A Comparison of Pretest, Stein-Type and Penalty Estimators in Logistic Regression Model

Orawan Reangsephet[1(✉)], Supranee Lisawadi[1], and Syed Ejaz Ahmed[2]

[1] Department of Mathematics and Statistics, Thammasat University, Bangkok, Thailand
num.stat@gmail.com

[2] Faculty of Mathematics and Science, Brock University, St. Catharines, ON, Canada

Abstract. Various estimators are proposed based on the preliminary test and Stein-type strategies to estimate the parameters in a logistic regression model when it is priori suspected that some parameters may be restricted to a subspace. Two different penalty estimators as LASSO and ridge regression are also considered. A Monte Carlo simulation experiment was conducted for different combinations, and the performance of each estimator was evaluated in terms of simulated relative efficiency. The positive-part Stein-type shrinkage estimator is recommended for use since its performance is robust regardless of the reliability of the subspace information. The proposed estimators are applied to a real dataset to appraise their performance.

Keywords: Monte Carlo simulation · Logistic regression model · Likelihood ratio test · Preliminary test estimator · Shrinkage estimator · Penalty estimator

1 Introduction

For the past few decades, simultaneous variable selection and estimation of submodel parameters has become popular. Many predictors exist to infer an interesting response in the initial model. Some of these predictors may be inactive and not influential; these should be excluded from the final model that represents a sparsity pattern in the predictor space to achieve parsimony, flexibility and reliability. Several researchers, following this information in statistical modeling, have used either the full model or a candidate submodel.

The logistic regression model also called the logit model, is the most widely used for an analysis of the independent binary response data in medical, engineering, and other studies. This model assumes that the logit of the response variable can be modelled by a linear combination of unknown parameters $x_i'\beta$

J. Xu et al. (eds.), *Proceedings of the Eleventh International Conference on Management Science and Engineering Management*, Lecture Notes on Multidisciplinary Industrial Engineering, DOI 10.1007/978-3-319-59280-0_2

where $\mathbf{x_i}' = (x_{i1}, x_{i2}, \cdots, x_{ip})$ is a $p \times 1$ vector of the p predictors for the i^{th} subject and $\beta = (\beta_1, \beta_2, \cdots, \beta_p)'$ is a $p \times 1$ vector of regression parameters. Detailed information on logistic regression can be found in the books by Hilbe [8] and Hosmer and Lameshow [10].

In this article, we consider the problem of estimating the logistic regression model when the response variable may be related to many predictors, some of which may be inactive. Prior information about inactive predictors may be incorporated in the full model to produce the candidate submodel.

The pretest (preliminary test) estimation strategy, is inspired by Bancroft, and the shrinkage estimation strategy, is inspired by Stein, efficiently combine both full model and submodel estimators in an optimal way to achieve an improved estimator. Numerous authors have discussed the pretest, shrinkage, and penalty estimation strategies in many fields including Ahmed and Amezziane [2], Ahmed and Yüzbaşı [4], Al-Momaniet et al. [5], Gao, Ahmed, and Feng [6], Hossain, Ahmed, and Doksum [12], and Yüzbaşı and Ahmed [16,17]. For a logistic regression model, shrinkage estimators and three penalty estimators as LASSO, adaptive LASSO and SCAD were considered by Hossain and Ahmed [11] and Lisawadi, Shah, and Ahmed [13] considered the pretest estimation.

As we know, ridge regression (Hoerl and Kennard [9]) has been widely used when there are many possible predictors to achieve the precision of an estimate. Ahmed et al. [3] found that the ridge regression is highly efficient and stable when there are many predictors with small effect. Hence, we suggest the ridge regression for a logistic regression model. In this article, we propose the pretest and shrinkage estimators in the logistic regression model when it is priori suspected that parameters may be restricted to a subspace and compares the resulting estimators to the classical maximum likelihood estimator as well as the penalty estimators, i.e. LASSO estimator and ridge regression. Monte Carlo simulation study is carried out using the simulated relative efficient to appraise the performance of the proposed estimators.

To further illustrate the proposed estimators in the logistic regression model, we apply the proposed estimator to the South African heart disease data set and provide a bootstrap approach to compute simulated relative efficiency (SRE) and simulated relative prediction error (SPE) of the estimators. The detail of this data set will be described in the Sect. 4. Hossain, Ahmed, and Doksum [12] also considered this data in the generalized linear model via the pretest estimator, positive-part Stein-type shrinkage estimator, and three penalty estimators as LASSO, adaptive LASSO, and SCAD The performance of these estimators are evaluated in terms of simulated relative efficient (SRE).

Under the prior information about inactive predictors, the full parameter vector β can be partitioned as $\beta = (\beta_1', \beta_2')'$ where β_1 and β_2 represent a $p_1 \times 1$ active parameter and a $p_2 \times 1$ inactive parameter subvector, respectively, such that $p = p_1 + p_2$. Therefore, our interest lies in the estimation of the active parameter subvector β_1 when the information on β_2 is readily available. In other words, this information about the inactive parameters may be used to estimate the active parameter subvector β_1 when their values are near to some specified value β_2^0. Without the loss of generality, it is plausible that β_2 may be set to a zero vector, $\beta_2 = 0$. Keep in mind that the candidate submodel estimator is more efficient than the full model estimator when the candidate submodel is correct. On the other hand, the submodel estimator may not be reliable and become considerably inefficient when the candidate submodel incorrectly represents the data at hand.

The remainder of this article is organized as follows; the model and the efficient estimation strategies are proposed in Sect. 2, the results of a Monte Carlo simulation study are reported in Sect. 3, real data applications are described in Sect. 4, and finally, discussions and conclusions are presented in Sect. 5.

