JOURNAL OF APPLIED STATISTICS
2021, VOL. 48, NO. 9, 1559–1578
https://doi.org/10.1080/02664763.2020.1769576

A new kind of stochastic restricted biased estimator for logistic regression model

M. I. Alheety[a], Kristofer Månsson[b] and B. M. Golam Kibria [c,*]

[a]Department of Mathematics, College of Education for Pure Sciences, University of Anbar, Ramadi, Iraq; [b]Department of Economics, Finance and Statistics, Jönköping University, Jonkoping, Sweden; [c]Department of Mathematics and Statistics, Florida International University, Miami, FL, USA

ABSTRACT
In the logistic regression model, the variance of the maximum likelihood estimator is inflated and unstable when the multicollinearity exists in the data. There are several methods available in literature to overcome this problem. We propose a new stochastic restricted biased estimator. We study the statistical properties of the proposed estimator and compare its performance with some existing estimators in the sense of scalar mean squared criterion. An example and a simulation study are provided to illustrate the performance of the proposed estimator.

ARTICLE HISTORY
Received 19 December 2018
Accepted 9 May 2020

KEYWORDS
Logistic regression; maximum likelihood estimator; mean squared error matrix; ridge regression; simulation study; stochastic restricted estimator

MATHEMATICS SUBJECT CLASSIFICATIONS
Primary 62J05; Secondary 62J07

1. Introduction

Logistic regression analysis studies the association between a dichotomous dependent variable and a set of independent (explanatory) variables, which could be discrete, categorical or continuous in nature. The main application of logistic regression analysis is the classification of individuals in different groups. Estimation of the parameters of the logistic regression model is very important. We will discuss about the logistic regression model and the parameter estimation in this section. The logistic regression model can be written in the following form:

$$y_i = \pi_i + \epsilon_i, \tag{1.1}$$

where y_i is a random variable distributed as a Bernoulli distribution with parameter π_i as:

$$\pi_i = \frac{exp(x_i'\beta)}{1 + exp(x_i'\beta)}; \tag{1.2}$$

CONTACT Kristofer Månsson kristofer.mansson@ju.se
*The author, B.M. Golam Kibria would like to dedicate this paper to Professor A.K. Md. E. Saleh, Carleton University (from where he did his Masters), Ottawa, Canada for introducing the topic of ridge regression and his constant encouragement, love and support throughout his career.

where x_i is the ith row of the data matrix X which has an $n \times (p+1)$ with p independent explanatory variables and ϵ_i is a random variable such that $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \pi_i(1-\pi_i) = w_i$.

The maximum likelihood estimator (MLE) of β is given as follows:

$$\hat{\beta} = (X'\hat{W}X)^{-1}X'\hat{W}Z \\ = S^{-1}X'\hat{W}Z, \tag{1.3}$$

where $S = (X'\hat{W}X)$; $\hat{W} = diag(\hat{\pi}_i(1-\hat{\pi}_i))$ and Z is the column vector with $Z_i = \log(\hat{\pi}_i) + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)}$. The maximum likelihood estimator $\hat{\beta}$ is asymptotically unbiased estimator of β.

When the independent variables are correlated, the variance of the MLE will be inflated and the logistic regression model becomes unstable, especially if there exists a strong correlation. Researchers used the biased estimation technique to reduce the effect of multicollinearity. One of the most useful biased estimators that are used for a long time is called ridge regression estimator for a linear regression model, proposed by Hoerl and Kennard [9], ridge logistic regression (denoted RiMLE) which was proposed by Schaefer *et al.* [20], while Aguilera *et al.* [1] proposed the principal component, logistic estimator. In the same context, Nja *et al.* [18] introduced the modified logistic ridge regression estimator, while Inan and Erdogan, [11] Asar and Genc [5] and Asar [3] introduced the Liu-type estimator. In addition to the sample information, there are some exact or stochastic restrictions for the unknown parameter of the model exist which may help to reduce the multicollinearity problem. Therefore, suppose that we have some prior information about β in the form of independent stochastic linear restrictions as:

$$h = H\beta \, e \sim (0, \ V) \tag{1.4}$$

where h is a j × 1 random vector, H is a j × p known matrix and V is assumed to be known and positive definite. Also, it is assumed that $\epsilon$is stochastically independent of e.

Duffy and Santner [7] proposed the restricted maximum likelihood estimator (RMLE) of β for the logistic regression model with the exact prior restrictions ($e = 0$) as follows:

$$\hat{\beta}_{RMLE} = \hat{\beta} + S^{-1}H(H'S^{-1}H)(h - H\hat{\beta}) \tag{1.5}$$

Varathanand Wijekoon [22] proposed a stochastic restricted maximum likelihood estimator (SRMLE) for the logistic regression model with the stochastic linear restrictions as follows:

$$\hat{\beta}_{SRMLE} = \hat{\beta} + S^{-1}H(V + H'S^{-1}H)(h - H\hat{\beta}) \tag{1.6}$$

Asar *et al.* [4] proposed a restricted ridge estimator (RRMLE) in the logistic regression model as follows:

$$\hat{\beta}_{RRMLE} = \hat{\beta}(k) + S(k)^{-1}H(H'S(k)^{-1}H)(h - H\hat{\beta}(k)), \tag{1.7}$$

where $\hat{\beta}(k) = S(k)^{-1}X'\hat{W}Z$ is the logistic ridge estimator, which is given by Schaffer *et al.* [20] and $S(k) = (\mathrm{S} + \mathrm{kI_p})$.

Varathan and Wijekoon [23] introduced a new biased estimator which is called stochastic restricted ridge maximum likelihood estimator (SRRMLE), and is defined as:

$$\hat{\beta}_{SRRMLE} = \hat{\beta}(k) + S^{-1}H(V + H'S^{-1}H)(h - H\hat{\beta}(k)) \tag{1.8}$$

Furthermore, Varathan and Wijekoon [24] introduced the following stochastic restricted Liu maximum likelihood estimator(SRLMLE):

$$\hat{\beta}_{SRLMLE} = (S + I)^{-1}(S + dI)\hat{\beta}_{SRMLE} \tag{1.9}$$

where d is the Liu shrinkage parameter.

Also, Varathan and Wijekoon [25] proposed the following Stochastic Restricted Liu-Type Logistic Estimator (SRLTLE):

$$\hat{\beta}_{SRLMLE} = (S + kI)^{-1}(S - dI)\hat{\beta}_{SRMLE} \tag{1.10}$$

Following Hubert and Wijekoon [10], we proposed a new general biased estimator which will be called the stochastic restricted two-parameters maximum likelihood estimator (SRTPMLE) when the linear stochastic restrictions are available for the logistic regression model. The organization of the paper is as follows. The proposed estimator and its superiority are given in Section 2. The estimation of the shrinkage parameters is outlined in Section 3. To compare the performance of the estimators, a simulation study has been conducted in Section 4. An application is given in Section 5. Finally, some concluding remarks are given in Section 6.

2. Proposed estimator and its superiority

In this section, we will propose a new estimator and study its performance compare to other existing estimators.

2.1. The proposed estimator

Since the combination of two different estimators might inherit the advantage of both estimators, we propose the following biased estimator using the estimator in (1.6) which we call the stochastic restricted two-parameters maximum likelihood estimator (SRTPMLE):

$$\hat{\beta}_{SRTPMLE} = F_{k,d}\hat{\beta}_{SRMLE}, \tag{2.1}$$

where $F_{k,d} = (I + kS^{-1})^{-1}(I - (1 - d)^2(S + I)^{-2})$, $0 < d < 1$ and $k \geq 0$. The expected value; the bias and the variance matrices of the SRTPMLE are obtained as follows:

$$E(\hat{\beta}_{SRTPMLE}) = F_{k,d}\beta \tag{2.2}$$

Therefore,

$$B_1 = Bias(\hat{\beta}_{SRTPMLE}) = (F_{k,d} - I)\beta \tag{2.3}$$

$$\text{and} \qquad Cov(\hat{\beta}_{SRTPMLE}) = F_{k,d}AF_{k,d}, \tag{2.4}$$

where $A = S^{-1} - S^{-1}H'(V + HS^{-1}H')^{-1}HS^{-1}$.

