

# SUBSAMPLED DATA BASED ALTERNATIVE REGULARIZED ESTIMATORS

Subir Ghosh<sup>1\*</sup>, Gabriel Ruiz<sup>2</sup>, and Brandon Wales<sup>3</sup>

<sup>1,2,3</sup>Department of Statistics, University of California Riverside USA .

## Abstract

Subsampling the data is used in this paper as a learning method about the influence of the data points for drawing inference on the parameters of a fitted logistic regression model. The alternative, alternative regularized, alternative regularized lasso, and alternative regularized ridge estimators are proposed for the parameter estimation of logistic regression models and are then compared with the maximum likelihood estimators. The proposed alternative regularized estimators are obtained by using a tuning parameter but the proposed alternative estimators are not regularized. The proposed alternative regularized lasso estimators are the averaged standard lasso estimators and the alternative regularized ridge estimators are also the averaged standard ridge estimators over subsets of groups where the number of subsets could be smaller than the number of parameters. The values of the tuning parameters are obtained to make the alternative regularized estimators very close to the maximum likelihood estimators and the process is explained with two real data as well as a simulated study. The alternative and alternative regularized estimators always have the closed form expressions in terms of observations that the maximum likelihood estimators do not have. When the maximum likelihood estimators do not have the closed form expressions, the alternative regularized estimators thus obtained provide the approximate closed form expressions for them.

**Key words and phrases:** Item response, lasso, logistic regression, maximum likelihood, regularized, ridge, subsampling, tuning parameter.

# 1. Introduction

Scientific investigations often involve fitting a logistic regression model to the binary response data. Subsampling the data is implemented as a learning tool for enhancing the understanding of the estimation process in fitting a logistic regression model by harnessing the strength of information from some special subsampled data. The logistic regression model is popular for describing the binary response data (Agresti, 2012; Bishop, Fienberg, & Holland, 1975; Chaloner & Larntz, 1989; Cox, 1958; Efron, 1975; Engelhardt, 1975; Khan & Shaw, 2011; Webb, Wilson & Chong, 2004). In computer age statistical inference (Efron & Hastie, 2016), the computationally intensive statistical methods play an important role in paving the pathways for statistical learning and discoveries (James, Witten, Hastie, & Tibshirani, 2013).

The item response theory (IRT) refers to a family of mathematical models for describing the relationship between latent traits (unobservable variables) and their realizations (observed outcome variables). A link is established between the properties of items on an instrument, subjects responding to these items and the underlying trait being measured. IRT assumes that the latent construct and items of a measure are organized in an unobservable continuum. The main goal is to establish the subject position on that continuum. Since the subjects are different, the positions on the continuum are not identical and are determined by the subjects included in the research study and item parameters. An item should discriminate the subjects on the basis of their responses on the continuum. Item response theory psychometrics are frequently based upon logistic regression model (Lord (1983a,b& c, 1986), Rasch (1960), Stone (1992), Baker and Kim (2017)). The item response function of the two parameter logistic model for a dichotomous item is defined as

$$P_{ij}(\theta_j, b_i, a_i) = \frac{e^{[a_i(\theta_j - b_i)]}}{1 + e^{[a_i(\theta_j - b_i)]}},$$

where  $\theta_j$  is the *ability* parameter of the individual  $j$ ,  $b_i$  is the *difficulty* parameter of the item  $i$  or the item parameter,  $a_i$  is the *discrimination* parameter and  $P_{ij}(\theta_j, b_i, a_i)$  is the probability of correct response for the individual  $j$  on the item  $i$ .

For the convenience of presentation, the following abbreviations are used for the estimators throughout the paper.

Abbreviation	Estimator
AE	Alternative Estimator
ARE	Alternative Regularized Estimator
ARLE	Alternative Regularized Lasso Estimator
ARRE	Alternative Regularized Ridge Estimator
MLE	Maximum Likelihood Estimator

Consider a response variable  $Y$  generated from a binary random variable taking two realized values 1 (success) and 0 (failure). One of the most popular models for studying the dependence of the response variable  $Y$  on  $q$  explanatory variables  $X_1, \dots, X_q$  in  $N$  groups, is the logistic regression model. Denote the vector of  $q$  explanatory variables by  $(X_1, \dots, X_q)'$ . The response variable  $Y$  for the  $i^{\text{th}}$  group is denoted by  $Y_i$  and the corresponding vector of explanatory variables by  $\mathbf{X}_i = (X_{i1}, \dots, X_{iq})'$  for  $i = 1, \dots, N$ . For the  $i^{\text{th}}$  group, the probability that the binary random variable taking the value 1 is  $p_i$  and the value 0 is  $(1 - p_i)$ . The total number of realized values of the binary random variable is  $n_i$ ,  $i = 1, \dots, N$ . The  $Y_i$  takes the value  $y_i$  as the number of observed value 1 and  $(n_i - y_i)$  as the number of observed value 0. The random variables  $Y_1, \dots, Y_N$  are independent binomial random variables having

$$E(Y_i) = n_i p_i, \text{ Var}(Y_i) = n_i p_i (1 - p_i), \text{ and Cov}(Y_i, Y_{i'}) = 0, \text{ for } i \neq i'.$$