2 Model and Estimation Strategies

Let $y_1, y_2, \cdots, y_n$ be independent binary response variables which contain only two possible outcomes, and $\mathbf{x_i} = (x_{i1}, x_{i2}, \cdots, x_{ip})'$ is a $p \times 1$ predictors vector for the i^{th} subject and $i = 1, 2, \cdots, n$. The simplest idea would be to let z_i be a linear function of the predictors, suppose

$$z_i = \mathbf{x_i'}\beta, \tag{1}$$

where β is a $p \times 1$ vector of regression coefficients. Thus, the logistic regression model assume that

$$P\left(y_i = 1|\mathbf{x}_i\right) = \pi\left(z_i\right) = \frac{\exp\left(z_i\right)}{1 + \exp\left(z_i\right)} = \frac{\exp\left(\mathbf{x}_i^{'}\beta\right)}{1 + \exp\left(\mathbf{x}_i^{'}\beta\right)}. \tag{2}$$

The log-likelihood function of the logistic regression model is given by

$$l\left(\beta\right) = \sum_{i=1}^{n} y_i \ln\left\{\pi(\mathbf{x}_i^{'}\beta)\right\} + \sum_{i=1}^{n}(1 - y_i) \ln\left\{1 - \pi\left(\mathbf{x}_i^{'}\beta\right)\right\}. \tag{3}$$

The derivative of the log-likelihood function with respect to β is obtained by solving the score equation:

$$\frac{\partial l\left(\beta\right)}{\partial\beta} = \sum_{i=1}^{n}\left[y_i - \pi(\mathbf{x}_i^{'}\beta)\right]\mathbf{x}_i = 0. \tag{4}$$

2.1 The Unrestricted and Restricted Maximum Likelihood Estimator

The unrestricted maximum likelihood estimator (UE) of the parameter vector β denoted by $\hat{\beta}^{UE}$ is obtained by solving the non-linear score Eq. 4, and this can be solved by using an iterative method like Newton-Raphson.

Under the certain regularity conditions of maximum likelihood estimator (MLE), Gourieroux and Monfort [7] showed that $\hat{\beta}^{UE}$ is a consistent estimator of β and asymptotically normally distributed with a variance-covariance matrix $(I(\beta))^{-1}$, where $I(\beta)$ is the information matrix which is defined as

$$I\left(\beta\right)=\sum_{i=1}^{n}\pi(\mathbf{x}_{i}^{'}\beta)\left\{1-\pi(\mathbf{x}_{i}^{'}\beta)\right\}\mathbf{x}_{i}\mathbf{x}_{i}^{'}. \tag{5}$$

The restricted maximum likelihood estimator (RE) of β denoted by $\hat{\beta}^{RE}$ can be obtained by maximizing the log-likelihood function (3) under the subspace restriction $\beta_2-\beta_2^0=0$.

2.2 The Linear Shrinkage Estimator

The linear shrinkage estimator (LS) of β denoted by $\hat{\beta}^{LS}$ is a linear combination of the unrestricted and restricted estimator, that is

$$\hat{\beta}^{LS}=\lambda\hat{\beta}^{RE}+\left(1-\lambda\right)\hat{\beta}^{UE},\lambda\in\left[0,1\right], \tag{6}$$

where λ defines the degree of confidence in the given prior information and is a fixed constant. The linear shrinkage estimator shrinks $\hat{\beta}^{UE}$ toward $\hat{\beta}^{RE}$. If $\lambda=0$, then LS simplifies to an unrestricted estimator, while it simplifies to a restricted estimator when $\lambda=1$. The performance of the linear shrinkage estimator is better than the unrestricted and restricted MLE in some part of the parameter space.

2.3 The Preliminary Test Estimator

The preliminary test estimator or pretest estimator (PT) of β denoted by $\hat{\beta}^{PT}$ is defined as

$$\hat{\beta}^{PT}=\hat{\beta}^{UE}-\left(\hat{\beta}^{UE}-\hat{\beta}^{RE}\right)I\left(\mathscr{L}_n\leq\mathscr{L}_{n,\alpha}\right), \tag{7}$$

where $I(.)$ is an indicator function, and $\mathscr{L}_{n,\alpha}$ is the α-level critical value of the exact distribution of a suitable test statistic $\mathscr{L}_n$ under $H_0:\beta_2=\beta_2^0$. For testing $H_0:\beta_2=\beta_2^0$, the likelihood ratio statistic $\mathscr{L}_n$ is suggested:

$$\mathscr{L}_n=-2\log\left(\frac{L\left(\hat{\beta}^{RE}\right)}{L\left(\hat{\beta}^{UE}\right)}\right)=2\left(l\left(\hat{\beta}^{UE}\right)-l\left(\hat{\beta}^{RE}\right)\right), \tag{8}$$

where $l(\hat{\beta}^{UE})$ and $l(\hat{\beta}^{RE})$ are values of the log-likelihood at the unrestricted and restricted estimates, respectively. Under H_0, the distribution of $\mathscr{L}_n$ converges to Chi-square distribution with p_2 degree of freedom as $n\to\infty$.

Clearly, the pretest estimator takes the value of the unrestricted estimator when the test statistic lies in a rejection region, otherwise, it takes the value of the restricted estimator. This estimator has limits due to the large size of the pretest.

2.4 The Shrinkage Pretest Estimator

The shrinkage pretest estimator (SP) of β denoted by $\hat{\beta}^{SP}$ is defined by replacing the restricted estimator with the linear shrinkage estimator in Eq. (7), that is

$$\hat{\beta}^{SP} = \hat{\beta}^{UE} - \left(\hat{\beta}^{UE} - \hat{\beta}^{LS}\right) I\left(\mathscr{L}_n \leq \mathscr{L}_{n,\alpha}\right). \tag{9}$$

An alternative form of the estimator is

$$\hat{\beta}^{SP} = \hat{\beta}^{UE} - \lambda\left(\hat{\beta}^{UE} - \hat{\beta}^{RE}\right) I\left(\mathscr{L}_n \leq \mathscr{L}_{n,\alpha}\right). \tag{10}$$

Ahmed [1] found that the shrinkage pretest estimator significantly improves upon the pretest estimator in terms of size α, and it dominates the unrestricted estimator in a large portion of the parameter space. For $\lambda = 1$, the pretest estimators are used to estimate the parameter, while we use a UE as $\lambda = 0$. Generally, the estimators based on the pretest strategy are biased and inefficient when the null hypothesis does not hold.