We use the concept of mean square error matrix MMSE as a criterion for the goodness of fit. Firstly, we give the concept of MMSE as follows:

Definition 2.1: Let β^* be any estimator for β, then the MMSE of β^* as given by:

$$MMSE(\hat{\beta}^*) = Cov\ (\hat{\beta}^*) + Bias(\hat{\beta}^*)\ Bias(\hat{\beta}^*)'.$$

Also, the concept of scalar mean squared error is nothing but the trace of MMSE; i.e.

$$mse(\hat{\beta}^*) = tr(MMSE(\hat{\beta}^*)).$$

In order to find the superiority of any estimator $\hat{\beta}^1$ compared to other estimator $\hat{\beta}^2$ under the MMSE criterion, the estimator $\hat{\beta}^2$ is better than $\hat{\beta}^1$ with respect to MMSE sense if and only if:

$$MMSE(\hat{\beta}^1) - MMSE(\hat{\beta}^2) \geq 0,$$

that is, the difference will be positive definite (pd) or semi-positive definite (spd) matrix.

2.2. The comparison between the MLE and SRTPMLE estimators

We make the MMSE comparison between the MLE and SRTPMLE but first, the MMSE of MLE and SRTPMLE are obtained respectively as follows:

$$\text{MMSE}(\hat{\beta}) = S^{-1} \tag{2.5}$$

and $$\text{MMSE}(\hat{\beta}_{\text{SRTPMLE}}) = F_{k,d}AF_{k,d} + B_1{B_1}', \tag{2.6}$$

The difference of MMSE values between them can be given by:

$$\begin{aligned}\Delta_1 &= \text{MMSE}(\hat{\beta}) - \text{MMSE}(\hat{\beta}_{\text{SRTPMLE}}) \\ &= S^{-1} - (F_{k,d}\ A\ F_{k,d} + \ B_1{B_1}') \\ &= M - N,\end{aligned}$$

where $M = S^{-1}$ and $N = F_{k,d}AF_{k,d} + B_1{B_1}'$. Now we need to give the following Lemma 2.1 to show that under which condition Δ_1 will be pd.

Lemma 2.1: (*See Wu,* [26])**:** *Suppose that M is a positive definite matrix and N is a nonnegative definite matrix. Then:*

$$M - N \geq 0 \leftrightarrow \lambda_{max}(NM^{-1}) \leq 1.$$

In order to apply Lemma 2.1 *for this case, we have to use the following theorem:*

Theorem 2.1: (*See Rao and Toutenburg* [19])**:** *Let A: n × n and B: n × n be any two matrices such that A is a positive definite and B is a non- negative definite matrices. Then A + B is non-negative definite matrix.*

Since $B_1{B_1}'$ is a nonnegative definite matrix, S^{-1} and $F_{k,d}AF_{k,d}$ are positive definite, then using Theorem 2.1*, N is a positive definite matrix and therefore using Lemma* 2.1*, if*

$\lambda_{max}(NM^{-1}) \leq 1$, *then* $M - N$ *is a positive definite matrix . So we can state the following theorem:*

Theorem 2.2: *For 0 < d < 1 and k≥ 0, under logistic regression model with stochastic restrictions, the SRTPMLE is superior to MLE in the MMSE if and only if* $\lambda_{max}(NM^{-1}) \leq 1$.

2.3. The comparison between the SRMLE and SRTPMLE estimators

Asymptotic properties of *SRMLE* is given as follows (see Varathan and Wijekoon [24]):

The expected value; the bias and the variance matrices of the SRMLE are obtained as follows:

$$E(\hat{\beta}_{SRMLE}) = E(\hat{\beta} + S^{-1}H(V + H'S^{-1}H)(h - H\hat{\beta})).$$

But $\hat{\beta}$ is asymptotically unbiased, therefore,

$$\begin{aligned} E(\hat{\beta}_{SRMLE}) &= E(\hat{\beta}) + \mathrm{S}^{-1}\mathrm{H}(\mathrm{V} + \mathrm{H}'\mathrm{S}^{-1}\mathrm{H})E(h) - \mathrm{S}^{-1}\mathrm{H}(\mathrm{V} + \mathrm{H}'\mathrm{S}^{-1}\mathrm{H})E(H\hat{\beta}) \\ &= \beta + \mathrm{S}^{-1}\mathrm{H}(\mathrm{V} + \mathrm{H}'\mathrm{S}^{-1}\mathrm{H})E(H\beta + e) - \mathrm{S}^{-1}\mathrm{H}(\mathrm{V} + \mathrm{H}'\mathrm{S}^{-1}\mathrm{H})\mathrm{H}\beta \\ &= \beta + \mathrm{S}^{-1}\mathrm{H}(\mathrm{V} + \mathrm{H}'\mathrm{S}^{-1}\mathrm{H})\mathrm{H}\beta - \mathrm{S}^{-1}\mathrm{H}(\mathrm{V} + \mathrm{H}'\mathrm{S}^{-1}\mathrm{H})\mathrm{H}\beta, \end{aligned}$$

where $\mathrm{E(e)} = 0$.

Therefore;

$$E(\hat{\beta}_{SRMLE}) = \beta.$$

Then

$$Bias(\hat{\beta}_{SRMLE}) = 0$$

$$\text{and} \quad Cov(\hat{\beta}_{SRMLE}) = A\,.$$

The MMSE of SRMLE is given as follows:

$$\mathrm{MMSE}(\hat{\beta}_{\mathrm{SRMLE}}) = A. \tag{2.7}$$

Therefore, the difference of MMSE values between SRMLE and SRTPMLE can be given by:

$$\begin{aligned} \Delta_2 &= \mathrm{MMSE}(\hat{\beta}_{\mathrm{SRMLE}}) - \mathrm{MMSE}(\hat{\beta}_{\mathrm{SRTPMLE}}) \\ &= A - (F_{k,d}\, A\, F_{k,d} + B_1 B_1{}') \\ &= S^{-1} - (D + F_{k,d}\, A\, F_{k,d} + B_1 B_1{}') \\ &= M - N_1, \end{aligned}$$

where $D = \mathrm{S}^{-1}H'(\mathrm{V} + \mathrm{HS}^{-1}\mathrm{H}')^{-1}\mathrm{HS}^{-1}$ and $N_1 = D + F_{k,d}AF_{k,d} + B_1B_1'$.

Since D and $F_{k,d}AF_{k,d}$ are positive definite and B_1B_1' is a non- negative definite matrices, then using Theorem 2.1; N_1 is a positive definite matrix. Therefore, by Lemma 2.1 we can state the following theorem:

Theorem 2.3: *For 0 < d < 1 and k≥ 0, under logistic regression model with stochastic restrictions* (1.4), *the SRTPMLE is superior to* SRMLE *in the MMSE if and only if* $\lambda_{max}(N_1M^{-1}) \leq 1$.

