Research Journal of Mathematics and Statistics 9(2): 2017 9(2): 11-19, 2017
DOI:10.19026/rjms.9.4715
ISSN: 2042-2024, e-ISSN: 2040-7505

Submitted: November 15, 2016 Accepted: February 13, 2017 Published: November 25, 2017

Research Article

Evaluating the Performance of Liu Logistic Regression Estimator

Rasha A. Farghali and Samah M. Abo-El-Hadid
Department of Mathematics, Insurance and Applied Statistics, Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt

Abstract: This study aims at comparing the performance of logistic Liu estimators with Maximum Likelihood (ML), Stien and ridge regression estimators using a Monte Carlo simulation, where the mean squared/absolute errors, $MSE(\beta)/\ MAE(\beta)$ mean squared/absolute error between the actual probability $\pi(x)$ and the estimated probability $\hat{\pi}(x)$, $MSE\big(\pi(x)\big)/\ MAE\big(\pi(x)\big)$ are used as performance criteria. An algorithm for simulation steps is included. An application of the effect of quantities of household wastes and its components on the probability of getting a running waste recycling factory is analyzed. Results from both the simulation and the application show that logistic Liu estimators are mostly preferred for correcting mutilcollinearity in logistic regression.

Keywords: Biased estimators, Liu estimators, logistic regression, multicollinearity, ridge regression estimators, stien estimators

INTRODUCTION

Logistic regression, which is considered a member of the generalized linear models family, allows one to predict a discrete outcome. Generally, the dependent or response variable y is dichotomous, such as presence/absence, success/failure,…, etc. whereas the independent or explanatory variables $x_1, x_2, \dots, x_p$ may be continuous, discrete, dichotomous, or a mix of these variables. The relationship between y and $x_1, x_2, \dots, x_p$ is estimated using maximum likelihood method. Maximum likelihood estimates (MLE) have minimum variances, but in the presence of multicollinearity they were inflated and have large variances.

Consequently, confidence intervals for regression parameters become wider and in testing hypothesis insignificant parameters are obtained (Schaefer *et al*., 1984; Schaefer, 1986; Agresti, 2002; Hosmer and Lemeshow, 2002) among others.

Biased estimators such as: Stien, ridge regression and Liu estimators were introduced for correcting multicollinearity. These biased estimators have a common advantage that is; all the explanatory variables are considered simultaneously without any reduction to improve the accuracy (Belsley *et al*., 1980; Belsley, 1991).

These biased estimators were first used for correcting multicollinearity in linear regression (Hoerl and Kennard, 1970 a, b; Dielman, 2005; Farag *et al*., 2012; Hamed *et al*., 2013; Rong, 2010). They aimed at achieving two goals:

- Reducing the Mean Squared Errors (MSEs) for the estimates of the parameters and
- Improving the conditioning of the information matrix, so that the obtained parameter estimates and their standard errors are smaller than ML estimates.

Stien estimators achieved the first goal but, it has a disadvantage which is, the shrinkage parameter d_{st} (James and Stien, 1961), was calculated using MLE, which is already inflated as a result of multicollinearity, so that Stien estimators and their standard errors are inflated. Ridge regression estimates have achieved the second goal but, they have a disadvantage that there is still no consensus regarding how to select the ridge parameter d_{ridge} (Le Cessie and Van Houwelingen, 1992; Kibria *et al*., 2012; Farag *et al*., 2012; Kan *et al*., 2013). Liu estimators were introduced to combine two different methods (Stien estimators and ridge regression estimators) to obtain the advantages of both estimators and avoid their disadvantages (Liu, 1993; Liu, 2003, 2004; Akdeniz and Erol, 2003; Rong, 2010).

The pervious estimators were extended to correct multicollinearity in binomial logistic regression and multinomial logistic regression but their performance in comparison to each other was not studied (Steyerberg *et al*., 2001; Aguilera *et al*., 2006; Camminatiello and Lucadamo, 2010; Farghali, 2012, 2014; Asar and Genc, 2016).

Corresponding Author: Rasha A. Farghali, Department of Mathematics, Insurance and Applied Statistics, Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt, Tel.: +20101626825

This study aims to evaluate the performance of logistic Liu estimators in comparison with MLE, Stien and ridge regression estimators using a Monte Carlo simulation, where the mean squared errors of parameters $MSE(\beta)$, mean absolute errors of parameters $MAE(\beta)$, mean squared error between the actual probability $\pi(x)$ and the estimated probability $\hat{\pi}(x)$, $MSE(\pi(x))$; mean absolute error between the actual probability $\pi(x)$ and the estimated probability $\hat{\pi}(x)$, $MAE(\pi(x))$ are used as performance criteria and also in the simulation study, factors including the degree of correlation, the sample size and the number of explanatory variables are varied. The estimator with the lowest standard errors and with the minimum $(MSE(\beta), MAE(\beta), MSE(\pi(x)), MAE(\pi(x)))$ is considered the best option for correcting multicollinearity in logistic regression.

Finally, the benefits of using logistic Liu estimator are shown using the data of municipal solid waste management in Egypt, where the effect of quantities of household wastes and its components (paper packing, plastic, glass and metal) on the probability of getting a running waste recycling factory is investigated.

METHODOLOGY

This section describes the binomial logistic regression model and the effect of multicollinearity on the parameters estimates and on its standard errors. Furthermore, different biased estimators are presented to correct multicollinearity in logistic regression.

