

Table 5: Estimated MSE for the Probit Model when $\rho=6$

ρ	n	ML	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
0.85	150	0.84	0.69	0.47	0.65	0.62	0.23	0.23	0.23	0.26	0.26	0.35
	200	0.56	0.48	0.39	0.48	0.47	0.20	0.20	0.21	0.25	0.25	0.32
	400	0.27	0.24	0.24	0.26	0.25	0.17	0.17	0.17	0.20	0.20	0.22
0.95	150	2.51	1.83	0.64	1.19	1.08	0.29	0.28	0.28	0.31	0.31	0.44
	200	1.68	1.25	0.62	1.04	0.97	0.26	0.26	0.26	0.32	0.32	0.46
	400	0.74	0.60	0.47	0.62	0.60	0.24	0.25	0.26	0.32	0.32	0.41
0.99	150	13.12	8.64	0.45	1.17	0.97	0.36	0.31	0.31	0.31	0.30	0.35
	200	8.80	5.80	0.54	1.33	1.13	0.30	0.29	0.29	0.31	0.32	0.39
	400	3.99	2.74	0.69	1.43	1.27	0.27	0.30	0.31	0.39	0.40	0.54

Table 6: Estimated MSE for the Probit Model when $\rho=8$

ρ	n	ML	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
0.85	200	0.99	0.84	0.29	0.54	0.50	0.34	0.30	0.30	0.23	0.23	0.25
	400	0.54	0.49	0.35	0.46	0.45	0.22	0.21	0.21	0.21	0.22	0.32
	600	0.33	0.31	0.25	0.30	0.30	0.15	0.16	0.16	0.18	0.18	0.24
0.95	200	7.58	4.94	0.58	1.33	1.17	0.70	0.61	0.60	0.50	0.49	0.51
	400	2.72	1.83	0.60	1.07	1.00	0.35	0.33	0.33	0.34	0.34	0.54
	600	1.05	0.90	0.53	0.80	0.76	0.27	0.27	0.28	0.32	0.32	0.49
0.99	200	35.50	22.10	0.72	1.04	0.88	0.91	0.71	0.69	0.64	0.59	0.50
	400	14.65	9.08	0.61	1.60	1.39	0.50	0.43	0.42	0.42	0.42	0.59
	600	7.92	4.93	0.70	1.67	1.48	0.40	0.38	0.39	0.42	0.43	0.68

between the explanatory variables and when many independent variables are included in the regression model.

4. EMPIRICAL APPLICATION

The benefit of applying ridge regression estimators are demonstrated by an empirical application where we use a dataset collected by the US National Institute of Diabetes and Digestive and Kidney Diseases.¹ This dataset is based on a sample from a population consisting of women who were at least 21 years old, of Pima Indian heritage and living near Phoenix, Arizona. The data were collected by the US National Institute of Diabetes and Digestive and Kidney Diseases and these women were then tested for diabetes according to the World Health Organization criteria (more information about the dataset may be found in [20]). The explanatory variables are the number of

pregnancies (X_1), plasma glucose concentration in an oral glucose tolerance test (X_2), diastolic blood pressure measured in hg (X_3), skin triceps skin fold thickness measured in mm (X_4), body mass index (X_5), diabetes pedigree function (X_6) and age (X_7). According to [20] these variables were chosen because they have been found to be significant risk factors for diabetes among Pimas and other populations. Hence, these explanatory variables explains the dependent variable which takes on the value one if the person has diabetes and zero otherwise.²

The results of the data analyses by using the logit and probit model are presented in Tables 7 and 8 respectively. The standard errors are obtained by using bootstrapping techniques. We clearly see that all of the significant coefficients are positive which means that all of the explanatory variables increases the probability of getting diabetes. This is natural since the variables were chosen by [20] where risk factors assumed to

¹This dataset is publicly available via the package MASS in R. It is denoted *PIMA.tr*.

²Diabetes was defined as a plasma glucose concentration greater the 200 mg/dl two hours following the ingestion of 75 gm of a carbohydrate solution.