2.4. The comparison between the SRLTLE and SRTPMLE estimators

We make the MMSE comparison between the SRLTLE and SRTPMLE, where the MMSE of SRLTLE is obtained as follows:

$$\mathrm{MMSE}(\hat{\beta}_{\mathrm{SRLTLE}}) = Z_{k,d}\, A\, Z_{k,d} + B_2 B_2{}', \tag{2.8}$$

where $B_2 = (Z_{k,d} - I)\beta$. The difference of MMSE values between SRLTLE and SRTPMLE can be given by:

$$\begin{aligned}\Delta_2 &= \mathrm{MMSE}(\hat{\beta}_{\mathrm{SRLTLE}}) - MMSE(\hat{\beta}_{\mathrm{SRTPMLE}}) \\ &= D_1 + B_2 B_2' - B_1 B_1',\end{aligned}$$

where

$$\begin{aligned}D_1 &= (S+kI)^{-1} SAS(S+kI)^{-1} + d^2(S+kI)^{-1}A(S+kI)^{-1} - 2\mathrm{d}(S{+}kI)^{-1}AS(S{+}kI)^{-1} \\ &= M_1 - N_2\end{aligned}$$

where

$$\begin{aligned}&M_1 = (S+kI)^{-1}SAS(S+kI)^{-1} + d^2(S+kI)^{-1}A(S+kI)^{-1} \text{ and} \\ &N_2 = 2d(S+kI)^{-1}AS(S+kI)^{-1}.\end{aligned}$$

By using Lemma 2.1 where M_1 and N_2 are positive definite matrices, if $\lambda_{max}(N_2 M_1{}^{-1}) \leq 1$,then $D_1 = M_1 - N_2$ is positive definite matrix. In order to determine the condition that makes Δ_2 positive definite, we have to use the following Lemma 2.2:

Lemma 2.2: *(See Trenkler G and Toutenburg H [21]):*
Let $\tilde{\beta}_j = A_j Y, j = 1, 2$ be two competing homogeneous linear estimators of β. Suppose that $D = Cov(\tilde{\beta}_1) - Cov(\tilde{\beta}_2) > 0$, where $Cov(\tilde{\beta}_j)$ denotes the covariance matrix of $\tilde{\beta}_j$. Then $\Delta = MMSE(\tilde{\beta}_1) - MMSE(\tilde{\beta}_2) \geq 0$ if and only if $d_2'(D + d_1{}'d_1)^{-1}d_2 \leq 1$, where d_j: $j = 1,2$ denotes the bias vector of $\tilde{\beta}_j$.

Hence, the following theorem can be stated.

Theorem 2.4: *If $\lambda_{max}(N_2 M_1{}^{-1}) \leq 1$, the estimator SRTPMLE is superior to SRLTLE if and only if*

$$B_1'(D + B_2{}'B_2)^{-1}B_1 \leq 1.$$

Since the proposed estimator depends on the unknown parameters, d and k, we will discuss their estimation techniques in the following section.

3. Estimating the shrinkage parameters d and k

In order to estimate the shrinkage parameters d and $k$and since there does not exist a definite rule of how to estimate them (see Månsson *et al.* [16]), we follow Hoerl and Kennard [9], Kibria [13], Khalaf and Shukur [12] and Muniz and Kibria [17], where the optimal value of the ridge parameter was estimated. Following Månsson *et al.* [15], the least mean squared error of SRTPMLE can be obtained by finding the optimal value of the biasing parameter d for fixed k and also for the biased parameter k for fixed d. Therefore, we have to find d and k that achieve the desired performance for the proposed estimator.

For the sake,

$$mse(\hat{\alpha}_{\text{SRTPMLE}}) = \text{tr}(\text{MMSE}(\hat{\alpha}_{\text{SRTPMLE}}))$$
$$= \sum_{i=1}^{p} \frac{(\lambda_i(\lambda_i+2)+d(2-d))\lambda_i^2 a_{ii}}{(\lambda_i+1)^2(\lambda_i+k)^2} + \sum_{i=1}^{p}\left[\frac{(\lambda_i(\lambda_i+2)+d(2-d))\lambda_i}{(\lambda_i+1)^2(\lambda_i+k)^2} - 1\right]^2 \alpha_i^2$$

where α_i^2 is defined as the ith element $P'\beta$ and P is a $p \times p$ eigenvectors matrix, such that $P'SP = \Lambda$, where Λ is a $p \times p$ diagonal matrix and its elements $\lambda_1, \lambda_2, \ldots., \lambda_p$ are the eigenvalues of $P'SP$, such that $\lambda_1 > \lambda_2 > \ldots > \lambda_p$ and $a_{ii} \geq 0$ is the diagonal element of the matrix A, which is defined under Equation (2.4). Let k be fixed, minimize $mse(\hat{\alpha}_{\text{SRTPMLE}})$ as a function and find the derivative with respect to $d$and after some simplifications, the optimal d will be given as follows:

$$d = 1 - \frac{\sum_{i=1}^{p} \frac{(\lambda_i a_{ii} - k\alpha_i^2)}{(\lambda_i+1)^2(\lambda_i+k)^2}}{\sum_{i=1}^{p} \frac{\lambda_i^2(a_{ii}+\alpha_i^2)}{(\lambda_i+1)^4(\lambda_i+k)^2}}. \tag{3.1}$$

The optimal value of d in (3.1) depends on the unknown parameter, α_i^2. Therefore, we replaced it with its unbiased estimator $\hat{\alpha}_i^2$ to get the following:

$$\hat{d} = 1 - \frac{\sum_{i=1}^{p} \frac{(\lambda_i a_{ii} - k\hat{\alpha}_i^2)}{(\lambda_i+1)^2(\lambda_i+k)^2}}{\sum_{i=1}^{p} \frac{\lambda_i^2(a_{ii}+\hat{\alpha}_i^2)}{(\lambda_i+1)^4(\lambda_i+k)^2}}. \tag{3.2}$$

where we bound it to be between 0 and 1. By the same way; for fixed d, the optimal value of k will be given as follows:

$$k_i = \frac{(\lambda_i(\lambda_i+2)+d(2-d))\lambda_i a_{ii} - (1-d)\lambda_i\alpha_i^2}{(\lambda_i+1)^2\alpha_i^2}, \tag{3.3}$$

where the value of k in (3.3) is getting by equating the numerator to zero as proposed by Hoerl and Kennard [9]. To get the estimated optimal value of k, we refer our readers to Kibria [13], Muniz and Kibria [17], Goktas and Sevinc [8] and Kibria and Banik [14] among others. We use firstly the following classical ridge estimators:

$$k_1 = \frac{1}{\sum \hat{\alpha}_i^2} \text{ and } k_2 = \frac{1}{\hat{\alpha}_{\max}^2} \tag{3.4}$$

which are versions of the once suggested by Hoerl and Kennard [9] for linear regression model. We however, do not use the residual variance since the estimated parameters are based on an iterative weighted least squares and it is theoretically equal to one. Then following Kibria [13], we consider the following two estimators:

$$k_3 = \frac{1}{\left(\prod_{i=1}^{p} \hat{\alpha}_i^2\right)^{\frac{1}{p}}} \text{ and } k_4 = Median\{m_i^2\}, \tag{3.5}$$

where $m_i = \sqrt{\frac{1}{\hat{\alpha}_i^2}}$. Another estimator used to estimate k is the following:

$$k_5 = \max(s_i), \tag{3.6}$$

first proposed by Alkhamisi *et al.* [2] for the linear regression model where $s_i = \frac{\lambda_i}{(n-p)+\lambda_i\hat{\alpha}_i^2}$. Finally following Muniz & Kibria [17], we consider the following estimators:

$$k_6 = \max\left(\frac{1}{m_i}\right),\ k_7 = \left(\prod_{i=1}^{p} \frac{1}{m_i}\right)^{\frac{1}{p}},\ k_8 = \left(\prod_{i=1}^{p} m_i\right)^{\frac{1}{p}}\ k_9 = median\left(\frac{1}{m_i}\right), \tag{3.7}$$

The above estimators were shown to work well in the presence of moderate to a high degree of multicollinearity (see, for examples, Kibria [13] and Muniz and Kibria [17]). For the new two-parameter estimator suggested in Equation (2.1), the estimated values of k will be plugged into Equation (3.2) in order to estimate the second parameter d. Furthermore in order to use the optimal value of ki presented in Equation 3.3 we use the following estimators:

$$k_{opt1} = \sum \hat{k}_i,\ k_{opt2} = median(\hat{k}_i) \text{ and } k_{opt3} = \max(\hat{k}_i), \tag{3.8}$$

where the values in Equation (3.3) are replaced with their estimated counterparts. Finally these values are plugged into Equation (3.2) to estimate the optimal value of d. As a start up-value for d which is needed we use the optimal value of ka simple Liu estimator derived in Månsson *et al.* [15]. This is in line with suggestions in Varathan and Wijekoon [24] and defined as follows:

$$D2 = \max\left(0, \text{median}\left(\frac{\hat{\alpha}_i^2 - 1}{\frac{1}{\hat{\lambda}_i} + \hat{\alpha}_i^2}\right)\right)$$

4. Simulation study

A Monte Carlo simulation study has been conducted to illustrate the performance of the proposed estimator, SRTPMLE over the existing estimators (ML, RML, SRML) in this section.