The binomial logistic regression model: Let the real relationship between the response variable y and the explanatory variables $x_1, x_2, \dots, x_p$ be as follows (Hosmer and Lemeshow, 2002; Agresti, 2002):

$$y_i = \frac{e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}} + \varepsilon_i \tag{1}$$

where,
i = 1, 2 …, n
n = Sample size
p = Number of the explanatory variables
x_{ij} = The measurement of the j^th explanatory variable for the i^th observation, $i = 1,2,\dots,n, j = 1,2,\dots,p$.
β_j = the j^th regression parameters, $j = 1,2,\dots,p$
ε_i = random error for the i^th observation, $i = 1,2,\dots,n$:

$$y_i = \begin{cases} 1 \text{ the i}^{\text{th}} \text{ observation has the property under} \\ \quad \text{consideration i} = 1,2,\dots,\text{n} \\ \quad 0 \quad \text{otherwise} \end{cases} \tag{2}$$

The fitted logistic regression model is as follows:

$$ln\left[\frac{\hat{\pi}_1(x)}{\hat{\pi}_0(x)}\right] = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \;\; i = 1,2,\dots,n \tag{3}$$

It is well known that the Maximum Likelihood Estimates (MLE) $\hat{\beta}_{MLE}$ for the logistic regression model are obtained by solving the following nonlinear system of equations using numerical methods as presented in Hosmer and Lemeshow (2002) and Agresti (2002):

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \left(\frac{e^{\hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}}}{1 + e^{\hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}}} \right) \tag{4}$$

$$\sum_{i=1}^{n} y_i x_{ij} = \sum_{i=1}^{n} \left(\frac{e^{\hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}}}{1 + e^{\hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}}} \right) \cdot x_{ij} \tag{5}$$

where,
i = 1, 2 …, n
j = 1, 2, …, p

And the covariance matrix is calculated as follows:

$$Var(\hat{\beta}_{MLE}) \approx (\grave{X}\hat{W}X)^{-1} \tag{6}$$

where, $(\grave{X}\hat{W}X)$ is the estimated weighted information matrix of order $(p+1) \times (p+1)$. X is a matrix of order $(n \times (p+1))$ that consists of the measurements of explanatory variables for each observation at each level of the response variable. $\hat{W}$ is a diagonal matrix of order $(n \times n)$ its general element $\hat{\pi}_1(x_i)(1-\hat{\pi}_1(x_i))$, $i = 1,2,\dots,n$.

$\hat{\beta}_{MLE}$ are unbiased estimators with minimum variances when the explanatory variables $x_1, x_2, \dots, x_p$ are uncorrelated, but often produce poor results because of the multicollinearity problem among the explanatory variables (Lesaffre and Marx, 1993; Tutz and Lieitenstorfer, 2006). Multicollinearity may produce signs opposite to the true signs of paired correlations and yields theoretically important variables with insignificant coefficients. Also, it affects the ability of prediction, wider confidence intervals and incorrect decisions for testing hypotheses for the regression parameters (Agresti, 2002; Hosmer and Lemeshow, 2002; Månsson *et al*., 2012, 2015).

Multicollinearity in binomial logistic regression: Many studies have been introduced for correcting multicollinearity in logistic regression models (Schaefer *et al*., 1984; Schaefer, 1986; Steyerberg *et al*., 2001; Aguilera *et al*., 2006; Camminatiello and Lucadamo, 2010; Farghali, 2012, 2014; Asar and Genc, 2016). They developed the methods that were used for correcting multicollinearity in linear regression (such as: Stien estimators, ridge regression estimators, principal components regression, Liu estimators and mathematical programming). Most of these studies evaluated the performance of a single method by comparing it with the MLE but not in comparison with other methods (Månsson and Shukur, 2011; Månsson *et al*., 2012, 2015).

In this study, we are concerned about evaluating the performance of logistic Liu estimators (with two different shrinkage parameters) in comparison with: MLE, logistic Stien estimators and logistic ridge regression estimators.

In Farghali (2012), the logistic Stien and logistic ridge regression estimators were introduced as follows:

$$\hat{\beta}_{St} = d_{St}\ \hat{\beta}_{MLE} \tag{7}$$

where, $0 < d_{St} < 1$

And the estimated standard errors $SE(\hat{\beta}_j)_{St}$ were calculated as follows:

$$SE(\hat{\beta}_j)_{St} = d_{St}\ SE(\hat{\beta}_j)_{MLE} \tag{8}$$

The shrinkage parameter for logistic Stien estimators was calculated as follows:

$$d_{St} = \frac{\hat{\beta}'_{MLE}\hat{\beta}_{MLE}}{\hat{\beta}'_{MLE}\hat{\beta}_{MLE} + tr(\grave{X}\widehat{W}X)^{-1}} \tag{9}$$

The logistic ridge regression estimators were as follows:

$$\hat{\beta}_R = (X'\hat{W}X + \hat{d}_{ridge}I)^{-1}(X'\hat{W}X)\hat{\beta}_{MLE} \tag{10}$$

where, $0 \le \hat{d}_{ridge} < 1$

The estimated covariance matrix was as follows:

$$Var(\hat{\beta}_R) = (\grave{X}\widehat{W}X + \hat{d}_{ridge}I)^{-1}(\grave{X}\widehat{W}X)(\grave{X}\widehat{W}X + \hat{d}_{ridge}I)^{-1} \tag{11}$$

And the ridge parameter $\hat{d}_{ridge}$ for logistic ridge regression was calculated as follows:

$$\hat{d}_{ridge} = \frac{p}{\hat{\beta}'_{MLE}\hat{\beta}_{MLE}} \tag{12}$$

Logistic Stien estimators have a disadvantage that the shrinkage parameter d_{St} is calculated using MLE, which is already inflated as a result of multicollinearity, so that Stien estimators and its standard errors are inflated. Also, Logistic ridge regression estimators have a disadvantage: there is still no consensus regarding how to select the ridge parameter d_{ridge} (El-Dash *et al*., 2011; Hamed *et al*., 2013; Farghali, 2012).

Logistic Liu estimator: The hope that the combination of two different methods (Stien estimators and ridge regression estimators) might inherit the advantages of both estimators and avoid their disadvantages motivated Liu (2003, 2004) and Liu (1993), to suggest another biased estimator for correcting multicollinearity in linear regression.