2.5 The Stein-Type Shrinkage Estimator

The Stein-type shrinkage estimator which combines the unrestricted and the restricted estimator in an optimal way to dominate the unrestricted estimator. The Stein-type shrinkage estimator (S) of β denoted by $\hat{\beta}^{S}$ is given as follows

$$\hat{\beta}^{S} = \hat{\beta}^{RE} + \left(1 - (p_2 - 2)\,\mathscr{L}_n{}^{-1}\right)\left(\hat{\beta}^{UE} - \hat{\beta}^{RE}\right), p_2 \geq 3, \tag{11}$$

alternatively,

$$\hat{\beta}^{S} = \hat{\beta}^{UE} - (p_2 - 2)\,\mathscr{L}_n{}^{-1}\left(\hat{\beta}^{UE} - \hat{\beta}^{RE}\right), p_2 \geq 3. \tag{12}$$

For some insight to this estimator, we refer to Hossain, Ahmed, and Doksum [12], Yüzbaşı and Ahmed [17] among others. The Stein-Type shrinkage estimator will provide uniform improvement over the unrestricted estimator. However, the Stein-type shrinkage estimator tends to over-shrink the unrestricted estimator towards the restricted estimator when the test statistic $\mathscr{L}_n$ is very small in comparison with $p_2 - 2$. To avoid the over-shrink behavior of this estimator, the truncated version is suggested which is called the positive-part Stein-type shrinkage estimator.

2.6 The Positive-Part Stein-Type Shrinkage Estimator

The positive-part Stein-type shrinkage estimator (S^{+}) of β denoted by $\hat{\beta}^{S^{+}}$ is a convex combination of the unrestricted and restricted estimator, that is

$$\hat{\beta}^{S^{+}} = \hat{\beta}^{RE} + \left(1 - (p_2 - 2)\,\mathscr{L}_n^{-1}\right)^{+}\left(\hat{\beta}^{UE} - \hat{\beta}^{RE}\right), p_2 \geq 3, \tag{13}$$

where $a^+ = \max\{0, a\}$. Alternatively, it can be written in the following conical form as

$$\hat{\beta}^{S^+} = \hat{\beta}^{RE} + \left(1 - (p_2 - 2)\,\mathscr{L}_n{}^{-1}\right)\left(\hat{\beta}^{UE} - \hat{\beta}^{RE}\right) I\left(\mathscr{L}_n < (p_2 - 2)\right), p_2 \geq 3. \quad (14)$$

The positive-part Stein-type shrinkage estimator is particularly important to control the over-shrinking inherent in $\hat{\beta}^S$.

2.7 The LASSO Estimator

Tibshirani [15] introduced the LASSO estimator of β which minimizes the negative log-likelihood in Eq. (3) under the L_1 constraint. It can be defined as

$$\hat{\beta}^{\text{LASSO}} = \text{argmin}_{\beta}\left\{-l\left(\beta\right) + \gamma\sum_{i=1}^{p}|\beta_i|\right\}, \gamma \geq 0, \quad (15)$$

where λ is the tuning parameter which controls the amount of a shrinkage. The LASSO shrinks some coefficients to exactly zero. Therefore, LASSO procedure performs variable selection and parameter estimation simultaneously.

2.8 The Ridge Regression Estimator

Hoerl and Kennard [9] proposed the ridge regression estimator of β which minimizes the negative log-likelihood in Eq. (3) under the L_2 constraint. It can be defined as

$$\hat{\beta}^{RIDGE} = \arg\min_{\beta}\left\{-l\left(\beta\right) + \gamma\sum_{i=1}^{p}{\beta_i}^2\right\}, \gamma \geq 0, \quad (16)$$

where λ is a tuning parameter which controls the amount of shrinkage. The ridge regression estimator always keeps all the predictors in the model; thus, this estimator cannot produce a parsimonious model.

3 Monte-Carlo Simulation Studies

In this section, we carry out a Monte Carlo simulation to compare the performance of the pretest, Stein-type and penalty estimators in terms of the quadratic risk, namely mean squared error (MSE). Our simulations are based on a logistic regression model with the sample size $n = 250$. A binary response data is generated from the following model

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \mathbf{x_i}^{'}\beta = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \quad i = 1, 2, \cdots, n, \quad (17)$$

where $p_i = P(y_i = 1|\mathbf{x_i})$ and the predictor values $\mathbf{x_i}$ have been drawn from a standardized multivariate normal distribution.

We consider the hypothesis $H_0 : \beta_2 = 0$. We partition the parameter vector β as $\beta = (\beta_1', \beta_2')'$ and where β_1 and $\beta_{\mathbf{2}}$ represent a $p_1 \times 1$ and a $p_2 \times 1$ vector, respectively, such that $p = p_1 + p_2$. We set the true value of $\beta = (\beta_1', \beta_2')' = (\beta_1', 0)'$ with $\beta_1' = (1.90, -1.05, 0.25, -0.78)$.

The value of λ is set to 0.25, 0.50, and 0.75. The value of significance level α is set to 0.01, 0.05, 0.10 and a higher value 0.35.

We now define the parameter $\varDelta^*$ representing the distance between the simulated model and the candidate submodel estimator by

$$\varDelta^* = \left(\beta - \beta^{(0)}\right)' \left(\beta - \beta^{(0)}\right) = \sum_{i=1}^{p} \left(\beta_i - \beta_i^{(0)}\right)^2, \tag{18}$$

where $\beta^{(0)} = \left(\beta_1^{'}, 0\right)^{'}$ and β is the true parameter in the simulated model. Samples were generated using $\varDelta^*$ between 0 and 4.