Table 7: Estimated Parameters using the Probit Model

Estimators		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
ML	$\hat{\beta}_j$	0.06	0.02	0.00	-0.00	0.05	1.06	0.02
	$se(\hat{\beta}_j)$	0.04	0.01	0.01	0.01	0.03	0.40	0.01
	t-value	1.37	3.69	0.19	-0.12	1.70	2.68	1.67
k ₁	$\hat{\beta}_j$	0.06	0.02	0.00	0.00	0.05	1.05	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.03	0.39	0.02
	t-value	1.38	4.56	0.26	-0.08	1.63	2.68	1.39
k ₂	$\hat{\beta}_j$	0.06	0.01	0.02	0.01	-0.01	0.48	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.02	0.21	0.01
	t-value	1.52	3.65	2.88	1.15	-0.57	2.31	1.53
k ₃	$\hat{\beta}_j$	0.06	0.02	0.02	0.01	0.02	0.78	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.02	0.28	0.01
	t-value	1.58	4.41	1.61	0.58	0.72	2.77	1.67
k ₄	$\hat{\beta}_j$	0.06	0.02	0.02	0.01	0.01	0.76	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.02	0.28	0.01
	t-value	1.59	4.36	1.70	0.62	0.64	2.67	1.66
k ₅	$\hat{\beta}_j$	0.06	0.01	-0.03	0.02	-0.02	0.24	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.02	0.13	0.01
	t-value	1.53	3.35	3.54	1.40	-1.18	1.81	1.44
k ₆	$\hat{\beta}_j$	0.06	0.02	0.02	0.01	0.02	0.78	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.02	0.28	0.01
	t-value	1.59	4.40	1.59	0.57	0.73	2.81	1.68
k ₇	$\hat{\beta}_j$	0.06	0.02	0.02	0.01	0.02	0.78	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.02	0.29	0.01
	t-value	1.59	4.37	1.62	0.58	0.73	2.67	1.68
k ₈	$\hat{\beta}_j$	0.06	0.01	0.02	0.01	-0.01	0.50	0.02
	$se(\hat{\beta}_j)$	0.03	0.00	0.01	0.01	0.02	0.22	0.01
	t-value	1.66	3.76	2.86	1.12	-0.49	2.26	1.54
k ₉	$\hat{\beta}_j$	0.06	0.02	0.02	0.01	0.01	0.77	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.02	0.28	0.01
	t-value	1.59	4.37	1.67	0.60	0.68	2.73	1.68
k ₁₀	$\hat{\beta}_j$	0.06	0.01	0.02	0.01	0.00	0.65	0.02
	$se(\hat{\beta}_j)$	0.04	0.00	0.01	0.01	0.02	0.26	0.01
	t-value	1.63	4.10	2.22	0.85	0.14	2.56	1.62

Note: $\hat{\beta}_j$ is the j th value (from 1, ..., p) of the parameter vector $\hat{\beta}$.