Månsson *et al*. (2012) suggested logistic Liu estimators to correct multicollinearity in binomial logistic regression. The logistic Liu estimators were as follows:

$$\hat{\beta}_{Liu} = (X'\hat{W}X + I)^{-1}(X'\hat{W}X + \hat{d}_{Liu}I)\hat{\beta}_{MLE} \tag{13}$$

where, $0 \le \hat{d}_{Liu} \le 1$

Also, they suggested different methods for estimating the shrinkage parameter d_{Liu}, one of these methods was as follows:

$$\hat{d}_{Liu} = Max\left(0, Min\left(\frac{\hat{\gamma}_j^2 - 1}{\frac{1}{q_j} + \hat{\gamma}_j^2}\right)\right) \tag{14}$$

where,

$$\hat{\gamma}_j = \sum_{t=1}^{p} v_{jt}\ (\hat{\beta}_j)_{MLE} \tag{15}$$

q_j is the jth eigenvalue of the standardized weighted information matrix $(X^{*'}\widehat{W}X^{*'})$, $j = 1,2,\ldots,p$. v_j is the jth eigenvector corresponding to the jth eigenvalue of the standardized estimated weighted information matrix, $(X^{*'}\widehat{W}X^{*'})$, $j = 1,2,\ldots,p$.

The shrinkage parameter $\hat{d}_{Liu}$ in Eq. (14) was estimated in two steps: first, they calculated the value of each individual parameter $\hat{d}_j$ as follows:

$$\hat{d}_j = \frac{\hat{\gamma}_j^2 - 1}{\frac{1}{q_j} + \hat{\gamma}_j^2} \tag{16}$$

Second, based on a simulation study, they reduced the p values obtained in (16) to a single value $\hat{d}_{Liu}$ as shown in (14).

The disadvantages of Månsson *et al*. (2012) were:

- They did not introduce an exact method for determining a single value $\hat{d}_{Liu}$ and
- They did not study the performance of the suggested estimator in comparison to other biased estimators.

Farghali (2014) suggested multinomial logistic Liu estimators to correct multicollinearity in multinomial logistic regression. Following Liu (1993), the estimated shrinkage parameter $\hat{d}_{Liu}$ was chosen to minimize the mean squared error of the parameters:

$$MSE_{Liu} = E\left[(\hat{\beta}_{Liu} - \beta)'(\hat{\beta}_{Liu} - \beta)\right] \tag{17}$$

So that, the estimated shrinkage parameter for Liu biased estimators was as follows:

$$\left(\hat{d}_{Liu}\right)_I = \frac{\sum_{K=0}^{C-1}\sum_{j=1}^{p}\left[\left(\hat{\gamma}_{Kj}^2-1\right)/\left(q_{Kj}+1\right)^2\right]}{\sum_{K=0}^{C-1}\sum_{j=1}^{p}\left[\left(q_{Kj}\hat{\gamma}_{Kj}^2+1\right)/q_{Kj}\left(q_{Kj}+1\right)^2\right]} \tag{18}$$

And the estimated covariance matrix was as follows:

$$Var\left(\hat{\beta}_{Liu}\right) = \left(\acute{X}\widehat{W}X + I\right)^{-1}\left(\acute{X}\widehat{W}X + \left(\hat{d}_{Liu}\right)_I I\right)\left(\acute{X}\widehat{W}X\right)^{-1}\left(\acute{X}\widehat{W}X + \left(\hat{d}_{Liu}\right)_I I\right)\left(\acute{X}\widehat{W}X + I\right)^{-1} \tag{19}$$

Thus, in Farghali (2014) a single value of $\left(\hat{d}_{Liu}\right)_I$ was obtained and $Var\left(\hat{\beta}_{Liu}\right)$ was introduced. Also, she extended Månsson *et al*. (2012) to correct multicollinearity in multinomial logistic regression.

The disadvantage of Farghali (2014) was that the performance of the suggested biased estimator was studied only by a set of hypothetical data.

In this study, logistic Liu estimator with $\left(\hat{d}_{Liu}\right)_I$ is introduced as a special case of the multinomial logistic Liu estimators by putting $C = 2$ in Eq. (18) and (19), we obtained the estimated shrinkage parameter $\left(\hat{d}_{Liu}\right)_I$ and the estimated covariance matrix $Var\left(\hat{\beta}_{Liu}\right)$ for binomial logistic regression. Simulation studies were conducted that evaluated the performance of logistic Liu estimator with both ($\hat{d}_{Liu}$ and $\left(\hat{d}_{Liu}\right)_I$) in comparison with MLE, logistic Stien estimators and logistic ridge regression estimators. Furthermore, the logistic Liu estimators were applied to a real-life dataset.

Judging the performance of the estimators: To investigate the performance of logistic Liu estimators in comparison with MLE, logistic Stien estimators and logistic ridge regression estimators we calculate ($MSE(\beta)$, $MAE(\beta)$, $MSE\left(\pi(x)\right)$, $MAE\left(\pi(x)\right)$) using the following equations:

$$MSE(\beta) = \frac{\sum_{i=1}^{R}\left(\hat{\beta}-\beta\right)_i'\left(\hat{\beta}-\beta\right)}{R} \tag{20}$$

$$MAE(\beta) = \frac{\sum_{i=1}^{R}\left|\hat{\beta}-\beta\right|_i}{R} \tag{21}$$

$$MSE\left(\pi(x)\right) = \frac{\sum_{i=1}^{R}\left(\hat{\pi}(x)-\pi(x)\right)_i'\left(\hat{\pi}(x)-\pi(x)\right)}{R} \tag{22}$$

$$MAE\left(\pi(x)\right) = \frac{\sum_{i=1}^{R}\left|\hat{\pi}(x)-\pi(x)\right|_i'}{R} \tag{23}$$

where, $\hat{\beta}$ is the estimator of β obtained from MLE, logistic Stien estimators and logistic ridge regression estimators and logistic Liu estimators with both ($\hat{d}_{Liu}$ and $\left(\hat{d}_{Liu}\right)_I$) and R equals 2000 which corresponds to the number of replicates used in the Monte Carlo simulation.