The number of replications in the simulation was initially varied and it was determined that $N = 1,000$ iterations were adequate to obtain a stable result for each combination of parameters.

Based on the simulated data, we estimated the MSE of all the proposed estimators. The performance of the estimators was evaluated using the notion of simulated relative efficient (SRE), which is the MSE relative to the MSE of $\hat{\beta}^{UE}$. For any estimator $\hat{\beta}^*$, the SRE of $\hat{\beta}^*$ with respect to $\hat{\beta}^{UE}$ is defined as

$$SRE\left(\hat{\beta}^{UE}, \hat{\beta}^*\right) = \frac{\text{SimulatedMSE}(\hat{\beta}^{UE})}{\text{SimulatedMSE}(\hat{\beta}^*)} = \frac{\text{Simulated} \sum_{i=1}^{p} \left(\beta_i - \beta_i^{UE}\right)^2}{\text{Simulated} \sum_{i=1}^{p} \left(\beta_i - \beta_i^*\right)^2}. \tag{19}$$

Keep in mind that an SRE is larger than the one that indicates the degree of superior of the estimator $\hat{\beta}^*$ over $\hat{\beta}^{UE}$.

3.1 Model with Correct Candidate Submodel ($\boldsymbol{\varDelta^* = 0}$)

First, the case when the candidate submodel is assumed to be correct, is $\varDelta^* = 0$. Various choices of active and inactive predictors are provided for $(p_1, p_2) = (4, 3), (4, 5), (4, 7), (4, 10)$, and $(4, 15)$, and the SRE results are reported in Tables 1, 2 and 3. The tuning parameter γ of the two penalty estimators is estimated using 10 fold-cross validation. The findings from Tables 1, 2 and 3 are summarized as follows:

We note that SREs of all the estimators increase as the number of inactive predictors $_2$ is increased for fixed λ and α. Interestingly, the restricted estimator is the best, and all estimators are superior to the unrestricted estimator for all configurations except the ridge regression estimator. The linear shrinkage estimator depends on the choice of λ. Its SRE decreases sharply to 1 as $\lambda \to 0$ and approaches to SRE of the restricted estimator for higher value of λ. The SREs of the pretest estimators depend on the size of the test α. For small α,

Table 1. The SREs of the estimators with respect to the UE for $\lambda = 0.25$ at $\Delta^* = 0$.

Estimator		Number of in active (p_2)				
		3	5	7	10	15
RE		1.693	2.104	2.735	3.544	6.120
LS		1.204	1.280	1.356	1.424	1.543
PT	$\alpha = 0.01$	1.631	2.002	2.542	3.162	4.558
	$\alpha = 0.05$	1.472	1.759	2.147	2.513	3.037
	$\alpha = 0.10$	1.369	1.606	1.804	2.108	2.398
	$\alpha = 0.35$	1.128	1.238	1.312	1.368	1.441
SP	$\alpha = 0.01$	1.190	1.263	1.335	1.396	1.489
	$\alpha = 0.05$	1.152	1.220	1.283	1.333	1.393
	$\alpha = 0.10$	1.125	1.187	1.227	1.278	1.325
	$\alpha = 0.35$	1.049	1.088	1.109	1.125	1.147
S		1.158	1.467	1.818	2.300	3.363
S^+		1.204	1.549	1.924	2.428	3.604
LASSO		1.226	1.293	1.470	1.819	2.383
Ridge		0.561	0.668	0.867	1.029	1.332

Table 2. The SREs of the estimators with respect to the UE for $\lambda = 0.50$ at $\Delta^* = 0$

Estimator		Number of in active (p_2)				
		3	5	7	10	15
RE		1.693	2.104	2.735	3.544	6.120
LS		1.416	1.610	1.839	2.069	2.556
PT	$\alpha = 0.01$	1.631	2.002	2.542	3.162	4.558
	$\alpha = 0.05$	1.472	1.759	2.147	2.513	3.037
	$\alpha = 0.10$	1.369	1.606	1.804	2.108	2.398
	$\alpha = 0.35$	1.128	1.238	1.312	1.368	1.441
SP	$\alpha = 0.01$	1.384	1.566	1.773	1.968	2.315
	$\alpha = 0.05$	1.298	1.454	1.622	1.765	1.954
	$\alpha = 0.10$	1.239	1.376	1.473	1.608	1.737
	$\alpha = 0.35$	1.089	1.162	1.206	1.240	1.286
S		1.158	1.467	1.818	2.300	3.363
S^+		1.204	1.549	1.924	2.428	3.604
LASSO		1.226	1.293	1.470	1.819	2.383
Ridge		0.561	0.668	0.867	1.029	1.332

Table 3. The SREs of the estimators with respect to the UE for $\lambda = 0.75$ at $\Delta^* = 0$

Estimator		Number of in active (p_2)				
		3	5	7	10	15
RE		1.693	2.104	2.735	3.544	6.120
LS		1.597	1.927	2.381	2.913	4.354
PT	$\alpha = 0.01$	1.631	2.002	2.542	3.162	4.558
	$\alpha = 0.05$	1.472	1.759	2.147	2.513	3.037
	$\alpha = 0.10$	1.369	1.606	1.804	2.108	2.398
	$\alpha = 0.35$	1.128	1.238	1.312	1.368	1.441
SP	$\alpha = 0.01$	1.546	1.848	2.245	2.669	3.556
	$\alpha = 0.05$	1.413	1.655	1.953	2.227	2.616
	$\alpha = 0.10$	1.326	1.53	1.689	1.925	2.159
	$\alpha = 0.35$	1.116	1.215	1.278	1.326	1.392
S		1.158	1.467	1.818	2.300	3.363
S^+		1.204	1.549	1.924	2.428	3.604
LASSO		1.226	1.293	1.470	1.819	2.383
Ridge		0.561	0.668	0.867	1.029	1.332

the pretest estimator is comparable to that of the restricted estimator. On the contrary, its SRE decreases as the size increases.

