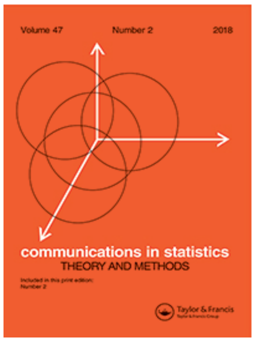

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Optimal generalized logistic estimator

Nagarajah Varathan & Pushpakanthie Wijekoon

To cite this article: Nagarajah Varathan & Pushpakanthie Wijekoon (2018) Optimal generalized logistic estimator, Communications in Statistics - Theory and Methods, 47:2, 463-474, DOI: 10.1080/03610926.2017.1307406

To link to this article: https://doi.org/10.1080/03610926.2017.1307406

Published online: 14 Sep 2017.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2018, VOL. 47, NO. 2, 463–474
https://doi.org/10.1080/03610926.2017.1307406

Optimal generalized logistic estimator

Nagarajah Varathan[a,b] and Pushpakanthie Wijekoon[c]

[a]Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka; [b]Department of Mathematics and Statistics, University of Jaffna, Jaffna, Sri Lanka; [c]Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka

ABSTRACT
In this paper, we propose a new efficient estimator namely Optimal Generalized Logistic Estimator (OGLE) for estimating the parameter in a logistic regression model when there exists multicollinearity among explanatory variables. Asymptotic properties of the proposed estimator are also derived. The performance of the proposed estimator over the other existing estimators in respect of Scalar Mean Square Error criterion is examined by conducting a Monte Carlo simulation.

ARTICLE HISTORY
Received 19 October 2016
Accepted 6 March 2017

KEYWORDS
Logistic regression; Multicollinearity; Biased estimator; Optimal generalized logistic estimator; Scalar mean square error.

1. Introduction

Logistic regression model is a popular method to model binary data in many application areas in statistics. However, unstable parameter estimators based on the maximum likelihood method occur when the covariates are highly correlated. This phenomenon is known as the multicollinearity among the predictor variables, and the remedial measures for parameter estimates were discussed in the literature. Some of the proposed estimators to overcome the multicolinearity are, namely, the Ridge Logistic Estimator (RLE) (Schaefer, Roi, and Wolfe 1984), Liu Logistic Estimator (LLE) (Liu 1993; Urgan and Tez 2008; and Mansson, Kibria, and Shukur 2012), Almost Unbiased Ridge Logistic Estimator (AURLE) (Wu and Asar 2016), Almost Unbiased Liu Logistic Estimator (AULLE) (Xinfeng 2015), and Liu type logistic estimator (Inan and Erdogan 2013). Further, Asar (2015) and Asar and Genç (2016) introduced some new shrinkage parameters for the Liu-type logistic estimator to improve its efficiency. However, when comparing the efficiency of these estimators, it was noted that none of the above estimators are always superior over the others. In this paper, we introduce a new estimator called the Optimal Generalized Logistic Estimator (OGLE) based on Quasi-Likelihood (QL) estimation technique (Wedderburn 1974), which shows more efficiency than all the estimators proposed in the literature.

The rest of the paper is organized as follows. The model specification and existing estimators are given in Section 2. The Optimal Generalized Logistic Estimator (OGLE) has been proposed and its asymptotic properties are derived in Section 3. In Section 4, the performance of the proposed estimator with respect to Scalar Mean Squared Error (SMSE) is compared with some existing estimators by performing a Monte Carlo simulation study. A real data example is analyzed in Section 5. Finally, the conclusion of the study is presented in Section 6.

CONTACT Nagarajah Varathan varathan10gmail.com Department of Mathematics and Statistics, University of Jaffna, Jaffna 40000, Sri Lanka.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.

2. Model specification and existing estimators

Consider the logistic regression model

$$y_i = \pi_i + \varepsilon_i, \qquad i = 1, \ldots, n \tag{1}$$

which follows Bernoulli distribution with parameter π_i as

$$\pi_i = \frac{\exp\left(x_i'\beta\right)}{1 + \exp\left(x_i'\beta\right)} \tag{2}$$

where x_i is the i^{th} row of X, which is an $n \times (p+1)$ data matrix with p predictor variables and β is a $(p+1) \times 1$ vector of coefficients, ε_i are independent with mean zero and variance $\pi_i(1-\pi_i)$ of the response y_i. The maximum likelihood estimator (MLE) of β can be obtained as follows:

$$\hat{\beta}_{MLE} = C^{-1}X'\hat{W}Z, \tag{3}$$

where $C = X'\hat{W}X$; Z is the column vector with i^{th} element equals $logit\,(\hat{\pi}_i) + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)}$ and $\hat{W} = diag[\hat{\pi}_i(1-\hat{\pi}_i)]$, which is an unbiased estimate of β. The covariance matrix of $\hat{\beta}_{MLE}$ is

$$Cov\,(\hat{\beta}_{MLE}) = \{X'\hat{W}X\}^{-1}. \tag{4}$$

To combat the multicollinearity in logistic regression, several estimators were proposed, based only on the sample information in the literature. Some of these estimators are Ridge Logistic Estimator (RLE) (Schaefer, Roi, and Wolfe 1984), Liu Logistic Estimator (LLE) (Liu 1993; Urgan and Tez 2008; and Mansson et al. 2011), Almost Unbiased Ridge Logistic Estimator (AURLE) (Wu and Asar 2016), and Almost Unbiased Liu Logistic Estimator (AULLE) (Xinfeng 2015). These estimators are defined as

$$RLE: \ \hat{\beta}_{RLE} = Z_k\hat{\beta}_{MLE}; \quad \text{where } Z_k = (I + kC^{-1})^{-1}, \ K \geq 0 \tag{5}$$

$$LLE: \ \hat{\beta}_{LLE} = Z_d\hat{\beta}_{MLE}; \quad \text{where } Z_d = (C+I)^{-1}(C+dI), \ 0 \leq d \leq 1 \tag{6}$$

$$AURLE: \ \hat{\beta}_{AURLE} = W_k\hat{\beta}_{MLE}; \quad \text{where } W_k = \left[I - k^2(C+kI)^{-2}\right], \ k \geq 0 \tag{7}$$

$$AULLE: \ \hat{\beta}_{AULLE} = W_d\hat{\beta}_{MLE}; \quad \text{where } W_d = \left[I - (1-d)^2\,(C+I)^{-2}\right], \ 0 \leq d \leq 1 \tag{8}$$