Monte carlo simulation: This section consists of a brief description of how the data are generated together with a result discussion.

The design of the experiment: The response variable of the logistic regression model is generated using pseudo-random numbers from the $Be\left(\pi(x_i)\right)$ distribution where:

$$\pi(x_i) = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \tag{24}$$

The parameter values of β are chosen so that $\beta'\beta = 1$ and $\beta_1 = \beta_2 = \cdots = \beta_p$ (Månsson and Shukur, 2011). To be able to generate data with different degrees of correlation, we use the following formula:

$$x_{ij} = (1-\rho^2)^{1/2} z_{ij} + \rho z_{ip} \tag{25}$$

where, z_{ij} are pseudo-random numbers generated using the standard normal distribution and ρ^2 represents the degree of correlation. In the design of the experiment, three different values of ρ are considered $\rho = 0.75, 0.85$ and 0.95.

The other factors that varied in the simulation study are the values of n and p. we use sample sizes corresponding to 50, 70, 100, 150 and 200 observations and regression models including 2 and 3 explanatory variables.

The proposed algorithm:

Step 1: Set sample size n; the total number of experiments R; number of the explanatory variables p; and the parameters β .

Step 2 : Let $r = 1$.

Step 3 : Generate data with different degrees of correlation according to formula (25).

Step 4 : The maximum likelihood estimates (MLE) $\hat{\beta}_{MLE}$ for the logistic regression model are obtained through solving nonlinear system of Eq. (4) and (5).

Step 5 : The shrinkage parameter for logistic Stein estimators is estimated using Eq. (9).

Step 6 : Logistic Stein estimators were obtained using formula (7).

Step 7 : Ridge parameter $\hat{d}_{ridge}$ for logistic ridge regression was calculated according to (12).

Step 8 : Logistic ridge regression estimators were calculated using formula (10).

Step 9 : The logistic Liu shrinkage parameters were estimated as in Eq. (14 and 18).

Step 10: The logistic Liu estimators were obtained using formula (13).

Step 11: Calculate mean squared errors of parameters $MSE(\beta)$, mean absolute errors of parameters

$MAE(\beta)$, mean squared error between the actual probability and the estimated probability $MSE(\pi(x))$; mean absolute error between the actual probability and the estimated probability $MAE(\pi(x))$ as in Eq. (20)-(23).

Step 12: Set $r = r + 1$, if $r = R$, stop.

RESULTS AND DISCUSSION

The simulated $MSE(\beta)$, $MAE(\beta)$, $MSE(\pi(x))$ and $MAE(\pi(x))$ for all of the estimators for different ρ, n and p are presented in Table 1 and 2.

From Table 1, in the case of high correlation coefficients (ρ = 0.75 and 0.85) with small and

Table 1: Simulation results when $p = 2$

ρ	n	Estimator	$MSE(\beta)$	$MAE(\beta)$	$MSE(\pi)$	$MAE(\pi)$
0.75	50	MLE	0.2498	0.3803	0.0079	0.0659
		Stein	0.1945	0.3514	0.0104	0.0777
		Ridge	0.1565	0.3106	0.0063	0.0591
		Liu ($\hat{d}_{Liu}$)	0.1244	0.2783	0.0048	0.0506
		Liu $(\hat{d}_{Liu})_I$	0.1632	0.3130	0.0057	0.0568
	70	MLE	0.1488	0.3004	0.0052	0.0537
		Stein	0.1269	0.2872	0.0073	0.0649
		Ridge	0.1041	0.2562	0.0043	0.0491
		Liu ($\hat{d}_{Liu}$)	0.0936	0.2433	0.0035	0.0439
		Liu $(\hat{d}_{Liu})_I$	0.1088	0.2600	0.0041	0.0479
	100	MLE	0.1057	0.2525	0.0039	0.046
		Stein	0.0938	0.2459	0.0051	0.0539
		Ridge	0.0801	0.2242	0.0032	0.0422
		Liu ($\hat{d}_{Liu}$)	0.0758	0.2185	0.0028	0.0393
		Liu $(\hat{d}_{Liu})_I$	0.0847	0.2283	0.0032	0.0419
	150	MLE	0.0636	0.1985	0.0024	0.0362
		Stein	0.0582	0.1928	0.0032	0.0425
		Ridge	0.0514	0.1799	0.0021	0.034
		Liu ($\hat{d}_{Liu}$)	0.0515	0.1801	0.0019	0.0326
		Liu $(\hat{d}_{Liu})_I$	0.0542	0.1839	0.0021	0.0339
	200	MLE	0.0464	0.1702	0.0018	0.0318
		Stein	0.0442	0.1682	0.0023	0.0362
		Ridge	0.0398	0.1589	0.0016	0.0302
		Liu ($\hat{d}_{Liu}$)	0.0401	0.1591	0.0016	0.0293
		Liu $(\hat{d}_{Liu})_I$	0.0414	0.1613	0.0016	0.0300
0.85	50	MLE	0.3211	0.4437	0.0074	0.064
		Stein	0.2390	0.3961	0.0133	0.0897
		Ridge	0.1748	0.3286	0.0052	0.0533
		Liu ($\hat{d}_{Liu}$)	0.1429	0.2991	0.0037	0.0452
		Liu $(\hat{d}_{Liu})_I$	0.1852	0.3363	0.0041	0.0486
	70	MLE	0.2201	0.3655	0.0053	0.0534
		Stein	0.1753	0.3368	0.0106	0.0800
		Ridge	0.1331	0.2848	0.0038	0.0455
		Liu ($\hat{d}_{Liu}$)	0.1209	0.2740	0.0030	0.0402
		Liu $(\hat{d}_{Liu})_I$	0.1434	0.2946	0.0032	0.0423
	100	MLE	0.1347	0.2883	0.0036	0.0441
		Stein	0.1159	0.2747	0.0076	0.0674
		Ridge	0.0908	0.2378	0.0027	0.0379
		Liu ($\hat{d}_{Liu}$)	0.0891	0.2368	0.0022	0.0348
		Liu $(\hat{d}_{Liu})_I$	0.0986	0.2474	0.0024	0.0365
	150	MLE	0.0895	0.2388	0.0023	0.0355
		Stein	0.0806	0.2308	0.0049	0.0541
		Ridge	0.0657	0.2051	0.0018	0.0311
		Liu ($\hat{d}_{Liu}$)	0.0679	0.2093	0.0016	0.0296
		Liu $(\hat{d}_{Liu})_I$	0.0716	0.2138	0.0017	0.0306
	200	MLE	0.0671	0.2049	0.0017	0.0304
		Stein	0.0617	0.1994	0.0036	0.0466
		Ridge	0.0523	0.1812	0.0013	0.0272
		Liu ($\hat{d}_{Liu}$)	0.0545	0.1857	0.0013	0.0264
		Liu $(\hat{d}_{Liu})_I$	0.0565	0.1884	0.0013	0.0271
0.95	50	MLE	0.8177	0.6932	0.0072	0.0623
		Stein	0.5254	0.5752	0.0254	0.1287
		Ridge	0.3385	0.4184	0.0043	0.0475
		Liu ($\hat{d}_{Liu}$)	0.2488	0.3622	0.0026	0.0366
		Liu $(\hat{d}_{Liu})_I$	0.2748	0.3839	0.0027	0.0374
	70	MLE	0.5027	0.5494	0.0053	0.0533
		Stein	0.3536	0.4772	0.0229	0.1222
		Ridge	0.2211	0.3437	0.0032	0.0409
		Liu ($\hat{d}_{Liu}$)	0.1815	0.3204	0.0019	0.0317