4.1. Simulation technique

The dependent variable is a binary variable which will be modeled using a logit regression model. Therefore it is generated using pseudo-random numbers from the

$Be(\pi_i)$distribution, where

$$\pi_i = \frac{\exp(x_i{}'\beta)}{1 + \exp(x_i{}'\beta)}. \tag{4.1}$$

The parameter values in Equation (4.1) are chosen so that sum of squared parameters equals one. This is an arbitrary choice and we have changed values and generated them using uniform distribution instead ($U(0,1)$ and $U(-1,1)$). This did not change the pattern of well-performing estimators. In the design of the experiment, we first vary the degree of correlation values as

Table 1. MLE and RLMLE with $p = 4$.

	MLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					ρ	$= 0.75$				
80	1.16	0.80	0.65	0.33	0.36	1.12	0.46	0.70	0.33	0.65
120	0.67	0.49	0.42	0.25	0.27	0.65	0.40	0.52	0.27	0.50
200	0.36	0.29	0.25	0.17	0.19	0.35	0.28	0.32	0.19	0.31
400	0.17	0.15	0.14	0.10	0.12	0.17	0.15	0.16	0.12	0.16
n					ρ	$= 0.85$				
80	1.97	1.27	1.00	0.38	0.42	1.86	0.49	0.88	0.40	0.78
120	1.13	0.77	0.63	0.28	0.31	1.08	0.49	0.73	0.33	0.68
200	0.60	0.45	0.38	0.19	0.21	0.58	0.40	0.50	0.25	0.47
400	0.28	0.23	0.21	0.12	0.14	0.28	0.24	0.26	0.16	0.26
n										
80	6.20	3.61	2.75	0.74	0.84	5.21	0.33	0.87	0.55	0.68
120	3.55	2.12	1.63	0.49	0.54	3.09	0.45	0.99	0.49	0.82
200	1.90	1.19	0.93	0.30	0.34	1.71	0.54	0.93	0.40	0.82
400	0.89	0.62	0.50	0.19	0.21	0.83	0.48	0.65	0.30	0.61
n					ρ	$= 0.99$				
80	32.63	17.93	13.22	3.27	3.72	17.03	0.10	0.25	0.86	0.18
120	18.90	10.45	7.73	1.94	2.23	10.73	0.12	0.43	0.78	0.28
200	10.22	5.69	4.22	1.11	1.24	6.41	0.21	0.69	0.68	0.49
400	3.92	2.24	1.68	0.48	0.53	2.82	0.34	0.80	0.46	0.63

Table 2. RMLE and RRMLE with $p = 4$.

	RMLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					ρ	$= 0.75$				
80	1.73	1.58	1.53	1.48	1.50	1.71	1.49	1.58	1.43	1.56
120	1.56	1.48	1.45	1.43	1.45	1.55	1.46	1.51	1.39	1.50
200	1.47	1.43	1.41	1.41	1.42	1.47	1.43	1.45	1.39	1.45
400	1.33	1.32	1.31	1.30	1.31	1.33	1.32	1.33	1.30	1.33
n										
80	1.99	1.71	1.62	1.48	1.50	1.96	1.50	1.66	1.44	1.62
120	1.72	1.56	1.51	1.43	1.45	1.70	1.49	1.59	1.41	1.57
200	1.57	1.49	1.45	1.40	1.42	1.56	1.49	1.53	1.40	1.52
400	1.38	1.35	1.34	1.30	1.32	1.38	1.36	1.37	1.32	1.37
n					ρ	$= 0.95$				
80	3.29	2.34	2.09	1.54	1.57	3.03	1.42	1.65	1.46	1.57
120	2.53	1.96	1.80	1.46	1.48	2.40	1.46	1.70	1.44	1.63
200	2.04	1.72	1.62	1.42	1.44	1.98	1.54	1.71	1.44	1.67
400	1.62	1.48	1.43	1.32	1.33	1.60	1.46	1.54	1.35	1.52
n					ρ	$= 0.99$				
80	10.92	6.11	4.91	2.10	2.24	6.88	1.34	1.38	1.52	1.36
120	7.26	4.20	3.44	1.78	1.87	4.97	1.32	1.43	1.49	1.38
200	4.84	3.05	2.59	1.60	1.65	3.68	1.36	1.57	1.48	1.48
400	2.53	1.80	1.60	1.20	1.22	2.17	1.20	1.42	1.20	1.34

$\rho = (0.75, 0.85, 0.95$ and $0.99)$, which represent a moderate to a high degree of multicollinearity. The independent variables with different degrees of correlation are generated as:

$$x_{ij} = (1 - \rho^2)^{(1/2)} z_{ij} + \rho z_{ip}, \ i = 1, 2, \dots n, \ j = 1, 2, \dots p \tag{4.2}$$

where z_{ij} are pseudo-random numbers generated using the standard normal distribution. We choose to discard the first 200 observations to reduce eventual start-up value effects. Secondly, we also choose to vary the number of independent variables denoted p and we consider models where p equals 4, 8 and 12. Thirdly, we change the number of observations

Table 3. SRMLE and SRRMLEwith $p = 4$.

	SRMLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					$\rho = 0.75$					
80	0.73	0.53	0.45	0.30	0.32	0.71	0.37	0.51	0.28	0.48
120	0.48	0.37	0.32	0.23	0.25	0.47	0.32	0.40	0.23	0.38
200	0.29	0.24	0.21	0.16	0.18	0.28	0.23	0.26	0.17	0.26
400	0.15	0.13	0.12	0.10	0.11	0.15	0.14	0.14	0.11	0.14
n					$\rho = 0.85$					
80	1.11	0.75	0.62	0.32	0.35	1.06	0.41	0.64	0.32	0.58
120	0.73	0.52	0.44	0.25	0.27	0.70	0.40	0.54	0.27	0.51
200	0.44	0.34	0.29	0.18	0.20	0.43	0.32	0.38	0.21	0.37
400	0.23	0.20	0.18	0.11	0.13	0.23	0.20	0.22	0.15	0.21
n					$\rho = 0.95$					
80	2.71	1.61	1.29	0.48	0.53	2.41	0.36	0.71	0.43	0.60
120	1.80	1.11	0.90	0.36	0.39	1.64	0.42	0.74	0.38	0.66
200	1.12	0.72	0.59	0.25	0.27	1.04	0.45	0.69	0.31	0.63
400	0.60	0.43	0.36	0.16	0.18	0.57	0.38	0.48	0.24	0.46
n					$\rho = 0.99$					
80	10.79	5.80	4.52	1.34	1.49	6.62	0.21	0.40	0.67	0.33
120	6.99	3.76	2.92	0.91	1.03	4.60	0.22	0.49	0.58	0.39
200	4.40	2.45	1.91	0.63	0.69	3.15	0.29	0.65	0.51	0.52
400	1.96	1.14	0.89	0.31	0.34	1.56	0.33	0.64	0.34	0.54

Table 4. SRTPMLE with $p = 4$.