Note that both Z_k and Z_d are clearly positive definite. Further, both W_k and W_d are matrices since $W_k = (C+kI)^{-2}C(C+2kI) > 0$, and $W_d = (C+I)^{-2}[C^2 + 2C + dI(2-d)] > 0$ (since $(2-d) > 0; 0 \leq d \leq 1$). It can be noticed from the estimators defined in the Eqs. (5)–(8) that RLE, LLE, AURLE, and AULLE are functions of $\hat{\beta}_{MLE}$. Consequently, estimators having this format can be represented as a general form, which is called as Generalized Logistic Estimator (GLE), and is defined as

$$\hat{\beta}_{GLE} = J_{(i)}\hat{\beta}_{MLE} \tag{9}$$

where $J_{(i)}$ is a positive definite matrix, and in this paper $J_{(i)}$ stands for Z_k, Z_d, W_k, and W_d.

The asymptotic properties of GLE are

$$\begin{aligned} E[\hat{\beta}_{GLE}] &= E[J_{(i)}\hat{\beta}_{MLE}] \\ &= J_{(i)}\beta, \end{aligned} \tag{10}$$

and the dispersion matrix;

$$D[\hat{\beta}_{GLE}] = Cov\,[J_{(i)}\hat{\beta}_{MLE}]$$

$$= J_{(i)}C^{-1}J'_{(i)}. \tag{11}$$

Then, the Bias vector and Mean square error matrix (MSE) are

$$\begin{aligned} B[\hat{\beta}_{GLE}] &= E[J_{(i)}\hat{\beta}_{MLE}] - \beta \\ &= (J_{(i)} - I)\beta, \end{aligned} \tag{12}$$

and

$$\begin{aligned} MSE[\hat{\beta}_{GLE}] &= D[\hat{\beta}_{GLE}] + B[\hat{\beta}_{GLE}]B'[\hat{\beta}_{GLE}] \\ &= J_{(i)}C^{-1}J'_{(i)} + (J_{(i)} - I)\beta\beta'(J_{(i)} - I)' \\ &= J_{(i)}C^{-1}J'_{(i)} + J_{(i)}\left(I - J_{(i)}^{-1}\right)\beta\beta'\left(I - J_{(i)}^{-1}\right)'J'_{(i)}. \end{aligned} \tag{13}$$

Consequently, the Scalar mean square error (SMSE) can be obtained as

$$\begin{aligned} SMSE[\hat{\beta}_{GLE}] &= tr[MSE(\hat{\beta}_{GLE})] \\ &= tr(J_{(i)}C^{-1}J'_{(i)}) + \beta'\left(I - J_{(i)}^{-1}\right)'J'_{(i)}J_{(i)}\left(I - J_{(i)}^{-1}\right)\beta. \end{aligned} \tag{14}$$

3. The proposed new estimator

Although in the Generalized Logistic Estimator (GLE) in (9), the component $J_{(i)}$ can take different choices corresponding to different type of estimators, finding an optimal choice of $J_{(i)}$ is more meaningful. To achieve this, first we minimize the Scalar mean square error of GLE with respect to $J_{(i)}$. Then the unknown parameters are estimated by using the Quasi-Likelihood (QL) estimation technique. The resulting estimator is called as Optimal Generalized Logistic Estimator (OGLE).

To minimize the SMSE of GLE, first we consider the derivative of Eq. (14) with respect to $J_{(i)}$ as

$$\frac{\partial\{SMSE[\hat{\beta}_{GLE}]\}}{\partial J_{(i)}} = \frac{\partial\{tr(J_{(i)}C^{-1}J'_{(i)})\}}{\partial J_{(i)}} + \frac{\partial\beta' L\beta}{\partial J_{(i)}} \tag{15}$$

where $L = (I - J_{(i)}^{-1})'J'_{(i)}J_{(i)}(I - J_{(i)}^{-1})$ and it can be simplified as

$$\begin{aligned} L &= \left(I - J_{(i)}^{-1}\right)'J'_{(i)}J_{(i)}\left(I - J_{(i)}^{-1}\right) \\ &= \left(I - J_{(i)}^{-1}\right)'\left[J'_{(i)}J_{(i)} - J'_{(i)}\right] \\ &= J'_{(i)}J_{(i)} - J_{(i)} - J'_{(i)} + I \end{aligned} \tag{16}$$

Applying (16) in (15), implies

$$\begin{aligned} \frac{\partial\{SMSE[\hat{\beta}_{GLE}]\}}{\partial J_{(i)}} &= \frac{\partial\left\{tr(J_{(i)}C^{-1}J'_{(i)})\right\}}{\partial J_{(i)}} + \frac{\partial\left\{\beta'J'_{(i)}J_{(i)}\beta - 2\beta'J_{(i)}\beta + \beta'\beta\right\}}{\partial J_{(i)}} \\ &= \frac{\partial\left\{tr(J_{(i)}C^{-1}J'_{(i)})\right\}}{\partial J_{(i)}} + \frac{\partial\left\{\beta'J'_{(i)}J_{(i)}\beta\right\}}{\partial J_{(i)}} - 2\frac{\partial\left\{\beta'J_{(i)}\beta\right\}}{\partial J_{(i)}} \end{aligned} \tag{17}$$

Now, we use the following results (see Rao and Toutenburg 1995, p. 385, 386) in Eq. (17),
R1. Let N and Y be any two matrices with proper order, then

$$\frac{\partial tr(YNY')}{\partial Y} = Y(N + N')$$

R2. If x is a vector of order $n \times 1$, y is another vector of order $m \times 1$, and C is an $n \times m$ matrix, then

$$\frac{\partial x'Cy}{\partial C} = xy'$$

R3. Let x be a $n \times 1$ vector, N a symmetric $t \times t$ matrix, and C a $t \times n$ matrix, then

$$\frac{\partial x'C'NCx}{\partial C} = 2NCxx'$$

By applying **R1, R2**, and **R3** in (17), we obtain

$$\frac{\partial \left\{tr(J_{(i)}C^{-1}J'_{(i)})\right\}}{\partial J_{(i)}} = 2J_{(i)}C^{-1}, \tag{18}$$

$$\frac{\partial \left\{\beta' J'_{(i)}J_{(i)}\beta\right\}}{\partial J_{(i)}} = 2J_{(i)}\beta\beta', \tag{19}$$

and

$$\frac{\partial\{\beta' J_{(i)}\beta\}}{\partial J_{(i)}} = \beta\beta' \tag{20}$$

respectively.