Table 1: Continue

ρ	n	Estimator	$MSE(\beta)$	$MAE(\beta)$	$MSE(\pi)$	$MAE(\pi)$
		Liu $(\hat{d}_{Liu})_I$	0.1922	0.3318	0.0020	0.0322
	100	MLE	0.3444	0.4610	0.0036	0.0435
		Stein	0.2622	0.4162	0.0210	0.1174
		Ridge	0.1666	0.3046	0.0022	0.0339
		Liu $(\hat{d}_{Liu})$	0.1538	0.3019	0.0014	0.0277
		Liu $(\hat{d}_{Liu})_I$	0.1596	0.3080	0.0014	0.0280
	150	MLE	0.2180	0.3679	0.0022	0.0345
		Stein	0.1781	0.3416	0.0179	0.1088
		Ridge	0.1196	0.2629	0.0014	0.0270
		Liu $(\hat{d}_{Liu})$	0.1208	0.2711	0.0009	0.0227
		Liu $(\hat{d}_{Liu})_I$	0.1264	0.2772	0.0009	0.0228
	200	MLE	0.1609	0.3196	0.0018	0.0308
		Stein	0.1350	0.2989	0.0150	0.0997
		Ridge	0.0935	0.2381	0.0011	0.0240
		Liu $(\hat{d}_{Liu})$	0.0986	0.2501	0.0008	0.0206
		Liu $(\hat{d}_{Liu})_I$	0.1021	0.2537	0.0008	0.0207

Table 2: Simulation results when $p = 3$

ρ	n	Estimator	$MSE(\beta)$	$MAE(\beta)$	$MSE(\pi)$	$MAE(\pi)$
0.75	50	MLE	0.6485	0.6045	0.0116	0.0803
		Stein	0.3790	0.4694	0.0149	0.0959
		Ridge	0.2546	0.3647	0.0073	0.0634
		Liu $(\hat{d}_{Liu})$	0.1716	0.3259	0.0055	0.0552
		Liu $(\hat{d}_{Liu})_I$	0.2763	0.3646	0.0069	0.0607
	70	MLE	0.3887	0.4883	0.0082	0.0673
		Stein	0.2442	0.3952	0.0108	0.0815
		Ridge	0.1675	0.3111	0.0052	0.0537
		Liu $(\hat{d}_{Liu})$	0.1543	0.3141	0.0046	0.0504
		Liu $(\hat{d}_{Liu})_I$	0.1671	0.3045	0.0049	0.0514
	100	MLE	0.2455	0.3902	0.0056	0.0553
		Stein	0.1709	0.3318	0.0079	0.0696
		Ridge	0.1199	0.2661	0.0037	0.0458
		Liu $(\hat{d}_{Liu})$	0.1284	0.2871	0.0036	0.0450
		Liu $(\hat{d}_{Liu})_I$	0.1172	0.2597	0.0035	0.0442
	150	MLE	0.1511	0.3064	0.0037	0.0448
		Stein	0.1153	0.2740	0.0052	0.0560
		Ridge	0.0843	0.2259	0.0026	0.0384
		Liu $(\hat{d}_{Liu})$	0.0998	0.2510	0.0027	0.0390
		Liu $(\hat{d}_{Liu})_I$	0.0822	0.2210	0.0025	0.0375
	200	MLE	0.1130	0.2647	0.0028	0.0387
		Stein	0.0916	0.2422	0.0039	0.0484
		Ridge	0.0691	0.2047	0.0021	0.0340
		Liu $(\hat{d}_{\mathrm{Liu}})$	0.0829	0.2281	0.0022	0.0349
		Liu$(\hat{d}_{\mathrm{Liu}})_{\mathrm{I}}$	0.0670	0.1997	0.0020	0.0334
0.85	50	MLE	1.4723	0.9316	0.0109	0.0771
		Stein	0.7304	0.6449	0.0187	0.1094
		Ridge	0.4663	0.4739	0.0057	0.0557
		Liu $(\hat{d}_{Liu})$	0.1911	0.3286	0.0036	0.0443
		Liu $(\hat{d}_{Liu})_I$	0.4957	0.4721	0.0054	0.0532
	70	MLE	0.9335	0.7487	0.0078	0.0651
		Stein	0.5056	0.5481	0.0156	0.1012
		Ridge	0.3274	0.4007	0.0043	0.0484
		Liu $(\hat{d}_{Liu})$	0.1819	0.3302	0.0031	0.0415
		Liu $(\hat{d}_{Liu})_I$	0.3309	0.3903	0.0040	0.0461
	100	MLE	0.5756	0.5941	0.0053	0.0533
		Stein	0.3384	0.4627	0.0123	0.0899
		Ridge	0.2151	0.3358	0.0030	0.0407
		Liu $(\hat{d}_{Liu})$	0.1653	0.3220	0.0025	0.0373
		Liu $(\hat{d}_{Liu})_I$	0.2120	0.3258	0.0028	0.0388
	150	MLE	0.3712	0.4787	0.0035	0.0436
		Stein	0.2401	0.3893	0.0085	0.0753
		Ridge	0.1556	0.2894	0.0021	0.0337
		Liu $(\hat{d}_{Liu})$	0.1515	0.3085	0.0020	0.0332
		Liu $(\hat{d}_{Liu})_I$	0.1485	0.2781	0.0020	0.0323
	200	MLE	0.2768	0.4173	0.0028	0.0384
		Stein	0.1905	0.3523	0.0066	0.0665
		Ridge	0.1260	0.2676	0.0017	0.0307