Moreover, the performance of the shrinkage pretest estimator depends on the choice of λ and α. Its performance becomes poorer with an increase in and a decrease in λ. Unsurprisingly, the pretest estimator outperforms the linear shrinkage estimator at $\Delta^* = 0$. As we would expect, the positive-part Stein-type shrinkage estimator outperforms the Stein-type shrinkage estimator in any situation.

Similar results are observed for the LASSO estimator which performs better than the estimator based on the Stein-type strategy for small p_2. On the other hand, the estimator based on the Stein-type strategy is preferable when there are many inactive predictors. However, the ridge regression estimator does not perform well, but its performance slowly improves as the number of inactive predictors increases.

3.2 Model with Correct and Incorrect Candidate Submodel ($\boldsymbol{\Delta^* \geq 0}$)

The penalty estimators are not included in the $\Delta^* > 0$ case because these estimators do not take advantage of the fact that the regression parameter lies in the subspace $\beta_2 = 0$. In this case, the simulation model has an active coefficient vector $\beta_1 = (0.5, -1.5, 0.2)$ and inactive coefficient vector $\beta_2 = (\beta_4, a)$ where β_4 is a scalar and assumes various values; thus, $\Delta^* = (\beta_4)^2$ when the candidate submodel is used. We choose $\beta_4 = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1.0, 1.5,$ and 2.0

and a is $k \times 1$ zero vector with different dimensions, $k = p_2 - 1$ to be the number of inactive predictors in the model and we use $p_2 = 5, 7, 10,$ and 15 in the simulations.

The choice of α was fixed to be 0.01, 0.05, 0.10, $\lambda = 0.75$ and $n = 250$. SREs of the proposed estimators are are reported in Table 4 are graphically represented in Figs. 1, 2 and 3. The findings are summarized as follows:

The restricted estimator $\hat{\beta}^{RE}$ outperforms all the estimators at and near $\Delta^* = 0$. In contrast, as Δ^* becomes larger than zero, the relative efficiency of $\hat{\beta}^{RE}$ as well as the linear shrinkage estimator decrease and become unbounded. However, the relative efficiency of all the other estimators remains bounded and approaches to 1. These show that an incorrect candidate submodel is fatal to the restricted and linear shrinkage estimators.

The pretest estimator is much better than the shrinkage pretest estimator when the candidate submodel is correct, i.e. $\Delta^* = 0$. On the contrary, the shrinkage pretest estimator does well relative to the pretest estimator in a small part of the parameter space. However, the SRE of both pretest and shrinkage pretest estimators approaches to one as Δ^* moves away from zero, but after becoming inferior to $\hat{\beta}^{UE}$, and later at some point, they join the SRE of one from below. In addition, they outshine the estimator based on Stein-type strategy where Δ^* is near zero and for small and moderate p_2. The estimators based on Stein-type strategy are superior to the unrestricted estimator in the entire range of Δ^* especially their gain in risk reduction is impressive as p_2 increases. Lastly, we found that the estimators based on Stein-type strategy perform better than all other estimators in the wider range of Δ^* and these estimators are little impacted by severe departure from the restriction.

4 Real Data Example: South African Heart Disease Data

In this section, we apply the proposed estimators to the South African heart disease data set. Rousseauw et al. [14] described a retrospective sample of males in a heart-disease high-risk region of the Western Cape, South Africa. This study comprised over 462 samples and the set of variables is described in Table 5.

We notice that the condition index (CI) value is calculated as 392.718 which implies the existence of multicollinearity in this data set. After applying the variable selection procedure based on AIC criterion, BIC criterion, and LASSO, the results are given in Table 6.

Table 6 shows that the candidate submodel based on AIC and BIC criteria contains 5 active predictors, while LASSO selection procedure contains 7 active predictors. Hence, we will consider the candidate submodel with 5 active predictors that is tobacco, famhist, ldl, typea, and age. The restricted subspace is $\beta_2' = (\beta_{\text{adiposity}}, \beta_{\text{obesity}}, \beta_{\text{alcohol}}, \beta_{\text{sbp}}) = (0, 0, 0, 0), p = 9, p_1 = 5$, and $p_2 = 4$.

To examine the performance of the proposed estimators for the candidate submodel, we draw $m = 250$ bootstrap rows with replacement $N = 1,000$ times from the data. The performance of the proposed estimators with respect to the

Table 4. The SREs of RE, LS, PT, SP, S, and $S+$ with respect to the UE for $\lambda = 0.75$