Substituting (18), (19), and (20) in (17), we get

$$\begin{aligned}\frac{\partial\{SMSE[\hat{\beta}_{GLE}]\}}{\partial J_{(i)}} &= 2J_{(i)}C^{-1} + 2J_{(i)}\beta\beta' - 2\,\beta\beta' \\ &= 2J_{(i)}(C^{-1} + \beta\beta') - 2\,\beta\beta' \end{aligned} \tag{21}$$

The matrix $C^{-1} + \beta\beta'$ is positive definite (see Rao and Toutenburg 1995, p. 366), and hence, non singular. Equating (21) to a null-matrix, we shall obtain an optimal choice for $J_{(i)}$ as

$$\tilde{J}_{(i)} = \beta\beta'(C^{-1} + \beta\beta')^{-1} \tag{22}$$

Now, we propose an Optimal Generalized Logistic Estimator (OGLE) as

$$\hat{\beta}_{OGLE} = \tilde{J}_{(i)}\hat{\beta}_{MLE} \tag{23}$$

Since $\tilde{J}_{(i)}$ in (23) contains an unknown parameter β, we use Quasi-Likelihood (QL) technique to estimate β. Application of the QL estimation technique for the logistic regression model (1) is discussed below.

Quasi-Likelihood (QL) estimation of β

Suppose that a scalar response y_i and a p dimensional vector of covariates x_i are observed for individuals $i = 1, 2, \ldots, n$. Further, suppose that the marginal density of the response y_i; $i = 1, 2, \ldots, n$ is of the exponential family form

$$f(y_i) = \exp\left[\left\{y_i\theta_i - a(\theta_i) + b(y_i)\right\}\right] \tag{24}$$

(Liang and Zeger 1986), where $\theta_i = h(\eta_i)$ with $\eta_i = x_i'\beta$, $a(.)$, $b(.)$ and $h(.)$ are of known functional form, and β is the $p \times 1$ vector of parameters of interest. Consequently, the mean and variance function of the response y_i as

$$E[y_i] = a'(\theta_i) \quad and \quad Var[y_i] = a''(\theta_i), \tag{25}$$

where $a'(\theta_i)$ and $a''(\theta_i)$ are first and second order derivatives of $a(\theta_i)$, respectively, with respect to θ_i. To estimate the parameter β under this independent setup, Wedderburn (1974) proposed the Quasi-Likelihood estimating equation given by

$$\sum_{i=1}^{n}\left[\frac{\partial a'(\theta_i)}{\partial \beta}\frac{\left(y_i - a'(\theta_i)\right)}{a''(\theta_i)}\right] = 0. \tag{26}$$

In the case of logistic regression,
$y_i \sim$ Bernoulli(π_i), where

$$\begin{aligned} \pi_i &= \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \\ f(y_i) &= \pi_i^{y_i}(1-\pi_i)^{(1-y_i)} \\ &= \left\{\frac{\pi_i}{1-\pi_i}\right\}^{y_i}(1-\pi_i) \\ &= \exp\left\{y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right) + \ln(1-\pi_i)\right\} \\ &= \exp\{y_i\theta_i - a(\theta_i) + b(y_i)\}, \end{aligned} \tag{27}$$

which is in the form of exponential family, where

$$\theta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = x_i'\beta, \qquad a(\theta_i) = -\ln(1-\pi_i)$$

and

$$b(y_i) = 0.$$

This implies

$$a'(\theta_i) = \frac{\exp(\theta_i)}{1+\exp(\theta_i)} = \pi_i, \tag{28}$$

and

$$a''(\theta_i) = \pi_i(1-\pi_i). \tag{29}$$

Consequently, the QL estimating Eq. (26) becomes

$$\sum_{i=1}^{n}\left[\frac{\partial \pi_i}{\partial \beta}\frac{(y_i-\pi_i)}{\pi_i(1-\pi_i)}\right] = 0. \tag{30}$$

$$\text{where } \frac{\partial \pi_i}{\partial \beta} = \pi_i(1-\pi_i)x_i.$$

By applying the New-Raphson iterative algorithm, one can obtain the QL estimator which is the convergent value of the following iterative equation

$$\begin{aligned} \hat{\beta}_{QL}(r+1) = \hat{\beta}_{QL}(r) = + &\left[\sum_{i=1}^{n}\left[\frac{\partial \pi_i}{\partial \beta}\{\pi_i(1-\pi_i)\}^{-1}\frac{\partial \pi_i}{\partial \beta'}\right]\right]_r^{-1} \\ &\times \left[\sum_{i=1}^{n}\left[\frac{\partial \pi_i}{\partial \beta}\frac{(y_i-\pi_i)}{\pi_i(1-\pi_i)}\right]\right]_r = 0 \end{aligned} \tag{31}$$

Where $[]_r$ denotes the expression within the square bracket is evaluated at $\beta = \hat{\beta}_{QL}(r)$, the estimate obtained for the r^{th} iteration.