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					$\rho = 0.75$				
80	0.43	0.30	0.36	0.36	0.68	0.22	0.36	0.20	0.32
120	0.24	0.23	0.30	0.30	0.46	0.22	0.33	0.16	0.30
200	0.12	0.16	0.22	0.23	0.28	0.19	0.24	0.12	0.23
400	0.08	0.10	0.13	0.14	0.14	0.12	0.14	0.09	0.13
n					$\rho = 0.85$				
80	0.44	0.37	0.32	0.32	0.97	0.20	0.38	0.18	0.32
120	0.25	0.28	0.26	0.27	0.66	0.22	0.38	0.15	0.34
200	0.12	0.20	0.20	0.21	0.41	0.23	0.32	0.13	0.30
400	0.08	0.14	0.12	0.14	0.22	0.17	0.20	0.10	0.20
n					$\rho = 0.95$				
80	0.46	0.66	0.25	0.27	2.14	0.16	0.32	0.13	0.26
120	0.25	0.48	0.20	0.22	1.45	0.16	0.38	0.12	0.30
200	0.11	0.33	0.16	0.18	0.90	0.19	0.41	0.11	0.34
400	0.07	0.22	0.10	0.12	0.52	0.23	0.36	0.11	0.33
n					$\rho = 0.99$				
80	0.52	2.06	0.31	0.39	3.98	0.16	0.11	0.09	0.11
120	0.30	1.35	0.23	0.28	3.07	0.08	0.13	0.08	0.10
200	0.12	0.91	0.17	0.20	2.37	0.08	0.23	0.08	0.17
400	0.04	0.45	0.10	0.11	1.33	0.12	0.33	0.07	0.25

denoted n. In order to receive some stability in the model, a common rule of thumb is to at least have 20 observations per variable. Therefore we use models consisting of $20p$, $30p$, $50p$ and $100p$ respectively. The restrictions when $p = 4$ are set as:

$$H = \begin{bmatrix} 1 & 0 & -2 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \text{ and h} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{4.3}$$

The restrictions when $p = 8$ are:

$$H = \begin{bmatrix} 1 & 0 & -2 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -2 & 0 & 1 \end{bmatrix} \text{ and h} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4.4}$$

Table 5. SRLMLE and SRTPMLE with $p = 4$.

	SRLMLE	k_{opt1}	k_{opt2}	k_{opt3}
n		$\rho = 0.75$		
80	0.90	0.59	0.37	0.74
120	0.37	0.53	0.30	0.70
200	0.31	0.45	0.23	0.62
400	0.23	0.33	0.14	0.48
n		$\rho = 0.85$		
80	1.27	0.54	0.34	0.69
120	0.70	0.50	0.28	0.66
200	0.65	0.43	0.21	0.60
400	0.26	0.32	0.14	0.48
n		$\rho = 0.95$		
80	2.36	0.54	0.58	0.58
120	1.87	0.42	0.30	0.53
200	0.95	0.37	0.19	0.51
400	0.67	0.30	0.12	0.44
n		$\rho = 0.99$		
80	10.44	2.52	4.81	1.25
120	8.17	1.19	2.29	0.55
200	4.35	0.50	0.88	0.35
400	2.06	0.23	0.24	0.26

Table 6. MLE and RiMLE with $p = 8$.

	MLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					$\rho = 0.75$					
160	0.47	0.38	0.30	0.09	0.10	0.45	0.27	0.38	0.15	0.36
240	0.55	0.48	0.41	0.17	0.19	0.54	0.45	0.52	0.29	0.51
400	0.45	0.40	0.35	0.16	0.17	0.45	0.39	0.43	0.25	0.43
800	0.21	0.20	0.18	0.10	0.11	0.21	0.20	0.21	0.15	0.21
n					$\rho = 0.85$					
160	2.55	1.98	1.50	0.31	0.39	2.41	0.93	1.64	0.56	1.52
240	0.96	0.80	0.64	0.18	0.22	0.93	0.66	0.84	0.37	0.82
400	0.81	0.68	0.56	0.17	0.20	0.78	0.60	0.73	0.35	0.71
800	0.38	0.34	0.30	0.12	0.14	0.37	0.33	0.36	0.22	0.36
n										
160	8.60	6.17	4.54	0.63	0.90	7.13	0.74	2.23	0.87	1.92
240	4.95	3.62	2.68	0.42	0.59	4.30	0.95	2.15	0.74	1.91
400	2.62	1.98	1.52	0.26	0.37	2.37	0.97	1.69	0.58	1.57
800	1.22	0.98	0.78	0.17	0.26	1.16	0.75	1.02	0.42	0.98
n					$\rho = 0.99$					
160	46.43	31.52	23.39	2.82	4.05	22.92	0.14	0.87	1.46	0.64
240	17.47	12.07	9.11	1.20	1.69	10.62	0.41	1.92	1.18	1.54
400	14.85	10.29	7.78	1.03	1.42	9.43	0.53	2.17	1.14	1.79
800	6.98	4.92	3.74	0.53	0.73	5.14	0.92	2.34	0.87	2.08

Finally, the restrictions when $p = 12$ are:

$$H = \begin{bmatrix} 1 & 0 & -2 & 1 & 1 & 1 & -1 & -1 & 0 & 1 & -2 & 1 \\ 1 & -1 & 1 & -1 & 1 & 0 & -2 & 1 & -1 & -1 & 1 & 1 \end{bmatrix} \text{ and h} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (4.5)$$

The error term from Equation (1.4) is generated using a standard normal distribution. Hence, where $e \sim N(0, V)$, $V = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. The results of the simulation study depends on the choice of stochastic restriction. Preliminary results shows that if we choose to generate an error with no variation the common restricted estimator will be preferred and as the variance of the error term increases the results are more in favor of a stochastic restricted estimator. We choose the value one in variance which is common in previous literature (see

Table 7. RMLE and RRMLE with $p = 8$.

	RMLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					ρ	$= 0.75$				
160	0.36	0.31	0.26	0.14	0.14	0.36	0.25	0.32	0.17	0.31
240	0.58	0.53	0.48	0.32	0.33	0.57	0.52	0.56	0.40	0.56
400	0.52	0.49	0.45	0.31	0.32	0.52	0.49	0.51	0.39	0.51
800	0.37	0.36	0.35	0.28	0.29	0.37	0.36	0.37	0.32	0.37
n					ρ	$= 0.85$				
160	1.69	1.40	1.13	0.42	0.48	1.63	0.87	1.29	0.59	1.23
240	0.83	0.74	0.64	0.33	0.36	0.82	0.68	0.78	0.47	0.76
400	0.75	0.67	0.59	0.32	0.35	0.74	0.64	0.71	0.45	0.70
800	0.49	0.46	0.43	0.29	0.30	0.49	0.46	0.48	0.38	0.48
n					ρ	$= 0.95$				
160	4.68	3.62	2.78	0.63	0.80	4.22	0.74	1.81	0.81	1.59
240	2.96	2.34	1.82	0.48	0.60	2.74	0.91	1.74	0.72	1.58
400	1.77	1.46	1.18	0.37	0.45	1.69	0.94	1.39	0.61	1.32
800	1.00	0.87	0.74	0.32	0.36	0.97	0.78	0.91	0.51	0.90
n					ρ	$= 0.99$				
160	22.28	16.66	12.62	2.10	2.90	14.71	0.29	0.83	1.22	0.65
240	9.06	6.92	5.35	1.02	1.36	6.96	0.48	1.63	1.03	1.33
400	7.92	6.05	4.66	0.91	1.19	6.23	0.58	1.84	1.01	1.54
800	3.94	3.09	2.43	0.56	0.71	3.40	0.89	1.99	0.83	1.78

Table 8. SRMLE and SRRMLE with $p = 8$.

	SRMLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					ρ	$= 0.75$				
160	0.34	0.29	0.23	0.09	0.10	0.33	0.22	0.29	0.13	0.28
240	0.45	0.40	0.35	0.17	0.18	0.45	0.39	0.43	0.26	0.43
400	0.39	0.35	0.31	0.15	0.16	0.38	0.34	0.37	0.23	0.37
800	0.19	0.18	0.17	0.10	0.11	0.19	0.18	0.19	0.14	0.19
n										
160	1.65	1.34	1.06	0.29	0.35	1.59	0.77	1.22	0.48	1.15
240	0.73	0.63	0.52	0.18	0.21	0.71	0.56	0.67	0.33	0.65
400	0.63	0.55	0.47	0.17	0.19	0.62	0.50	0.59	0.31	0.58
800	0.32	0.30	0.26	0.11	0.13	0.32	0.29	0.31	0.21	0.31
n					ρ	$= 0.95$				
160	4.68	3.61	2.75	0.53	0.72	4.20	0.66	1.75	0.73	1.53
240	2.94	2.30	1.77	0.37	0.50	2.71	0.83	1.67	0.63	1.51
400	1.73	1.40	1.11	0.25	0.33	1.63	0.84	1.31	0.51	1.24
800	0.90	0.76	0.62	0.16	0.21	0.87	0.65	0.80	0.38	0.78
n					ρ	$= 0.99$				
160	22.30	16.67	12.63	2.05	2.86	14.69	0.20	0.78	1.18	0.60
240	9.06	6.91	5.33	0.95	1.30	6.93	0.41	1.58	0.97	1.29
400	7.91	6.02	4.64	0.83	1.12	6.19	0.50	1.79	0.95	1.49
800	3.91	3.04	2.38	0.46	0.62	3.35	0.81	1.91	0.74	1.71

for example, Varathan and Wijekoon [23,24]). In practice, the choice of restriction and particular stochastic restriction should come from theory along with some testing if this theory holds. An example of an often used theory is common returns to scale in a Cobb–Douglas production function. The measure of performance is the following commonly used MSE:

$$MSE = \frac{\sum_{i=1}^{10000} (\hat{\beta} - \beta)^2}{10000} \tag{4.6}$$

where $\hat{\beta}$ is the estimated parameter using various methods and β is the true value of the parameter. We use 10,000 replicates.

Table 9. SRTPMLE with $p = 8$.

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					$\rho = 0.75$				
160	0.08	0.16	0.10	0.09	0.32	0.14	0.24	0.07	0.23
240	0.14	0.27	0.20	0.18	0.44	0.33	0.41	0.17	0.40
400	0.14	0.24	0.18	0.16	0.37	0.30	0.35	0.16	0.35
800	0.11	0.15	0.12	0.11	0.19	0.17	0.19	0.11	0.18
n					$\rho = 0.85$				
160	0.23	0.68	0.25	0.23	1.43	0.40	0.86	0.19	0.78
240	0.13	0.37	0.17	0.16	0.68	0.42	0.59	0.17	0.57
400	0.14	0.34	0.16	0.15	0.60	0.39	0.53	0.17	0.52
800	0.13	0.21	0.10	0.10	0.32	0.26	0.30	0.14	0.30
n					$\rho = 0.95$				
160	0.25	1.62	0.18	0.23	3.30	0.17	0.77	0.17	0.62
240	0.13	1.07	0.14	0.17	2.17	0.28	0.93	0.17	0.79
400	0.10	0.70	0.11	0.13	1.37	0.43	0.91	0.17	0.82
800	0.14	0.42	0.08	0.09	0.79	0.45	0.67	0.17	0.64
n					$\rho = 0.99$				
160	0.34	7.67	0.41	0.70	9.42	0.08	0.11	0.12	0.08
240	0.09	3.27	0.20	0.33	4.84	0.05	0.43	0.15	0.30
400	0.08	2.82	0.17	0.28	4.36	0.06	0.56	0.16	0.41
800	0.07	1.46	0.10	0.16	2.50	0.22	0.98	0.18	0.81

Table 10. SRLMLE and SRTPMLE with $p = 8$.

	SRLMLE	k_{opt1}	k_{opt2}	k_{opt3}
n		$\rho = 0.75$		
80	0.36	0.23	0.09	0.30
120	0.51	0.59	0.18	0.84
200	0.44	0.57	0.16	0.82
400	0.24	0.46	0.11	0.73
n		$\rho = 0.85$		
80	1.94	0.62	0.24	0.85
120	0.74	0.56	0.16	0.81
200	0.71	0.55	0.15	0.80
400	0.32	0.45	0.10	0.72
n		$\rho = 0.95$		
80	2.36	0.54	0.58	0.58
120	1.87	0.42	0.30	0.53
200	0.95	0.37	0.19	0.51
400	0.67	0.30	0.12	0.44
n		$\rho = 0.99$		
80	34.56	1.87	10.72	0.58
120	15.44	0.37	1.59	0.43
200	14.53	0.33	1.04	0.44
400	5.21	0.25	0.27	0.45

4.2. Results discussion

The simulated MSEs of the estimators for different sample sizes (between 80 and 1200) are presented in Tables 1–5 for $p = 4$, in Tables 6–10 for $p = 8$ and in Tables 11–15 for $p = 12$. From these tables, we observed that as the degree of correlation among the independent variables and number of independent variables increases, the estimated MSE also inflated. The unrestricted MLE (i.e. the ordinary MLE based on iterative reweighted least squares)and the ridge version of it (RiMLE) for different number of independent variablesand sample sizes are presented in Tables 1, 6 and 10. It appears from these tables is that all ridge regression estimators dominate ML estimator in the smaller MSE sense. The RMLE

Table 11. MLE and RiMLE with $p = 12$.

	MLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					ρ	$= 0.75$				
240	1.61	1.38	1.07	0.24	0.29	1.56	1.03	1.39	0.53	1.34
360	0.95	0.84	0.68	0.19	0.23	0.93	0.75	0.88	0.42	0.87
600	0.52	0.47	0.40	0.15	0.17	0.51	0.46	0.50	0.30	0.50
1200	0.03	0.03	0.03	0.01	0.02	0.03	0.03	0.03	0.02	0.03
n					ρ	$= 0.85$				
240	2.99	2.45	1.84	0.28	0.39	2.82	1.29	2.16	0.69	2.04
360	1.72	1.45	1.13	0.21	0.28	1.66	1.07	1.47	0.55	1.42
600	0.94	0.82	0.67	0.16	0.21	0.92	0.74	0.87	0.41	0.86
1200	0.06	0.05	0.05	0.02	0.02	0.06	0.05	0.06	0.04	0.06
n					ρ	$= 0.95$				
240	10.27	7.85	5.81	0.59	0.91	8.47	1.16	3.46	1.11	3.04
360	3.74	2.92	2.20	0.25	0.38	3.25	0.90	1.94	0.60	1.78
600	3.23	2.57	1.97	0.26	0.38	2.94	1.33	2.26	0.74	2.14
1200	0.04	0.04	0.03	0.00	0.01	0.04	0.03	0.04	0.01	0.04
n					ρ	$= 0.99$				
240	57.06	41.74	31.79	2.79	4.23	28.05	0.22	1.77	1.98	1.33
360	33.49	24.61	18.94	1.77	2.66	18.76	0.44	2.90	1.81	2.31
600	18.15	13.41	10.38	1.00	1.52	11.60	0.90	3.77	1.50	3.21
1200	2.82	2.12	1.65	0.17	0.26	2.11	0.47	1.16	0.38	1.05

Table 12. RMLE and RRMLE with $p = 12$.

	RMLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					ρ	$= 0.75$				
240	2.38	2.20	1.96	1.29	1.33	2.34	1.95	2.22	1.53	2.19
360	1.88	1.79	1.65	1.24	1.27	1.86	1.73	1.83	1.44	1.82
600	1.55	1.51	1.45	1.21	1.23	1.54	1.50	1.54	1.35	1.53
1200	0.17	0.17	0.17	0.15	0.15	0.17	0.17	0.17	0.16	0.17
n					ρ	$= 0.85$				
240	3.36	2.98	2.52	1.31	1.40	3.26	2.17	2.83	1.65	2.74
360	2.46	2.26	2.00	1.25	1.31	2.42	1.99	2.29	1.54	2.25
600	1.88	1.78	1.65	1.22	1.26	1.86	1.73	1.83	1.44	1.82
1200	0.19	0.19	0.18	0.15	0.16	0.19	0.19	0.19	0.17	0.19
n										
240	8.32	6.83	5.39	1.57	1.82	7.38	2.04	3.93	1.99	3.60
360	3.43	2.90	2.35	0.89	0.99	3.17	1.43	2.29	1.17	2.16
600	3.52	3.08	2.61	1.27	1.37	3.36	2.21	2.94	1.67	2.85
1200	0.06	0.06	0.05	0.03	0.03	0.06	0.05	0.06	0.04	0.06
n					ρ	$= 0.99$				
240	38.39	29.76	22.70	3.27	4.40	23.00	1.23	2.49	2.64	2.13
360	23.82	18.70	14.51	2.49	3.21	16.00	1.44	3.45	2.54	2.97
600	13.61	10.83	8.56	1.88	2.30	10.28	1.81	4.21	2.29	3.72
1200	2.40	1.97	1.60	0.51	0.58	2.04	0.76	1.36	0.68	1.26

and RRMLE for different number of independent variables and sample sizes are shown in Tables 2, 7 and 11, which again indicating that the RRMLE outperform the RMLE. The SRMLE and SRRMLEfor different number of independent variables and sample sizes are presented are shown in Tables 3, 8 and 12 for $p = 4$, 8 and 12, respectively. If we review Tables 2, 3, 6, 7, 10 and 11, we can see that the SRML estimator and its ridge version uniformly dominate corresponding MLE, RMLE estimator and its ridge versions. The results of new proposed estimator in (1.9) is displayed in Tables 4, 5, 9, 10, 14 and 15 for $p = 4$, 8 and 12, respectively. This is a two parameter shrinkage estimator and the results shows

Table 13. SRMLE and SRRMLE with $p = 12$.

	SRMLE	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					$\rho = 0.75$					
240	1.39	1.20	0.95	0.24	0.29	1.35	0.93	1.22	0.49	1.19
360	0.85	0.76	0.62	0.19	0.22	0.83	0.68	0.79	0.39	0.78
600	0.48	0.44	0.38	0.15	0.17	0.47	0.43	0.47	0.28	0.46
1200	0.03	0.03	0.03	0.01	0.02	0.03	0.03	0.03	0.02	0.03
n					$\rho = 0.85$					
240	2.45	2.06	1.58	0.28	0.37	2.34	1.18	1.88	0.64	1.79
360	1.47	1.26	1.00	0.21	0.27	1.42	0.98	1.29	0.52	1.25
600	0.83	0.74	0.61	0.16	0.20	0.81	0.68	0.78	0.39	0.77
1200	0.05	0.05	0.04	0.02	0.02	0.05	0.05	0.05	0.03	0.05
n					$\rho = 0.95$					
240	7.68	6.15	4.64	0.56	0.84	6.70	1.09	3.09	1.03	2.74
360	2.92	2.37	1.81	0.24	0.36	2.64	0.84	1.74	0.56	1.60
600	2.60	2.15	1.68	0.25	0.37	2.43	1.25	2.00	0.69	1.91
1200	0.04	0.03	0.02	0.00	0.01	0.03	0.02	0.03	0.01	0.03
n					$\rho = 0.99$					
240	38.39	29.67	22.50	2.50	3.70	22.77	0.26	1.70	1.85	1.31
360	23.56	18.37	14.11	1.61	2.38	15.60	0.46	2.67	1.68	2.15
600	13.08	10.27	7.94	0.93	1.39	9.68	0.87	3.44	1.39	2.93
1200	2.14	1.70	1.32	0.17	0.25	1.77	0.45	1.07	0.35	0.97

Table 14. SRTPMLE with $p = 12$.

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
n					$\rho = 0.75$				
240	0.19	0.66	0.27	0.23	1.26	0.65	1.04	0.24	0.99
360	0.18	0.46	0.23	0.19	0.80	0.56	0.73	0.22	0.71
600	0.19	0.31	0.17	0.14	0.47	0.39	0.45	0.19	0.44
1200	0.02	0.02	0.01	0.01	0.03	0.03	0.03	0.02	0.03
n					$\rho = 0.85$				
240	0.18	1.02	0.22	0.20	1.94	0.65	1.34	0.24	1.24
360	0.16	0.69	0.18	0.16	1.27	0.67	1.06	0.23	1.01
600	0.19	0.45	0.13	0.13	0.77	0.55	0.71	0.22	0.69
1200	0.02	0.04	0.01	0.01	0.05	0.05	0.05	0.02	0.05
n					$\rho = 0.95$				
240	0.17	2.83	0.15	0.22	4.46	0.30	1.49	0.23	1.25
360	0.07	1.11	0.08	0.10	1.78	0.34	1.00	0.16	0.90
600	0.13	1.08	0.09	0.12	1.78	0.68	1.31	0.25	1.23
1200	0.01	0.02	0.00	0.00	0.03	0.02	0.03	0.01	0.02
n									
240	0.26	14.38	0.40	0.77	14.20	0.06	0.22	0.18	0.14
360	0.12	9.02	0.26	0.51	9.94	0.04	0.60	0.22	0.41
600	0.06	5.03	0.16	0.31	6.06	0.12	1.35	0.25	1.03
1200	0.03	0.83	0.03	0.06	1.10	0.15	0.59	0.09	0.52

Table 15. SRLMLE and SRTPMLE with $p = 12$.

	SRLMLE	k_{opt1}	k_{opt2}	k_{opt3}
n		ρ	$= 0.75$	
80	1.56	0.74	0.23	0.94
120	0.82	0.70	0.19	0.92
200	0.49	0.63	0.14	0.89
400	0.03	0.07	0.01	0.11
n				
80	1.74	0.68	0.20	0.91
120	1.37	0.65	0.16	0.90
200	0.90	0.60	0.13	0.87
400	0.05	0.07	0.01	0.11
n				
80	6.51	0.50	0.36	0.80
120	2.30	0.31	0.13	0.51
200	1.95	0.48	0.12	0.80
400	0.04	0.01	0.00	0.02
n		ρ	$= 0.99$	
80	31.50	1.12	14.72	0.52
120	19.07	0.41	3.81	0.53
200	10.95	0.30	0.86	0.56
400	1.99	0.09	0.09	0.19

a further improvement over MLE, RMLE and SRMLE and the ridge versions of these estimators. If we review Tables 13 and 14, we can see the remarkable improvement of two parameter stochastic restricted estimator over the SRML estimator. If we review all Tables (Tables 1–12), we can see that the estimators based on k_1, k_3, k_6 and k_8 are performing better compared to the rest.

5. Application

To illustrate the findings of the paper, we use a dataset from Cameron and Trivedi [6], where a recreation demand function is estimated using a logit model. The dataset is based on a survey and the dependent variable takes on the value one if the individual has taken a boat trip on Lake Somerville otherwise zero. We estimate this demand function including the cost of taking a boat trip on this particular lake and also the cost of some competing or substitute boating attractions at Lakes Conroe and Lake Houston. We focus on the high-income groups defined in the dataset as groups 7, 8 and 9 which leaves us with 59 observations. The correlation matrix of the independent variables is presented in Table 16. From Table 16, we can see that all correlation coefficients are above 0.9 which is expected since the cost of competitors are not independent. Thus this data can be used to illustrate the results of the paper. Due to the potential substitution effects, we use the following

Table 16. Correlation matrix.

	Lake Conroe	*Lake Somerville*	*Lake Houston*
Lake Conroe	1		
Lake Somerville	0.965	1	
Lake Houston	0.989	0.945	1

Table 17. Estimated logistic regression coefficients and bootstrap standard error.