p_2	Δ^*	RE	LS	PT			SP			S	S+
				$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$		
5	0.0	2.581	2.311	2.428	2.057	1.806	2.195	1.909	1.705	1.597	1.726
	0.1	1.328	1.611	1.141	1.046	1.016	1.299	1.139	1.082	1.279	1.309
	0.2	0.884	1.227	0.836	0.882	0.907	0.976	0.957	0.960	1.187	1.195
	0.3	0.666	0.988	0.773	0.873	0.920	0.882	0.928	0.954	1.150	1.152
	0.4	0.54	0.833	0.795	0.910	0.958	0.879	0.947	0.975	1.126	1.127
	0.5	0.452	0.716	0.863	0.948	0.968	0.917	0.967	0.980	1.111	1.111
	0.8	0.309	0.515	0.958	0.988	0.996	0.973	0.992	0.998	1.088	1.088
	1.0	0.255	0.434	0.984	1.000	1.000	0.989	1.000	1.000	1.081	1.081
	1.5	0.184	0.322	1.000	1.000	1.000	1.000	1.000	1.000	1.074	1.074
	2.0	0.152	0.269	1.000	1.000	1.000	1.000	1.000	1.000	1.070	1.070
7	0.0	3.065	2.66	2.873	2.477	2.079	2.523	2.232	1.924	1.959	2.147
	0.1	1.701	1.962	1.406	1.215	1.147	1.550	1.286	1.197	1.499	1.540
	0.2	1.183	1.548	1.000	0.973	0.975	1.140	1.042	1.021	1.349	1.359
	0.3	0.909	1.29	0.876	0.927	0.955	0.985	0.978	0.984	1.282	1.284
	0.4	0.748	1.116	0.857	0.939	0.963	0.938	0.971	0.981	1.240	1.241
	0.5	0.643	0.993	0.891	0.962	0.979	0.944	0.979	0.989	1.217	1.217
	0.8	0.449	0.735	0.969	0.991	0.996	0.981	0.994	0.997	1.169	1.169
	1.0	0.376	0.632	0.989	0.997	1.000	0.993	0.998	1.000	1.155	1.155
	1.5	0.278	0.483	1.000	1.000	1.000	1.000	1.000	1.000	1.139	1.139
	2.0	0.222	0.394	1.000	1.000	1.000	1.000	1.000	1.000	1.129	1.129
10	0.0	4.693	3.702	3.987	3.087	2.396	3.277	2.680	2.175	2.612	2.942
	0.1	2.43	2.661	1.982	1.531	1.366	2.124	1.604	1.414	1.944	2.048
	0.2	1.669	2.124	1.301	1.135	1.074	1.499	1.230	1.135	1.686	1.730
	0.3	1.263	1.755	1.029	0.991	0.983	1.197	1.079	1.037	1.569	1.584
	0.4	1.024	1.510	0.911	0.931	0.949	1.048	0.997	0.99	1.485	1.49
	0.5	0.868	1.335	0.857	0.926	0.955	0.974	0.974	0.983	1.433	1.434
	0.8	0.610	1.001	0.890	0.959	0.980	0.942	0.977	0.989	1.326	1.327
	1.0	0.508	0.858	0.928	0.980	0.992	0.959	0.988	0.995	1.300	1.300
	1.5	0.368	0.648	0.996	1.000	1.000	0.998	1.000	1.000	1.262	1.262
	2.0	0.295	0.529	1.000	1.000	1.000	1.000	1.000	1.000	1.236	1.236
15	0.0	6.502	4.665	4.986	3.510	2.651	3.879	2.976	2.372	3.58	3.961
	0.1	3.570	3.566	2.568	1.735	1.462	2.570	1.742	1.469	2.569	2.662
	0.2	2.466	2.926	1.575	1.258	1.164	1.703	1.310	1.197	2.163	2.196
	0.3	1.906	2.506	1.219	1.083	1.046	1.330	1.132	1.075	1.949	1.956
	0.4	1.549	2.188	1.062	1.018	1.011	1.152	1.053	1.029	1.818	1.819
	0.5	1.323	1.951	0.993	0.993	0.992	1.057	1.017	1.005	1.712	1.713
	0.8	0.927	1.503	0.971	0.992	0.997	0.991	0.997	0.999	1.560	1.560
	1.0	0.793	1.331	0.984	0.995	0.998	0.993	0.998	0.999	1.500	1.500
	1.5	0.574	1.015	0.999	0.999	1.000	0.999	0.999	1.000	1.423	1.423
	2.0	0.458	0.827	1.000	1.000	1.000	1.000	1.000	1.000	1.380	1.380

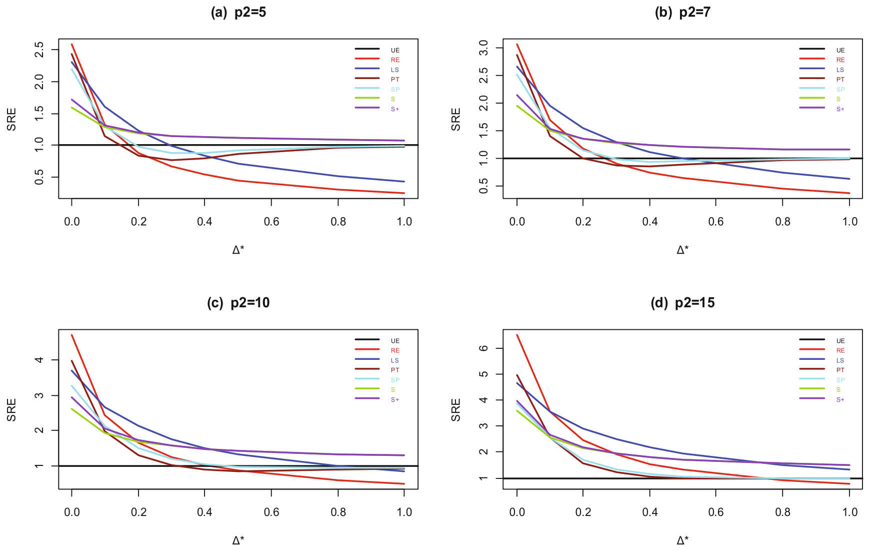

Fig. 1. SREs of RE, LS, PT, SP, S, and S^+ with respect to the UE when the candidate subspace misspecifies β_4 as zero as of $\Delta^* = (\beta_4)^2$. Here, $p_1 = 3$, $\alpha = 0.01$.