Table 8: Estimated Parameters using the Logit Model

Estimators		x_1	x_2	x_3	x_4	x_5	x_6	x_7
ML	$\hat{\beta}_j$	0.10	0.03	0.00	0.00	0.08	1.81	0.04
	$se(\hat{\beta}_j)$	0.08	0.01	0.02	0.02	0.05	0.80	0.03
	t-value	1.34	3.90	0.21	-0.08	1.83	2.27	1.39
k_1	$\hat{\beta}_j$	0.10	0.03	0.01	0.00	0.08	1.78	0.04
	$se(\hat{\beta}_j)$	0.08	0.01	0.02	0.02	0.05	0.71	0.03
	t-value	1.35	3.91	0.28	-0.04	1.76	2.52	1.60
k_2	$\hat{\beta}_j$	0.08	0.02	0.04	0.03	-0.04	0.29	0.02
	$se(\hat{\beta}_j)$	0.05	0.01	0.01	0.02	0.03	0.15	0.02
	t-value	1.57	3.07	3.30	1.46	-1.30	1.90	1.40
k_3	$\hat{\beta}_j$	0.09	0.02	0.04	0.02	-0.02	0.70	0.03
	$se(\hat{\beta}_j)$	0.06	0.01	0.01	0.02	0.03	0.30	0.02
	t-value	1.57	3.37	2.79	1.22	-0.70	2.31	1.48
k_4	$\hat{\beta}_j$	0.09	0.02	0.04	0.02	-0.02	0.67	0.03
	$se(\hat{\beta}_j)$	0.06	0.01	0.01	0.02	0.03	0.29	0.02
	t-value	1.62	3.44	2.86	1.26	-0.76	2.27	1.48
k_5	$\hat{\beta}_j$	0.07	0.02	0.05	0.03	-0.04	0.06	0.02
	$se(\hat{\beta}_j)$	0.05	0.01	0.01	0.02	0.03	0.05	0.02
	t-value	1.57	3.04	3.50	1.56	-1.48	1.37	1.42
k_6	$\hat{\beta}_j$	0.09	0.02	0.04	0.03	-0.03	0.54	0.03
	$se(\hat{\beta}_j)$	0.05	0.01	0.01	0.02	0.03	0.25	0.02
	t-value	1.63	3.32	3.07	1.35	-0.96	2.22	1.45
k_7	$\hat{\beta}_j$	0.09	0.02	0.04	0.03	-0.03	0.54	0.03
	$se(\hat{\beta}_j)$	0.05	0.01	0.01	0.02	0.03	0.27	0.02
	t-value	1.63	3.30	3.07	1.36	-0.97	2.02	1.45
k_8	$\hat{\beta}_j$	0.08	0.02	0.05	0.03	-0.04	0.23	0.02
	$se(\hat{\beta}_j)$	0.05	0.01	0.01	0.02	0.03	0.14	0.02
	t-value	1.62	3.13	3.37	1.50	-1.33	1.70	1.41
k_9	$\hat{\beta}_j$	0.09	0.02	0.04	0.03	-0.03	0.54	0.03
	$se(\hat{\beta}_j)$	0.05	0.01	0.01	0.02	0.03	0.26	0.02
	t-value	1.63	3.31	3.07	1.35	-0.97	2.12	1.45
k_{10}	$\hat{\beta}_j$	0.09	0.02	0.04	0.03	-0.03	0.46	0.03
	$se(\hat{\beta}_j)$	0.05	0.01	0.01	0.02	0.03	0.22	0.02
	t-value	1.62	3.26	3.16	1.40	-1.09	2.12	1.43

Note: $\hat{\beta}_j$ is the j th value (from 1, ..., p) of the parameter vector $\hat{\beta}$.

have a positive affect on the probability of receiving diabetes. This result holds for both the probit and the logit models. We also see that a decrease of the bootstrapped standard errors for all variables which clearly shows the benefit of applying ridge regression methods. We also see an increase of the t-statistics for most cases which further shows the usefulness of applying the ridge regression estimation technique. An example of such a case is X_3 for which the variable is insignificant of both the probit and logit model when using ML. However, when we use ridge regression (besides the classical k_1 estimator suggested by [5, 6] it often becomes significant. Hence, the general result shows that the ridge regression may reduce the standard errors of the estimated parameters. The newly proposed ridge regression estimators that are combinations of the previous ridge parameter performs well and show a greater stability than the old ones. Especially the new ridge regression estimator k_g is always either the best (with the lowest standard error) or close to the optimal ridge parameter when looking at the cases that minimizes the standard errors and maximizes the t-statistics. Hence, the result from this example shows the benefit of the ridge regression estimator. However, a much more thorough investigation is needed in order to give conclusive evidence of to what degree these risk factors effect the probability of receiving diabetes.

5. SUMMARY AND CONCLUDING REMARKS

This paper proposed some new estimators for estimating the ridge parameter for binary choice models. These new estimators are investigated along with some of the existing popular estimators using Monte Carlo simulations. In the design of the experiment we mainly focuses on the effect of multicollinearity and at the same time we also change the number of observations and the number of explanatory variables. The results from the simulation study clearly indicate that the ridge regression estimators outperform ML estimator in the sense of smaller MSE criterion. Furthermore, it is also shown that the proposed estimators have in general better MSE properties than the rest estimators from the previous research. This is in particular true in the presence of high degree of collinearity between the explanatory variables and when there are many explanatory variables included in the regression model. The benefit of ridge regression techniques is then shown by an empirical example where it is investigated how some risk factors affects the probability of receiving diabetes.