Table 2: Continue

ρ	n	Estimator	$MSE(\beta)$	$MAE(\beta)$	$MSE(\pi)$	$MAE(\pi)$
		Liu ($\hat{d}_{Liu}$)	0.1393	0.2977	0.0017	0.0311
		Liu $(\hat{d}_{Liu})_I$	0.1207	0.2569	0.0016	0.0295
0.95	50	MLE	13.1243	2.7574	0.0110	0.0762
		Stein	5.6883	1.5957	0.0269	0.1333
		Ridge	3.7743	1.2215	0.0052	0.0515
		Liu ($\hat{d}_{Liu}$)	0.8743	0.4629	0.0028	0.0378
		Liu $(\hat{d}_{Liu})_I$	3.8994	1.1837	0.0049	0.0490
	70	MLE	8.3497	2.2426	0.0079	0.0647
		Stein	3.6582	1.3250	0.0246	0.1274
		Ridge	2.5085	1.0176	0.0037	0.0442
		Liu ($\hat{d}_{Liu}$)	0.6333	0.4187	0.0021	0.0335
		Liu $(\hat{d}_{Liu})_I$	2.5484	0.9826	0.0036	0.0424
	100	MLE	5.0464	1.7618	0.0053	0.0532
		Stein	2.1919	1.0626	0.0236	0.1256
		Ridge	1.4948	0.7978	0.0026	0.0365
		Liu ($\hat{d}_{Liu}$)	0.3391	0.3472	0.0015	0.0285
		Liu $(\hat{d}_{Liu})_I$	1.4688	0.3472	0.0024	0.0352
	150	MLE	3.0670	1.3780	0.0035	0.0429
		Stein	1.3718	0.8670	0.0216	0.1206
		Ridge	0.9271	0.6333	0.0017	0.0296
		Liu ($\hat{d}_{Liu}$)	0.2434	0.3238	0.0010	0.0238
		Liu $(\hat{d}_{Liu})_I$	0.8742	0.5923	0.0016	0.0285
	200	MLE	2.3486	1.2100	0.0027	0.0372
		Stein	1.1001	0.7918	0.0194	0.1144
		Ridge	0.7421	0.5752	0.0013	0.0257
		Liu ($\hat{d}_{Liu}$)	0.2149	0.3255	0.0008	0.0208
		Liu $(\hat{d}_{Liu})_I$	0.6713	0.5279	0.0012	0.0246

moderate samples sizes (n = 50, 70 and 100), the best options are Liu estimators with $\hat{d}_{Liu}$ followed by ridge regression estimators and Liu estimators with $(\hat{d}_{Liu})_I$ respectively according to $MSE(\beta)$ and $MAE(\beta)$, while according to $MSE(\pi(x))$ and $MAE(\pi(x))$, Liu estimators with both ($\hat{d}_{Liu}$ and $(\hat{d}_{Liu})_I$) are the best option.

In the case of high multicollinearity (ρ = 0.95) with small and moderate samples sizes (n = 50, 70 and 100), Liu estimators with both ($\hat{d}_{Liu}$and $(\hat{d}_{Liu})_I$) showed their best performance by means of the reduction of $MSE(\beta)$, $MAE(\beta)$, $MSE(\pi(x))$ and $MAE(\pi(x))$, while for large samples sizes (n = 150 and 200) the best options are ridge regression estimators followed by Liu estimators with both ($\hat{d}_{Liu}$ and $(\hat{d}_{Liu})_I$) respectively for all correlation coefficient according to $MSE(\beta)$ and $MAE(\beta)$, while due to $MSE(\pi(x))$ and $MAE(\pi(x))$ the best option is Liu estimators with both ($\hat{d}_{Liu}$and $(\hat{d}_{Liu})_I$).