Note that QL estimator is also an alternative estimator for $\hat{\beta}_{MLE}$ in (3). Now in the Optimal Generalized Logistic Estimator (OGLE) in (23), $\tilde{J}_{(i)}$ can be estimated using the QL estimate, $\hat{\beta}_{QL}$, obtained from the Eq. (31) as

$$\begin{aligned}\widehat{\tilde{J}}_{(i)} &= \tilde{J}_{(i)}|\hat{\beta}_{QL}\\ &= \hat{\beta}_{QL}\hat{\beta}'_{QL}(C^{-1} + \hat{\beta}_{QL}\hat{\beta}'_{QL})^{-1}\end{aligned} \tag{32}$$

The asymptotic properties of OGLE:

$$\begin{aligned}B(\hat{\beta}_{OGLE}) &= E[\hat{\beta}_{OGLE}] - \beta\\ &= (\widehat{\tilde{J}}_{(i)} - I)\beta;\end{aligned} \tag{33}$$

$$D(\hat{\beta}_{OGLE}) = \widehat{\tilde{J}}_{(i)}C^{-1}\widehat{\tilde{J}}'_{(i)}; \tag{34}$$

$$\begin{aligned}MSE(\hat{\beta}_{OGLE}) &= \widehat{\tilde{J}}_{(i)}C^{-1}\widehat{\tilde{J}}'_{(i)} + (\widehat{\tilde{J}}_{(i)} - I)\,\beta\beta'(\widehat{\tilde{J}}_{(i)} - I)'\\ &= \widehat{\tilde{J}}_{(i)}C^{-1}\widehat{\tilde{J}}'_{(i)} + \widehat{\tilde{J}}_{(i)}\left(I - \widehat{\tilde{J}}^{-1}_{(i)}\right)\beta\beta'\left(I - \widehat{\tilde{J}}^{-1}_{(i)}\right)'\widehat{\tilde{J}}'_{(i)};\end{aligned} \tag{35}$$

and

$$SMSE(\hat{\beta}_{OGLE}) = tr(\widehat{\tilde{J}}_{(i)}C^{-1}\widehat{\tilde{J}}'_{(i)}) + \beta'\left(I - \widehat{\tilde{J}}^{-1}_{(i)}\right)'\widehat{\tilde{J}}'_{(i)}\widehat{\tilde{J}}_{(i)}\left(I - \widehat{\tilde{J}}^{-1}_{(i)}\right)\beta. \tag{36}$$

In the next section, by conducting a simulation study, we investigate the relative performance of the proposed optimal estimator over the other existing estimators with respect to the scalar mean square error sense.

4. The performance of the new estimator

A Monte Carlo simulation study is conducted to study the performance of the proposed optimal estimator based on the other existing estimators under different levels of multicollonearity. Following McDonald and Galarneau (1975), Gibbons (1981), Kibria (2003), and Muniz and Kibria (2009), the predictor variables are generated using the following equation:

$$x_{ij} = (1-\rho^2)^{1/2}z_{ij} + \rho z_{i,p+1}, \qquad i = 1,\ 2, \ldots,\ n,\ j = 1,\ 2, \ldots,\ p \tag{37}$$

where z_{ij} pseudorandom numbers from standardized normal distribution and ρ^2 represents the correlation between any two explanatory variables. Four explanatory variables are generated using (37) and four different values of ρ corresponding to 0.80, 0.90, 0.95, and 0.99 are considered. Further, to understand the effect of the sample size n, three different values 20, 50, and 100 are taken. The dependent variable y_i in (1) is obtained from the binary (π_i) distribution, where $\pi_i = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}$. The parameter values of $\beta_1,\ \beta_2, \ldots,\ \beta_p$ for each vector of estimator $\hat{\beta}_{RLE},\ \hat{\beta}_{LLE},\ \hat{\beta}_{AURLE},\ \widehat{\beta}_{AULLE},\ \hat{\beta}_{QL}$, and $\hat{\beta}_{OGLE}$ considered in this study are chosen so that $\sum_{j=1}^{p}\beta_j^2 = 1$ and $\beta_1 = \beta_2 = \cdots = \beta_p$ (Şiray, Toker, and Kaçiranlar 2015). Further, for the ridge parameter k and the Liu parameter d, some selected values are chosen so that $0 < k < 1$ and $0 < d < 1$. The simulation is repeated 2000 times by generating new pseudorandom numbers and the simulated SMSE values of MLE, QL, LRE, LLE, AURLE, AULLE,

Table 1. The correlation matrix of the design matrix.

	x1	x2	x3	x4
x1	1.000	0.998	0.971	0.970
x2	0.998	1.000	0.960	0.958
x3	0.971	0.960	1.000	0.987
x4	0.970	0.958	0.987	1.000

and OGLE are obtained using the following equation:

$$S\hat{M}SE(\hat{\beta})\frac{1}{2000}\sum_{r=1}^{2000}(\hat{\beta}_r-\beta)'(\hat{\beta}_r-\beta) \tag{38}$$

where $\hat{\beta}_r$ is any estimator considered in the r^{th} simulation. The results of the simulation are reported in Table B1–B12 (Appendix B) and also displayed in Figure A1–A3 (Appendix A). It can be observed from Figure A1–A3, increase in degree of correlation between two explanatory variables ρ inflates the estimated SMSE of all the estimators, and in general, increase in the sample size n decreases the estimated SMSE of all the estimators. According to Table B1–B12, the proposed estimator OGLE has smaller scalar mean square values compared to all the other estimators—MLE, QL, LRE, LLE, AURLE, and AULLE with respect to all $\rho = 0.8$, 0.9, 0.95, and 0.99, and $n = 20$, 50, and 100 considered in this study. Moreover, the performance of the QL estimator is better compared to the MLE for all sample sizes n and ρ values, in the mean square error sense. It was further noted from the simulation results, when the multicollinearity is very high, the LRE performs better compared to MLE, QL, LLE, AURLE, and AULLE for large k, d values.