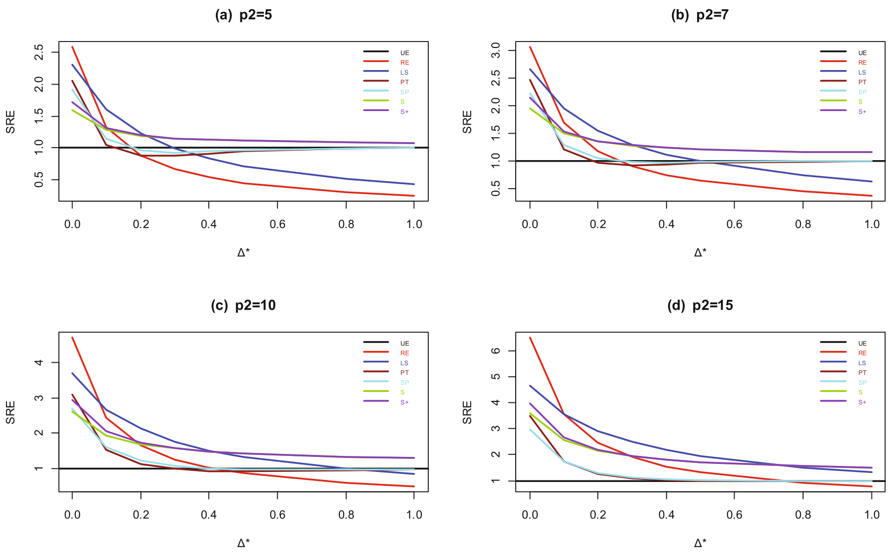

Fig. 2. SREs of RE, LS, PT, SP, S, and S+ with respect to the UE when the candidate subspace misspecifies β_4 as zero as of $\Delta^* = (\beta_4)^2$. Here, $p_1 = 3$, $\alpha = 0.05$.

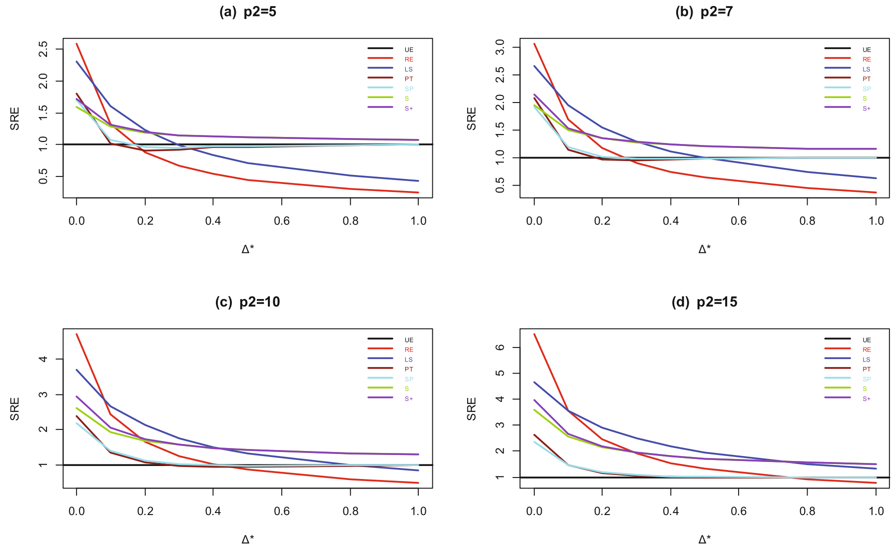

Fig. 3. SREs of RE, LS, PT, SP, S, and S+ with respect to the UE when the candidate subspace misspecifies β_4 as zero as of $\Delta^* = (\beta_4)^2$. Here, $p_1 = 3$, $\alpha = 0.10$.

Table 5. List of variables

Variable	Description
Response variable	
Chd	Coronary heart disease
Predictor variable	
Tobacco	Cumulative tobacco (kg)
Famhist	Family history of heart disease, a factor with levels absent and present
Ldl	Low density lipoprotein cholesterol
Typea	Type-A behavior
Age	Age at onset
Adiposity	Adiposity
Obesity	Obesity
Alcohol	Current alcohol consumption
Sbp	Systolic blood pressure

Table 6. Full and candidate sub-models for South African heart disease data

Selection criterion	Response	Active predictor
Full model	Chd	tobacco, famhist, ldl, typea, age, adiposity, obesity, alcohol, sbp
LASSO	Chd	tobacco, famhist, ldl, typea, age, obesity, sbp
AIC,BIC	Chd	tobacco, famhist, ldl, typea, age

unrestricted estimator is evaluated by simulated relative efficiency (SRE) and simulated relative prediction error (SPE) of the estimators which is defined as:

$$\text{SRE}\left(\hat{\beta}^{UE}, \hat{\beta}^{*}\right) = \frac{\text{Simulated } \sum_{i=i}^{p} \left(\beta_i^{\text{true}} - \hat{\beta}_i^{UE}\right)}{\text{Simulated } \sum_{i=1}^{p} \left(\beta_i^{\text{true}} - \hat{\beta}_i^{*}\right)}$$

$$\text{and SPE}\left(\hat{\beta}^{UE}, \hat{\beta}^{*}\right) = \frac{\text{Simulated } \sum_{i=1}^{m} \left(Y_i - \pi(\mathbf{x}'_i \hat{\beta}^{UE})\right)}{\text{Simulated } \sum_{i=1}^{m} \left(Y_i - \pi(\mathbf{x}'_i \hat{\beta}^{*})\right)}, i = 1, 2, \cdots, m.$$

Note that SPE is less than one; this means the unrestricted estimator is doing better. This study assumed the empirical distribution $\hat{F}$ based on 462 actual observations to be the true distribution and the resulting logistic regression coefficient $\hat{\beta}$'s to be the true parameter values. We assumed $\alpha = 0.01$ and $\lambda = 0.50$. The results of the point estimates, standard errors, SREs, and SPEs of the estimators are shown in Table 7.

Table 7 reveals that the restricted estimator is the best, and all the estimators outperform the unrestricted estimator. The performance of linear shrink-

Table 7. Estimates (first row) and standard errors (second row) of the coefficients for active predictors. The SRE and SPE columns give the relative efficiency and relative prediction error of the estimators with respect to UE, respectively.