As observed from Table 2, at ρ = 0.75 and small samples sizes (n = 50 and 70) the best options are Liu estimators with $\hat{d}_{Liu}$ followed by ridge regression estimators and Liu estimators with $(\hat{d}_{Liu})_I$, respectively according to the $MSE(\beta)$, While according to $MAE(\beta)$, $MSE(\pi(x))$ and $MAE(\pi(x))$, Liu estimators with both ($\hat{d}_{Liu}$and $(\hat{d}_{Liu})_I$) are the best option. While for moderate and large samples sizes (n = 100, 150 and 200) the best option is Liu estimators with $(\hat{d}_{Liu})_I$ according to the four criteria. In the case of high multicollinearity (ρ = 0.85 and 0.95) with all samples sizes Liu estimators with $\hat{d}_{Liu}$ showed its best performance by means of the reduction of the four criteria.

Thus it can be seen that, Liu estimators with both ($\hat{d}_{Liu}$ and $(\hat{d}_{Liu})_I$) are mostly preferred for correcting multicollinearity problem in binomial logistic regression.

Real data application: In this section, a real data set taken from the annual statistical book for environment in Egypt (September 2014) is used for comparing the different methods for correcting multicollinearity in logistic regression. A logistic regression model is estimated, where the response variable is defined as follows:

$$y_i = \begin{cases} 1 \text{ the i}^{\text{th}} \text{ governorate has a running waste} \\ \quad \text{recycling factory i} = 1,2,\ldots,25 \\ \quad 0 \quad \text{otherwise} \end{cases}$$

The data was collected from the entire Egyptian governorate (25 governorates) during the year (2011).

This response variable will be explained by the following explanatory variables: the quantity of home wastes by tons/year (X_1), the quantity of packing paper wastes by tons/year(X_2), the quantity of plastic wastes by tons/year (X_3), the quantity of glass wastes by tons/year (X_4) and the quantity of metal wastes by tons/year (X_5), respectively. Hence, in this real data application, the effect of changing the type and the

Table 3: The estimated parameters and the standard errors of the different estimators

Variables	MLE estimators	Stein estimators	Ridge regression estimators	Liu estimators $\hat{d}_{Liu}$	Liu estimators $(\hat{d}_{Liu})_I$
x_1	0.698 (1.954)	0.314 (0.879)	0.232 (1.189)	0.111(0.372)	0.316 (0.899)
x_2	-4.371 (3.956)	-1.965 (1.779)	-0.931(1.162)	-0.075 (0.268)	-1.573 (1.504)
x_3	4.783 (4.358)	2.150 (1.959)	1.135 (1.127)	0.227 (0.250)	1.816 (1.634)
x_4	-0.427 (3.290)	-0.192 (1.479)	0.255 (1.264)	0.071 (0.269)	-0.103 (1.295)
x_5	-0.549 (1.815)	-0.247 (0.816)	-0.536 (1.050)	-0.170 (0.409)	-0.302 (0.853)
Estimated bias parameter	________	$\hat{d}_{St}$= 0.450	$\hat{d}_R$= 0.116	$\hat{d}_{Liu}$= 0.001	$(\hat{d}_{Liu})_I$= 0.349

quantity of wastes on the number of running waste recycling factories is explored. The bivariate correlations between the explanatory variables are as follows:

$$\begin{bmatrix} 1 & 0.965 & 0.965 & 0.950 & 0.880 \\ 0.965 & 1 & 0.990 & 0.969 & 0.924 \\ 0.965 & 0.990 & 1 & 0.979 & 0.924 \\ 0.950 & 0.969 & 0.979 & 1 & 0.959 \\ 0.880 & 0.924 & 0.924 & 0.959 & 1 \end{bmatrix}$$

The above correlation matrix; showed that all the bivariate correlations are greater than 0.88 which means that there is a problem of multicollinearity. The logistic regression model is estimated using the computer software R by applying the IWLS algorithm.

For correcting multicollinearity, logistic Stien estimators, logistic ridge regression estimators and logistic Liu estimators (with both $\hat{d}_{Liu}$ in Eq. (14) and $(\hat{d}_{Liu})_I$ in Eq. (18)) were applied. The results are shown in Table 3.

It can be noticed that the quantity of household wastes (X_1) and the quantity of plastic wastes (X_3) have a positive impact on the running waste recycling factory, where, the quantity of packing paper wastes (X_2) and the quantity of metal wastes (X_5) have negative impact. For the quantity of glass wastes (X_4) it has negative impact on the running waste recycling factory for all estimators except the Liu estimators (with $\hat{d}_{Liu}$), these means that one can increase the probability of having a running waste recycling factory by increasing the quantities of household wastes and plastic wastes and decreasing the quantities of packing paper wastes, metal wastes and glass wastes.

Table 3 indicates, that the lowest parameter estimates and its standard errors are obtained by logistic Liu estimators with both ($\hat{d}_{Liu}$ and $(\hat{d}_{Liu})_I$), while the largest are obtained by the MLE estimates which suffer from multicollinearity. It means that logistic Liu estimators with both ($\hat{d}_{Liu}$ and $(\hat{d}_{Liu})_I$) are mostly preferred than other estimators to correct mutilcollinearity in logistic regression which ensure the simulation results.

CONCLUSION

In this study, a new shrinkage parameter for logistic Liu estimator, named $(\hat{d}_{Liu})_I$, which provides an alternative method for dealing with multicollinearity in logistic regression, was introduced. We have designed an algorithm for a Monte Carlo experiment by generating random numbers for explanatory variables and the response variable. We have considered several sample sizes, degrees of correlation and number of the explanatory variables. We have compared the logistic Liu estimators with both ($\hat{d}_{Liu}$and$(\hat{d}_{Liu})_I$) MLE and other estimators (Stien and ridge regression) that were used to correct multicollinearity in binomial logistic regression. The MSE(β), MAE(β), $MSE(\pi(x))$ *and* $MAE(\pi(x))$are used as performance criterion. The results showed that logistic Liu estimators with both ($\hat{d}_{Liu}$and$(\hat{d}_{Liu})_I$) are much more robust to the correlation than other estimators to correct mutilcollinearity in logistic regression. Therefore, the MLE should not be used in the presence of severe multicollinearity, as it becomes unstable with large variances and it has large MSE(β).Logistic Liu estimators with both ($\hat{d}_{Liu}$and$(\hat{d}_{Liu})_I$) has the best performance in the simulation than other estimators. Thus, these results agreed with Kibria *et al*. (2012) and Farghali (2014).