5. A real data example

To illustrate the performance of the new estimator OGLE, we consider a real data application, which is obtained from the Statistics Sweden website (http://www.scb.se/). This example was used by Mansson et al. (2012), Asar and Genç (2016), Wu and Asar (2016), and Varathan and Wijekoon (2016) to illustrate the results of their papers. The data describe the information of 100 municipalities of Sweden. The variables considered in this study are Population ($x1$), Number of unemployed people ($x2$), Number of newly constructed buildings ($x3$), Number of bankrupt firms ($x4$), and Net population change (y). The response variable y is defined as

$$y = \begin{cases} 1 & \text{if there is an increase in the population;} \\ 0 & o/w \end{cases}$$

The correlation matrix of the design matrix $x = [x1,\ x2,\ x3,\ x4]$ is given in Table 1. It can be observed from Table 1 that all the correlations among the explanatory variables are very high (greater than 0.95). The corresponding VIF values for the data are 488.17, 344.26, 44.99, and 50.71. VIF measures how much the variance of the estimated regression coefficients is inflated as compared to when the predictor variables are not linearly related. According to the literature, multicollinearity is high if VIF > 10. Hence, a clear high multicollinearity exists in the data set. Further, the condition number, which is used as a measure of the degree of multicollinearity is obtained as 188. This also indicates the sign of severe multicollinearity in this data set. The SMSE values of MLE, QL, LRE, LLE, AURLE, AULLE, and OGLE for some selected values of biasing parameters k, d in the range $0 < k,\ d < 1$ are given in Table B13. Results in Table B13 clearly show that the new estimator OGLE performs well compared to the estimators of MLE, QL, LRE, LLE, AURLE, and AULLE in the SMSE sense, with respect to the selected values of $k,\ d$ in the range $0 < k,\ d < 1$. Moreover, the estimators MLE, AURLE, and AULLE give a nearly equal performance with respect to the SMSE sense, for the given values of $k,\ d$.

6. Concluding remarks

In this paper, we proposed an Optimal Generalized Logistic Estimator (OGLE) for logistic regression model when there exists multicollinearity among explanatory variables. The relative performance of the proposed optimal estimator as compared to some existing estimators was analyzed by conducting a Monte Carlo simulation study. Further, a real data application is given to illustrate the behavior of the proposed estimator. The empirical results of this paper show that, in the scalar mean square error sense, the proposed estimator OGLE is superior over the estimators; MLE, QL, LRE, LLE, AURLE, and AULLE, which are based only on the sample information.

References

Asar, Y. 2015. Some new methods to solve multicollinearity in logistic regression. *Com-Munications in Statistics – Simulation and Computation*. 10.1080/03610918.2015.1053925.

Asar, Y., and A. Genç. 2016. New shrinkage parameters for the Liu-type logistic estimators. *Communications in Statistics- Simulation and Computation* 45 (3):1094–103. 10.1080/03610918.2014.995815.

Gibbons, D. G. 1981. A simulation study of some ridge estimators. *Journal of the American Statistical Association* 76:131–39. 10.1080/01621459.1981.10477619.

Inan, D., and B. E. Erdogan. 2013. Liu-Type logistic estimator. *Communications in Statistics-Simulation and Computation* 42 (7):1578–86. 10.1080/03610918.2012.667480.

Kibria, B. M. G. 2003. Performance of some new ridge regression estimators. *Communications in Statistics: Theory and Methods* 32:419–35.

Liang, K.-Y., and S. L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22. 10.1093/biomet/73.1.13.

Liu, K. 1993. A new class of biased estimate in linear regression. *Communications in Statistics: Theory and Methods* 22:393–402.

Mansson, G., B. M. G. Kibria, and G. Shukur. 2012. On Liu estimators for the logit regression model. *The Royal Institute of Technology, Centre of Excellence for Science and Innovation Studies (CESIS), Sweden*, Paper No. 259.

McDonald, G. C., and D. I. Galarneau. 1975. A Monte Carlo evaluation of some ridge type estimators. *Journal of the American Statistical Association* 70:407–16.

Muniz, G., and B. M. G. Kibria. 2009. On some ridge regression estimators: an Empirical Comparisons. *Communications in Statistics- Simulation and Computation* 38:621–30. 10.1080/03610910802592838.

Rao, C. R., and H. Toutenburg. 1995. *Linear Models: LeastSquares and Alternatives*. Second Edition. New York: Springer-Verlag.

Schaefer, R. L., L. D. Roi, and R. A. Wolfe. 1984. A ridge logistic estimator. *Communications in Statistics – Theory and Methods* 13:99–113. 10.1080/03610928408828664.

Şiray, G. U., S. Toker, and S. Kaçiranlar. 2015. On the restricted Liu estimator in logistic regression model. *Communications in Statistics- Simulation and Computation* 44:217–32. 10.1080/03610918.2013.771742.

Urgan, N. N., and M. Tez. 2008. Liu estimator in logistic regression when the data are collinear. *International Conference*. "*Continuous Optimization and Knowledge-Based Technologies*". pp. 323–27.

Varathan, N., and P. Wijekoon. 2016. Logistic Liu Estimator under stochastic linear restrictions. *Statistical Papers*. 10.1007/s00362-016-0856-6.

Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61:439–47.

Wu, J., and Y. Asar. 2016. On almost unbiased ridge logistic estimator for the logistic regression model. *Hacettepe Journal of Mathematics and Statistics* 45 (3):989–98.
Xinfeng, C. 2015. On the almost unbiased ridge and Liu estimator in the logistic regression model. *International Conference on Social Science, Education, Management and Sports Education*. Atlantis Press: 1663–65.

Appendix A

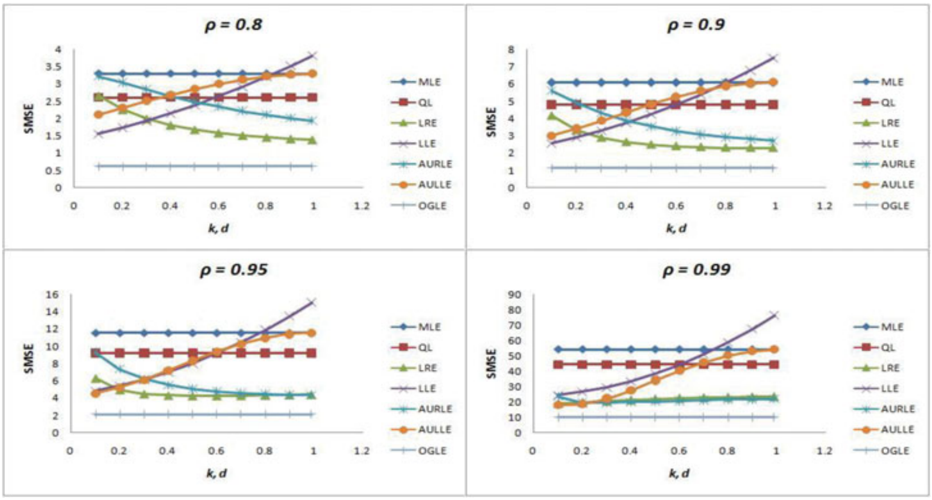

Figure A1. Estimated SMSE values for MLE, QL, LRE, LLE, AURLE, AULLE, and OGLE for $n = 20$.