Estimator	β_{tobacco}	β_{famhist}	β_{ldl}	β_{typea}	β_{age}	SRE	SPE
UE	0.084	0.948	0.186	0.042	0.048	1.000	1.000
	0.041	0.340	0.087	0.018	0.018		
RE	0.083	0.917	0.167	0.038	0.052	2.156	1.364
	0.039	0.324	0.077	0.017	0.015		
LS	0.085	0.933	0.176	0.040	0.050	1.653	1.360
	0.040	0.330	0.079	0.018	0.015		
PT	0.084	0.922	0.170	0.039	0.051	1.836	1.230
	0.039	0.329	0.079	0.018	0.016		
SP	0.085	0.935	0.178	0.040	0.050	1.507	1.274
	0.040	0.333	0.081	0.018	0.016		
S	0.084	0.938	0.180	0.041	0.049	1.309	1.156
	0.040	0.334	0.082	0.018	0.016		
S^{+}	0.084	0.938	0.180	0.041	0.049	1.361	1.191
	0.040	0.334	0.082	0.018	0.016		
LASSO	0.071	0.788	0.147	0.029	0.042	1.360	1.051
	0.038	0.353	0.083	0.020	0.014		
Ridge	0.072	0.761	0.146	0.028	0.035	1.552	1.210
	0.030	0.311	0.068	0.018	0.015		

age, pretest, and shrinkage pretest estimators is dominated by the restricted estimator as our data in the resampling scheme are generated from an empirical distribution where the candidate submodel is correct, that is $\Delta^* = 0$.

Furthermore, the positive-part Stein-type shrinkage estimator outshines the LASSO estimator when there are moderate or relatively large numbers of the inactive predictor in the model. In fact, the true parameter values are not exactly zero, and there is the multicollinearity problem. Unsurprisingly, the ridge regression estimator performs well. Lastly, the LASSO estimator shows good performance in terms of SRE but not in terms of SPE because of the instability estimation.

5 Discussion and Conclusions

In this article, we compared various estimators based on pretest and Stein-type strategy and two penalty estimators to the unrestricted and restricted maximum likelihood estimators in the context of the logistic regression model under the restriction of parameter. We established the properties of the proposed estimators via Monte Carlo simulation study.

By using Monte Carlo simulation study, we found that the performance of a restricted maximum likelihood estimator depends heavily on the quality of the candidate submodel. The restricted maximum likelihood estimator is the best estimator when the candidate submodel is correct or nearly correct. For any scenario, the estimators based on Stein-type strategy outperform the unrestricted maximum likelihood in the entire parameter space, especially the truncated version of the Stein-type shrinkage estimator. On the contrary, the performance of the estimators based on the preliminary test procedure lacks this property. The LASSO estimator is preferable to the Stein-type estimators when the number of inactive predictors is small. In contrast, the Stein-type estimators dominate the LASSO estimator only when the number of inactive predictors is relatively large. The ridge regression cannot produce a parsimonious model, thus the ridge regression estimator does not perform well when the model is sparse. However, the positive-part Stein-type shrinkage estimator is robust. It would be interesting, therefore, to investigate the relative performances of adaptive LASSO and smoothly clipped absolute deviation (SCAD) estimators. We will leave this for further consideration.

Acknowledgements. We would like to thank all the referees, the Editor, and the Associate Editor for their valuable suggestions on the revision of the article.

References

1. Ahmed SE (1992) Shrinkage preliminary test estimation in multivariate normal distributionsm. J Stat Comput Simul 43(3–4):177–195
2. Ahmed SE, Amezziane M (2016) Shrinkage-based semiparametric density estimation. Statistical Methodology

3. Ahmed SE, Raheem E, Hossain S (2011) Absolute Penalty Estimation. Springer International Encyclopedia of statistical science, New York
4. Ahmed SE, Yüzbaşı B (2016) Big data analytics: integrating penalty strategies. International Journal of Management Science & Engineering Management 11(2016):105–115
5. Al-Momani M, Hussein AA, Ahmed SE (2016) Penalty and related estimation strategies in the spatial error model: estimation strategies in sem model
6. Gao X, Ahmed SE, Feng Y (2016) Post selection shrinkage estimation for high-dimensional data analysis. Appl Stoch Models Bus Ind
7. Gourieroux CS, Monfort A (1981) Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. J Econometrics 17(1):83–97
8. Hilbe JM (2009) Logistic Regression Models. Chapman and Hall, Boca Raton
9. Hoerl AE, Kennard RW (2000) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 42(1):80–86
10. Hosmer DW, Lemeshow S (2000) Applied Logistic Regression. Wiley
11. Hossain S, Ahmed S (2014) Shrinkage estimation and selection for a logistic regression model
12. Hossain S, Ahmed SE, Doksum KA (2014) Shrinkage, pretest, and penalty estimators in generalized linear models. Stat Methodol 24(2):52–68
13. Lisawadi S, Shah MKA, Ahmed SE (2016) Model selection and post estimation based on a pretest for logistic regression models. J Stat Comput Simul 86(17):1–17
14. Rossouw JE, Du PJ, et al (1983) Coronary risk factor screening in three rural communities. The CORISs baseline study. S Afr Med J 64(12):430–436
15. Tibshirani R (2011) Regression shrinkage and selection via the lasso. J Roy Stat Soc 58(3):267–288
16. Yüzbaşı B, Ahmed SE (2015) Shrinkage ridge regression estimators in high-dimensional linear models. Springer, Heidelberg
17. Yüzbaşı B, Ahmed SE (2016) Shrinkage and penalized estimation in semi-parametric models with multicollinear data. J Stat Comput Simul

Springer
http://www.springer.com/978-3-319-59279-4

Proceedings of the Eleventh International Conference
on Management Science and Engineering Management
Zhu, M.; Gen, M.; Hajiyev, A.; Cooke, F.L. (Eds.)
2018, XXIV, 1801 p. 604 illus. In 2 volumes, not
available separately., Hardcover
ISBN: 978-3-319-59279-4