Finally, the estimators are applied to a real dataset, where the effect of changing the type and the quantity of wastes on the number of running waste recycling factories is explored, to show that the logistic Liu estimators with both ($\hat{d}_{Liu}$and$(\hat{d}_{Liu})_I$) are practical.

REFERENCES

Agresti, A., 2002. Categorical Data Analysis. 2nd Edn., Wiley and Sons Inc., New York, USA.

Aguilera, A.M., M. Escabias and M.J. Valderrama, 2006. Using principal components for estimating logistic regression with high-dimensional multicollinear data. Comput. Stat. Data An., 50(8): 1905-1924.

Akdeniz, F. and H. Erol, 2003. Mean squared error matrix comparisons of some biased estimators in linear regression. Commun. Stat.-Theor. M., 32(12): 2389-2413.

Asar, Y. and A. Genc, 2016. New shrinkage parameters for the Liu-type logistic estimators. Commun. Stat. Simulat., 45(3): 1094-1103.

Belsley, D., 1991. A guide to using the collinearity diagnostics. Comput. Sci. Econ. Manage., 4(1): 33-50.

Belsley, D.A., E. Kuh and R.E. Welsch, 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley and Sons Inc., New York, USA.

Camminatiello, I. and A. Lucadamo, 2010. Estimating multinomial logit model with multicollinear data. Asian J. Math. Stat., 3(2): 39-101.

Dielman, T., 2005. Applied Regression Analysis: A Second Course in Business and Economic Statistics. 4th Edn., Thomson Brooks-Cole, Australia.

El-Dash, A., A. El-Hefnawy and A. Farghali, 2011. Goal programming technique for correcting multicollinearity problem in multinomial logistic regression. Proceedings of the 46th Annual Conference on Statistics, Computer Sciences and Operation Research, December 26-29, pp: 72-87.

Farghali, R., 2012. Using multi-objective programming to correct the collinearity for logistic regression analysis with application. Ph.D. Thesis, Mathematics and Applied Statistics Department, Faculty of Commerce and Business Administration, Helwan University, Egypt. (In Arabic)

Farghali, R., 2014. A suggested biased estimator for correcting multicollinearity in. multinomial logistic regression. Egypt. Stat. J., 58: 183-197.

Farag, A.R.A., R. Hamed and A. El-Hefnawy, 2012. Selection of the ridge parameter using mathematical programming. M.Sc. Thesis, Statistics Department, Faculty of Economics and Political Science, Cairo University, Egypt.

Hamed, R., A. El-Hefnawy and A. Farag, 2013. Selection of the ridge parameter using mathematical programming. Commun. Stat. Simul. C., 42(6): 1409-1423.

Hoerl, A.E. and R.W. Kennard, 1970a. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1): 55-67.

Hoerl, A.E. and R.W. Kennard, 1970b. Ridge regression: Applications to nonorthogonal problems. Technometrics, 12(1): 69-82.

Hosmer, D. and S. Lemeshow, 2000. Applied Logistic Regression. 2nd Edn., Wiley and Sons Inc., New York, USA.

James, W. and C. Stien, 1961. Estimation with quadratic loss. Proceedings of the 4th Berkely Symposium on Mathematical Statistics and Probability, 1: 361-379.

Kan, B., Ö. Alpu and B. Yazici, 2013. Robust ridge and robust Liu estimator for regression based on the LTS estimator. J. Appl. Stat., 40(3): 644-655.

Kibria, B.M.G., K. Månsson and G. Shukur, 2012. Performance of some logistic ridge regression estimators. Comput. Econ., 40(4): 401-414.

Liu k., 1993. A new class of biased estimate in linear regression" Commun. Stat. Theory, 22(2); 393-402.

Lesaffre, E. and B.D. Marx, 1993. Collinearity in generalized linear regression. Commun. Stat. Theory., 22(7): 1933-1952.

Le Cessie, S. and J.C. Van Houwelingen, 1992. Ridge estimators in logistic regression. J. Roy. Stat. Soc. C App., 41(1): 191-201.

Liu, K., 2003. Using Liu-type estimator to combat collinearity. Commun. Stat. Theory, 32(5): 1009-1020.

Liu, K., 2004. More on Liu-type estimator in linear regression. Commun. Stat. Theory, 33(11): 2723-2733.

Månsson, K. and G. Shukur, 2011. On ridge parameters in logistic regression. Commun. Stat. Theory, 40(18): 3366-3381.

Månsson, K., B.M. Golam Kibria and G. Shukur, 2012. On Liu estimators for the logit regression model. Econ. Model., 29(4): 1483-1488.

Månsson, K., B.M. Golam Kibria and G. Shukur, 2015. Performance of some weighted Liu estimators for logit regression model: An application to Swedish accident data. Commun. Stat. Theory, 44(2): 363-375.

Rong, J.Y., 2010. Adjustive Liu-type estimators in linear regression models. Commun. Stat. Simulat., 39(6): 1162-1173.

Schaefer, R.L., 1986. Alternative estimators in logistic regression when the data are collinear. J. Stat. Comput. Sim., 25(1-2): 75-91.

Schaefer, R.L., L.D. Roi and R.A. Wolfe, 1984. A ridge logistic estimator. Commun. Stat. Theory, 13(1): 99-113.

Steyerberg, E.W., M.J.C. and J.D.F. Habbema, 2001. Application of shrinkage techniques in logistic regression analysis: A case study. Stat. Neerl., 55(1): 76-88.

Tutz, G. and F. Leitenstorfer, 2006. Response shrinkage estimators in binary regression. Comput. Stat. Data An., 50(10): 2878-2901.