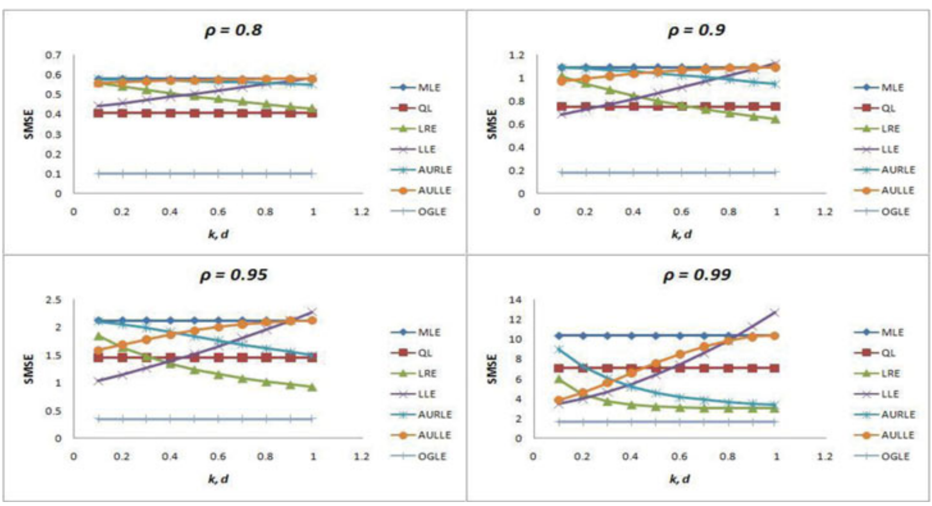

Figure A2. Estimated SMSE values for MLE, QL, LRE, LLE, AURLE, AULLE, and OGLE for $n = 50$.

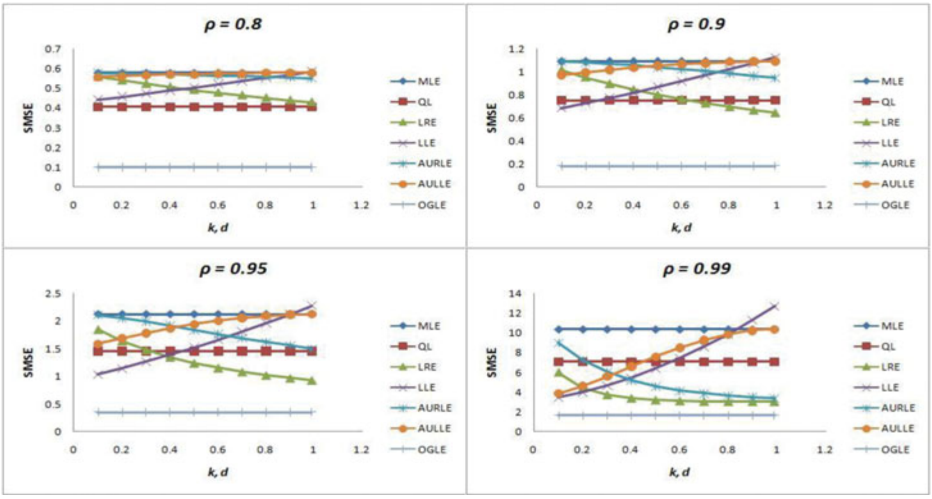

Figure A3. Estimated SMSE values for MLE, QL, LRE, LLE, AURLE, AULLE, and OGLE for $n = 100$.

Appendix B

Table B1. The estimated MSE values for different k, d when $n = 20$ and $\rho = 0.80$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	3.3038	3.3038	3.3038	3.3038	3.3038	3.3038	3.3038	3.3038	3.3038	3.3038
QL	2.6065	2.6065	2.6065	2.6065	2.6065	2.6065	2.6065	2.6065	2.6065	2.6065
LRE	2.6672	2.2674	2.0016	1.8173	1.6857	1.5893	1.5176	1.4636	1.4225	1.3940
LLE	1.5515	1.7306	1.9286	2.1454	2.3811	2.6357	2.9092	3.2015	3.5127	3.8089
AURLE	3.2145	3.0369	2.8427	2.6587	2.4936	2.3486	2.2226	2.1133	2.0186	1.9440
AULLE	2.1162	2.3090	2.5040	2.6919	2.8646	3.0153	3.1382	3.2292	3.2850	3.3037
OGLE	0.6159	0.6159	0.6159	0.6159	0.6159	0.6159	0.6159	0.6159	0.6159	0.6159

Table B2. The estimated MSE values for different k, d when $n = 20$ and $\rho = 0.90$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	6.0929	6.0929	6.0929	6.0929	6.0929	6.0929	6.0929	6.0929	6.0929	6.0929
QL	4.8110	4.8110	4.8110	4.8110	4.8110	4.8110	4.8110	4.8110	4.8110	4.8110
LRE	4.1834	3.3338	2.8946	2.6474	2.5015	2.4136	2.3608	2.3301	2.3140	2.3077
LLE	2.5820	2.9128	3.2997	3.7428	4.2421	4.7975	5.4090	6.0767	6.8006	7.5000
AURLE	5.6109	4.9150	4.3355	3.8907	3.5530	3.2949	3.0957	2.9403	2.8181	2.7300
AULLE	3.0166	3.4181	3.8801	4.3622	4.8293	5.2519	5.6057	5.8719	6.0370	6.0924
OGLE	1.1315	1.1315	1.1315	1.1315	1.1315	1.1315	1.1315	1.1315	1.1315	1.1315