REFERENCES

- [1] Van Den Noortgate W, De Boeck P, Meulders M. Cross-classification multilevel logistic models in psychometrics. *J Educat Behav Statist* 2003; 28: 369-86. <http://dx.doi.org/10.3102/10769986028004369>
- [2] Cameron AC, Trivedi PK. *Microeconometrics: Methods and Applications*. Cambridge University Press 2005. <http://dx.doi.org/10.1017/CBO9780511811241>
- [3] Schaefer RL, Roi LD, Wolfe RA. A ridge logistic estimator. *Commun Statist Theory Methods* 1984; 13: 99-113. <http://dx.doi.org/10.1080/03610928408828664>
- [4] Månsson K, Shukur G. On ridge parameters in logistic regression. *Commun Statist Theory Methods* 2011a; 40: 3366-81. <http://dx.doi.org/10.1080/03610926.2010.500111>
- [5] Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* 1970a; 12: 55-67. <http://dx.doi.org/10.1080/00401706.1970.10488634>
- [6] Hoerl AE, Kennard RW. Ridge regression: application to non-orthogonal problems. *Technometrics* 1970b; 12: 69-82. <http://dx.doi.org/10.1080/00401706.1970.10488635>
- [7] Hoerl AE, Kennard RW, Baldwin KF. Ridge regression: some simulation. *Commun Statist Theory Methods* 1975; 4: 105-23. <http://dx.doi.org/10.1080/03610927508827232>
- [8] McDonald GC, Galameau DI. A Monte Carlo evaluation of some ridge-type estimators. *J Am Statist Assoc* 1975; 70: 407-16. <http://dx.doi.org/10.1080/01621459.1975.10479882>
- [9] Lawless JF, Wang P. A simulation study of ridge and other regression estimators. *Commun Statist Theory Methods* 1976; 5: 307-23. <http://dx.doi.org/10.1080/03610927608827353>
- [10] Gibbons DG. A simulation study of some ridge estimators. *J Am Statist Assoc* 1981; 76: 131-39. <http://dx.doi.org/10.1080/01621459.1981.10477619>
- [11] Kibria BMG. Performance of some new ridge regression estimators. *Commun Statist Theory Methods* 2003; 32: 419-35.
- [12] Khalaf G, Shukur G. Choosing ridge parameters for regression problems. *Commun Statist Theory Methods* 2005; 34: 1177-82. <http://dx.doi.org/10.1081/STA-200056836>
- [13] Alkhamisi MA, Khalaf G, Shukur G. Some modifications for choosing ridge parameter. *Commun Statist Theory Methods* 2006; 35: 1-16. <http://dx.doi.org/10.1080/03610920600762905>
- [14] Alkhamisi M, Shukur G. Developing ridge parameters for SUR model. *Commun Statist Theory Methods* 2008; 37: 544-64. <http://dx.doi.org/10.1080/03610920701469152>
- [15] Muniz G, Kibria BMG. On some ridge regression estimators: An empirical comparisons. *Commun Statist Simulat Comput* 2009; 38: 621-30. <http://dx.doi.org/10.1080/03610910802592838>
- [16] Arashi M, Valizadeh T. Performance of Kibria's methods in partial linear ridge regression model. *To Appear in Statistical Papers* 2014.
- [17] Aslam M. Performance of Kibria's method for the heteroscedastic ridge regression model: Some Monte Carlo evidence. *Commun Statist Simulat Comput* 2014; 43: 673-86. <http://dx.doi.org/10.1080/03610918.2012.712185>

- [18] Hefnawya AE, Faraga A. A combined nonlinear programming model and Kibria method for choosing ridge parameter regression. *Commun Statist Simulat Comput* 2014; 43: 1442-70.
<http://dx.doi.org/10.1080/03610918.2012.735317>
- [19] Kibria BMG, Månsson K, Shukur G. Performance of some logistic ridge regression estimators. *Comput Econom* 2011; 40: 401-14.
<http://dx.doi.org/10.1007/s10614-011-9275-x>
- [20] Smith JW, Everhart JE, Dickson WC, Knowler WC, Johannes RS. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Symposium on Computer Applications in Medical Care*. In *Proceedings of the Symposium on Computer Applications in Medical Care* Computer Society Press 1988.

Received on 20-04-2014

Accepted on 31-05-2014

Published on 05-08-2014

<http://dx.doi.org/10.6000/1929-6029.2014.03.03.5>