	MLE	RMLE	SRMLE	RiMLE	SRLMLE	RRMLE	SRRMLE	SRTPMLE	SRLMLE
Lake Conroe	0.098(0.600)	0.075(0.776)	0.095(0.571)	0.126(0.376)	0.116(0.381)	0.197(0.548)	0.132(0.360)	0.109(0.250)	0.119(0.301)
Lake Somerville	−0.491(0.360)	−0.474(0.365)	−0.489(0.299)	−0.358(0.294)	−0.388(0.304)	−0.455(0.364)	−0.366(0.260)	−0.246(0.237)	−0.321(0.256)
Lake Houston	0.498(0.529)	0.506(0.608)	0.500(0.530)	0.334(0.382)	0.294(0.365)	0.346(0.437)	0.335(0.390)	0.228(0.310)	0.268(0.361)
k	NA	NA	NA	2.05	NA	2.05	2.05	2.05	1.59
d	NA	NA	NA	NA	0.00	NA	NA	0.00	0.00

Notes: We show the estimated parameters and the bootstrapped standard errors in parenthesis. The bootstrapping is based on 2000 replicates.

restriction:

$$H = \begin{bmatrix} 0 & 0.5 & 1 & 0.5 \end{bmatrix} \text{ and r} = [0] \tag{5.1}$$

Hence, we assume that a substitution effect exists between all three lakes. Furthermore, we assume that the stochastic part follows a normal distribution with zero mean and variance equal to 0.5. Hence not a large dispersion and therefore it is assumed the restriction holds approximately. To decide the restriction and the stochastic deviation from the theoretical restriction the researcher should usually use theoretical arguments such as the substitution effect or some prior knowledge in the field investigated. The values that we have chosen for the restriction is based on imposing a substitution effect. For the variance of the stochastic part, no theoretical arguments exist so it is determined by a grid search between 0 and 1 where the value minimizing the MSE is chosen. The estimated regression coefficients and the corresponding (within parenthesis) bootstrap's standard error are provided in Table 17. We bootstrap different individuals of the data and then based on the bootstrapped coefficients we calculate the standard errors. We can see that the coefficients for Lakes Conroe and Lake Houston are positive as expected since an increase in the price will increase the probability that one takes a boat trip on the competitor Lake Somerville instead. Furthermore, the coefficient for Lake Somerville is negative which is also expected since an increased cost theoretically should lead to lower demand for a boat trip. When looking at the bootstrapped standard we may see that they decrease when using a shrinkage estimator. In this particular example, we focus only on the classical Hoerl and Kennard [9] estimator since it was shown to be the optimal choice for the new estimator. This is expected when looking at the analytical results and the simulation results in this paper. We can also see that the standard errors are smaller for the ridge regression estimators and the lowest for the new estimator, SRTPMLE. This is in line with the simulated results where a substantial reduction in the estimated MSE was found.

6. Summary and concluding remarks

In this paper, we introduced a new stochastic restricted biased estimator for the logistic regression model. We explored the conditions for the superiority of the proposed estimator over the UMLE, RMLE, SRMLE, URML, RRML and SRRMLE estimators in the MMSE sense. Furthermore, A simulation study and numerical example have been given to illustrate the performance of the proposed estimator comparing to other estimators in this paper. We observed that the ridge regression estimators based on k_1, k_3, k_6 and k_8 are performing better compared to the rest.

Acknowledgements

The authors are thankful to referees and Associate editor for their valuable comments/suggestions, which certainly improved the quality and presentation of the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

B. M. Golam Kibria http://orcid.org/0000-0002-6073-1978

References

[1] A.M. Aguilera, M. Escabias, and M.J. Valderrama, *Using principal components for estimating logistic regression with high-dimensional multicollinear data*. Comput. Stat. Data. Anal. 50 (2006), pp. 1905–1924.
[2] M. Alkhamisi, G. Khalaf, and G. Shukur, *Some modifications for choosing ridge parameters*. Commun. Stat. Theor. M. 35 (2006), pp. 2005–2020.
[3] Y. Asar, *Some New methods to solve multicollinearity in logistic regression*. Commun. Stat. Simul. Comput. 46 (2017), pp. 2576–2586.
[4] Y. Asar, M. Arashi, and J. Wu, *Restricted ridge estimator in the logistic regression model*. Commun. Stat. Simul. Comput. 46 (2017), pp. 6538–6544.
[5] Y. Asar, and A. Genc, *New shrinkage parameters for the Liu- type logistic estimators*. Commun. Stat. Simul. Comput. 45 (2016), pp. 1094–1103.
[6] A.C. Cameron, and P.K. Trivedi, *Regression Analysis of Count Data*, 2nd ed., Econometric Society Monograph No.53, Cambridge University Press, Cambridge, 2013.
[7] D.E. Duffy, and T.J. Santner, *On the small sample prosperities of norm-restricted maximum likelihood estimators for logistic regression models*. Commun. Stat. Theory Methods 18 (1989), pp. 959–980.
[8] A. Goktas, and V. Sevinc, *Two new ridge parameters and a guide for selecting an appropriate ridge parameter in linear regression*. Gaz. Univ. J. Sci. 29 (2016), pp. 201–211.
[9] A.E. Hoerl, and R.W. Kennard, *Ridge regression: biased estimation for non-orthogonal problems*. Technometrics. 12 (1970), pp. 55–67.
[10] M.H. Hubert, and P. Wijekoon, *Improvement of the Liu estimator in linear regression model*. Stat. Pap. 47 (2006), pp. 471–479.
[11] D. Inan, and B.E. Erdogan, *Liu-type logistic estimator*. Commun. Stat. Simul. Comput. 42 (2013), pp. 1578–1586.
[12] G. Khalaf, and G. Shukur, *Choosing ridge parameter for regression Problems*. Commun. Stat. Theor. M. 34 (2005), pp. 1177–1182.
[13] B.M.G. Kibria, *Performance of some new ridge regression estimators*. Commun. Stat. Simul. Comput. 39 (2003), pp. 1162–1173.
[14] B.M.G. Kibria, and S. Banik, *Some ridge regression estimators and their performance*. J. Mod. Appl. Stat. Methods. 15 (2016), pp. 206–238.
[15] K. Månsson, B.M.G. Kibria, and G. Shukur, *On Liu estimators for the logit regression model*. Econ. Model 29 (2012), pp. 1483–1488.
[16] K. Månsson, B.M.G. Kibria, and G. Shukur, *A restricted Liu estimator for binary regression models and its application to an applied demand system*. J. Appl. Stat. 43 (2016), pp. 1119–1127.
[17] G. Muniz, and B.M.G. Kibria, *On some ridge regression estimators: An Empirical Comparisons*. Commun. Stat. Simul. Comput. 38 (2009), pp. 621–630.
[18] M.E. Nja, U.P. Ogoke, and E.C. Nduka, *The logistic regression model with a modified weight function*. J. Stat. Econom. Method 2 (2013), pp. 161–171.
[19] C.R. Rao, and H. Toutenburg, *Linear Models: Least Squares and Alternatives*, Spriger-Verlag, New York, 1995.
[20] R.L. Schaeffer, L.D. Roi, and R.A. Wolfe, *A ridge logistic estimator*. Commun. Stat. Theory Methods 13 (1984), pp. 99–113.
[21] G. Trenkler, and H. Toutenburg, *Mean square error matrix comparisons between biased estimators*. Stat. Pap. 31 (1990), pp. 165–179.
[22] N. Varathan, and P. Wijekoon, *Stochastic restricted maximum likelihood estimator in logistic regression model*. Open. J. Stat. 5 (2015), pp. 837–851.
[23] N. Varathan, and P. Wijekoon, *Ridge estimator in logistic regression under stochastic linear restriction*. Br. J. Math. Comput. Sci. 15 (2016), pp. 1–14.

[24] N. Varathan, and P. Wijekoon, *Logistic Liu estimator under stochastic linear restrictions*. Stat. Pap. 60 (2016), pp. 595–612.
[25] N. Varathan, and P. Wijekoon, *Liu-type logistic estimatorstochastic linear restrictions*. Ceylon J. Sci. 47 (2019), pp. 21–34.
[26] J. Wu, *Modified restricted Liu estimator in logistic regression model*. Comput. Stat. 31 (2016), pp. 1557–1567.