Table B3. The estimated MSE values for different k, d when $n = 20$ and $\rho = 0.95$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	11.5467	11.5467	11.5467	11.5467	11.5467	11.5467	11.5467	11.5467	11.5467	11.5467
QL	9.2315	9.2315	9.2315	9.2315	9.2315	9.2315	9.2315	9.2315	9.2315	9.2315
LRE	6.3043	4.9491	4.4728	4.2952	4.2404	4.2414	4.2688	4.3086	4.3534	4.3951
LLE	4.8257	5.3983	6.1174	6.9831	7.9953	9.1541	10.4594	11.9113	13.5098	15.0738
AURLE	9.3534	7.3891	6.2225	5.5170	5.0751	4.7911	4.6062	4.4859	4.4089	4.3651
AULLE	4.5768	5.2268	6.1589	7.2387	8.3496	9.3930	10.2882	10.9725	11.4009	11.5453
OGLE	2.1487	2.1487	2.1487	2.1487	2.1487	2.1487	2.1487	2.1487	2.1487	2.1487

Table B4. The estimated MSE values for different k, d when $n = 20$ and $\rho = 0.99$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	54.2219	54.2219	54.2219	54.2219	54.2219	54.2219	54.2219	54.2219	54.2219	54.2219
QL	44.4963	44.4963	44.4963	44.4963	44.4963	44.4963	44.4963	44.4963	44.4963	44.4963
LRE	19.1369	19.7380	20.7354	20.7354	22.0880	22.5313	22.8787	23.1572	23.3848	23.5567
LLE	24.5915	26.5799	29.5394	29.5394	38.3718	44.2447	51.0887	58.9038	67.6900	76.4279
AURLE	23.5329	19.6725	19.3765	19.7442	20.2267	20.6916	21.1075	21.4715	21.7885	22.0389
AULLE	18.1782	18.6758	22.2328	27.6798	34.0036	40.3468	46.0081	50.4413	53.2573	54.2122
OGLE	10.1770	10.1770	10.1770	10.1770	10.1770	10.1770	10.1770	10.1770	10.1770	10.1770

Table B5. The estimated MSE values for different k, d when $n = 50$ and $\rho = 0.80$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	1.1972	1.1972	1.1972	1.1972	1.1972	1.1972	1.1972	1.1972	1.1972	1.1972
QL	0.9506	0.9506	0.9506	0.9506	0.9506	0.9506	0.9506	0.9506	0.9506	0.9506
LRE	1.1146	1.0425	0.9792	0.9234	0.8741	0.8303	0.7914	0.7566	0.7255	0.7002
LLE	0.7438	0.7918	0.8416	0.8931	0.9465	1.0016	1.0585	1.1172	1.1777	1.2336
AURLE	1.1939	1.1850	1.1719	1.1557	1.1373	1.1172	1.0960	1.0743	1.0522	1.0322
AULLE	1.0601	1.0877	1.1126	1.1345	1.1534	1.1690	1.1813	1.1901	1.1954	1.1972
OGLE	0.1847	0.1847	0.1847	0.1847	0.1847	0.1847	0.1847	0.1847	0.1847	0.1847

Table B6. The estimated MSE values for different k, d when $n = 50$ and $\rho = 0.90$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	2.2492	2.2492	2.2492	2.2492	2.2492	2.2492	2.2492	2.2492	2.2492	2.2492
QL	1.7663	1.7663	1.7663	1.7663	1.7663	1.7663	1.7663	1.7663	1.7663	1.7663
LRE	1.9574	1.7314	1.5535	1.4119	1.2979	1.2053	1.1296	1.0672	1.0157	0.9768
LLE	1.0825	1.2002	1.3260	1.4598	1.6016	1.7515	1.9095	2.0756	2.2497	2.4133
AURLE	2.2265	2.1717	2.0993	2.0183	1.9344	1.8512	1.7705	1.6938	1.6215	1.5604
AULLE	1.6689	1.7787	1.8808	1.9734	2.0546	2.1231	2.1775	2.2171	2.2411	2.2491
OGLE	0.3286	0.3286	0.3286	0.3286	0.3286	0.3286	0.3286	0.3286	0.3286	0.3286

Table B7. The estimated MSE values for different k, d when $n = 50$ and $\rho = 0.95$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	4.3561	4.3561	4.3561	4.3561	4.3561	4.3561	4.3561	4.3561	4.3561	4.3561
QL	3.4116	3.4116	3.4116	3.4116	3.4116	3.4116	3.4116	3.4116	3.4116	3.4116
LRE	3.3465	2.7191	2.3105	2.0357	1.8467	1.7149	1.6222	1.5570	1.5114	1.4827
LLE	1.6997	1.94973	2.2300	2.5406	2.8816	3.2530	3.6547	4.0868	4.5492	4.9914
AURLE	4.2065	3.9088	3.5828	3.2743	2.9989	2.7595	2.5542	2.3794	2.2309	2.1166
AULLE	2.4119	2.7346	3.0570	3.3649	3.6462	3.8905	4.0892	4.2359	4.3258	4.3558
OGLE	0.6202	0.6202	0.6202	0.6202	0.6202	0.6202	0.6202	0.6202	0.6202	0.6202

Table B8. The estimated MSE values for different k, d when $n = 50$ and $\rho = 0.99$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	21.1970	21.1970	21.1970	21.1970	21.1970	21.1970	21.1970	21.1970	21.1970	21.1970
QL	16.6063	16.6063	16.6063	16.6063	16.6063	16.6063	16.6063	16.6063	16.6063	16.6063
LRE	8.7904	6.8935	6.6305	6.7506	6.9641	7.1883	7.3985	7.5881	7.7569	7.8924
LLE	8.5248	9.4676	10.7349	12.3268	14.2432	16.4842	19.0497	21.9398	25.1544	28.3250
AURLE	14.5517	10.0876	8.0735	7.1683	6.7783	6.6407	6.6316	6.6887	6.7795	6.8750
AULLE	6.8128	7.8577	9.6505	11.8647	14.2173	16.4687	18.4231	19.9279	20.8743	21.1938
OGLE	2.9652	2.9652	2.9652	2.9652	2.9652	2.9652	2.9652	2.9652	2.9652	2.9652

Table B9. The estimated MSE values for different k, d when $n = 100$ and $\rho = 0.80$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	0.5783	0.5783	0.5783	0.5783	0.5783	0.5783	0.5783	0.5783	0.5783	0.5783
QL	0.4066	0.4066	0.4066	0.4066	0.4066	0.4066	0.4066	0.4066	0.4066	0.4066
LRE	0.5587	0.5404	0.5232	0.5072	0.4921	0.4780	0.4647	0.4522	0.4404	0.4303
LLE	0.4437	0.4584	0.4734	0.4887	0.5042	0.5201	0.5361	0.5525	0.5691	0.5843
AURLE	0.5780	0.5769	0.5752	0.5730	0.5703	0.5672	0.5638	0.5600	0.5560	0.5523
AULLE	0.5566	0.5612	0.5653	0.5688	0.5718	0.5743	0.5762	0.5776	0.5784	0.5787
OGLE	0.1009	0.1009	0.1009	0.1009	0.1009	0.1009	0.1009	0.1009	0.1009	0.1009

Table B10. The estimated MSE values for different k, d when $n = 100$ and $\rho = 0.90$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	1.0909	1.0909	1.0909	1.0909	1.0909	1.0909	1.0909	1.0909	1.0909	1.0909
QL	0.7541	0.7541	0.7541	0.7541	0.7541	0.7541	0.7541	0.7541	0.7541	0.7541
LRE	1.0183	0.9548	0.8988	0.8494	0.8054	0.7662	0.7311	0.6996	0.6712	0.6479
LLE	0.6871	0.7302	0.7749	0.8211	0.8688	0.9181	0.9689	1.0212	1.0751	1.1249
AURLE	1.0881	1.0807	1.0696	1.0558	1.0401	1.0229	1.0047	0.9860	0.9669	0.9497
AULLE	0.9734	0.9969	1.0182	1.0370	1.0532	1.0666	1.0772	1.0848	1.0894	1.0909
OGLE	0.1828	0.1828	0.1828	0.1828	0.1828	0.1828	0.1828	0.1828	0.1828	0.1828

Table B11. The estimated MSE values for different k, d when $n = 100$ and $\rho = 0.95$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	2.1225	2.1225	2.1225	2.1225	2.1225	2.1225	2.1225	2.1225	2.1225	2.1225
QL	1.4556	1.4556	1.4556	1.4556	1.4556	1.4556	1.4556	1.4556	1.4556	1.4556
LRE	1.8541	1.6461	1.4819	1.3503	1.2434	1.1557	1.0830	1.0222	0.9710	0.9316
LLE	1.0313	1.1425	1.2610	1.3869	1.5202	1.6609	1.8090	1.9645	2.1273	2.2803
AURLE	2.1022	2.0532	1.9882	1.9153	1.8397	1.7645	1.6916	1.6219	1.5561	1.5004
AULLE	1.5966	1.6948	1.7869	1.8707	1.9446	2.0070	2.0568	2.0931	2.1151	2.1224
OGLE	0.3468	0.3468	0.3468	0.3468	0.3468	0.3468	0.3468	0.3468	0.3468	0.3468

Table B12. The estimated MSE values for different k, d when $n = 100$ and $\rho = 0.99$.

k/d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
MLE	10.3789	10.3789	10.3789	10.3789	10.3789	10.3789	10.3789	10.3789	10.3789	10.3789
QL	7.0826	7.0826	7.0826	7.0826	7.0826	7.0826	7.0826	7.0826	7.0826	7.0826
LRE	6.0176	4.4448	3.7460	3.4006	3.2210	3.1266	3.0790	3.0582	3.0533	3.0371
LLE	3.4749	4.0172	4.6848	5.4778	6.3960	7.4395	8.6083	9.9024	11.3218	12.7064
AURLE	8.9759	7.2616	6.0160	5.1588	4.5664	4.1508	3.8545	3.6403	3.4839	3.3788
AULLE	3.8512	4.6207	5.5646	6.5836	7.5915	8.5155	9.2959	9.8866	10.2542	10.3777
OGLE	1.6505	1.6505	1.6505	1.6505	1.6505	1.6505	1.6505	1.6505	1.6505	1.6505

Table B13. The SMSE values of estimators for the real data example.

	MLE	QL	LRE	LLE	AURLE	AULLE	OGLE
$k;d = 0.1$	0.0007596203	0.0003231272	0.0007595157	0.0007586909	0.0007596203	0.0007596197	0.0000171850
$k;d = 0.2$	0.0007596203	0.0003231272	0.0007594114	0.0007587955	0.0007596203	0.0007596198	0.0000171850
$k;d = 0.3$	0.0007596203	0.0003231272	0.0007593074	0.0007589000	0.0007596202	0.0007596199	0.0000171850
$k;d = 0.4$	0.0007596203	0.0003231272	0.0007592036	0.0007590046	0.0007596202	0.0007596200	0.0000171850
$k;d = 0.5$	0.0007596203	0.0003231272	0.0007591001	0.0007591092	0.0007596201	0.0007596201	0.0000171850
$k;d = 0.6$	0.0007596203	0.0003231272	0.0007589968	0.0007592138	0.0007596200	0.0007596202	0.0000171850
$k;d = 0.7$	0.0007596203	0.0003231272	0.0007588938	0.0007593184	0.0007596199	0.0007596202	0.0000171850
$k;d = 0.8$	0.0007596203	0.0003231272	0.0007587911	0.0007594230	0.0007596198	0.0007596203	0.0000171850
$k;d = 0.9$	0.0007596203	0.0003231272	0.0007586886	0.0007595276	0.0007596197	0.0007596203	0.0000171850
$k;d = 0.99$	0.0007596203	0.0003231272	0.0007585966	0.0007596217	0.0007596195	0.0007596203	0.0000171850