The realized value of the vector  $\mathbf{X}_i$  is  $\mathbf{x}_i = (x_{i1}, \dots, x_{iq})'$ , for  $i = 1, \dots, N$ . In the logistic regression model

$$\log_e \left( \frac{p_i}{1 - p_i} \right) \stackrel{\text{def}}{=} \alpha + \beta' \mathbf{x}_i, p_i \stackrel{\text{def}}{=} \frac{e^{\alpha + \beta' \mathbf{x}_i}}{1 + e^{\alpha + \beta' \mathbf{x}_i}}, \quad (1.1)$$

where  $\alpha$  and  $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_q)$  are the unknown parameters. The likelihood function is denoted by  $L(\alpha, \boldsymbol{\beta})$ . The score functions of the first derivatives of the log-likelihood function  $\ell(\alpha, \boldsymbol{\beta}) (= \log L(\alpha, \boldsymbol{\beta}))$  with respect to  $\alpha$  and the elements of  $\boldsymbol{\beta}$  are denoted by  $\mathbf{U}(\alpha, \boldsymbol{\beta})$ . The MLEs of  $\alpha$  and  $\boldsymbol{\beta}$ , denoted by  $\hat{\alpha}$  and  $\hat{\boldsymbol{\beta}}$ , satisfy the MLEEs  $\mathbf{U}(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = \mathbf{0}$ . The Newton–Raphson and Fisher scoring iterative methods provide the fast numerical values of  $\hat{\alpha}$  and  $\hat{\boldsymbol{\beta}}$  using the computational softwares, they do not provide any closed form expressions of  $\hat{\alpha}$  and  $\hat{\boldsymbol{\beta}}$ . Consequently, the exact statistical inferences about them are not possible with their numerical values particularly in small samples.

Section 2 considers two cases  $N = (q + 1)$  and  $N \geq (q + 1)$ . For the first case  $N = (q + 1)$ , the closed form expressions of MLEs are obtained when the observed values of  $y_i$  are not equal to 0 and  $n_i$ ,  $i = 1, \dots, N$ . The AEs are then proposed for the second case  $N \geq (q + 1)$  when the closed form expressions of MLEs are not available. The AREs are introduced in Section 3 as an approximation for MLEs. The closed form expressions of AREs are available and are dependent on a tuning parameter. Two real data examples are also presented in this section to illustrate the proposed AREs. In Section 4, the ARLEs and the ARREs are introduced for  $N \geq$  or  $\leq$  or  $= (q + 1)$  and compared for the second illustrative example. A simulation study is given in Section 5. The closing Section 6 draws conclusions.

## 2. MLEs and AEs

Denote an  $(N \times (q + 1))$  matrix  $\mathbf{X}$  and a  $(1 \times N)$  vector  $\mathbf{y}$  by

$$\mathbf{X}' = \begin{bmatrix} 1 & \dots & 1 & \dots & 1 \\ \mathbf{x}_1 & \dots & \mathbf{x}_i & \dots & \mathbf{x}_N \end{bmatrix}, \mathbf{y} = (y_1, \dots, y_N)'$$

Although the closed form expressions of MLEs  $\hat{\alpha}$  and  $\hat{\boldsymbol{\beta}}$  are not available for the general case  $N \geq (q + 1)$ , they are available for the special situation  $N = (q + 1) = \text{Rank}(\mathbf{X})$  and  $y_i \neq 0$  and  $y_i \neq n_i$  for  $i = 1, \dots, N$ . Alternative estimators (AEs) of  $\alpha$  and  $\boldsymbol{\beta}$  are proposed for the general case  $N \geq (q + 1)$  using the subsampled data and harnessing the learning from the situation  $N = (q + 1) = \text{Rank}(\mathbf{X})$ .

### 2.2. $N = (q + 1) = \text{Rank}(\mathbf{X})$ , $y_i \neq 0$ , $n_i$ , $i = 1, \dots, N$

For the special situation  $N = q + 1$ , the MLEEs  $\mathbf{U}(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = \mathbf{0}$  become

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\hat{\mathbf{y}}. \quad (2.1)$$

When  $\text{Rank}(\mathbf{X}) = N = (q + 1)$ , it follows from (2.1) that  $\mathbf{y} = \hat{\mathbf{y}}$ . In addition, when  $y_i \neq 0$  and  $y_i \neq n_i$  for  $i = 1, \dots, (N = (q + 1))$ , it can be seen

$$y_i = \hat{y}_i = \frac{n_i e^{\hat{\alpha} + \hat{\boldsymbol{\beta}}' \mathbf{x}_i}}{1 + e^{\hat{\alpha} + \hat{\boldsymbol{\beta}}' \mathbf{x}_i}}, \hat{\alpha} + \hat{\boldsymbol{\beta}}' \mathbf{x}_i = \log \frac{y_i}{n_i - y_i}. \quad (2.2)$$

Denote

$$\boldsymbol{\beta}^* = (\alpha, \beta_1, \dots, \beta_q)', \widehat{\boldsymbol{\beta}}^* = (\widehat{\alpha}, \widehat{\beta}_1, \dots, \widehat{\beta}_q)', \mathbf{y}^* = \left( \log \frac{\widehat{y}_1}{n_1 - \widehat{y}_1}, \dots, \log \frac{\widehat{y}_N}{n_N - \widehat{y}_N} \right)'$$

It can be seen from (2.3) that

$$\mathbf{X}\widehat{\boldsymbol{\beta}}^* = \mathbf{y}^*, \widehat{\boldsymbol{\beta}}^* = \mathbf{X}^{-1}\mathbf{y}^*. \tag{2.3}$$

The (2.4) provides the closed form expression of  $\widehat{\boldsymbol{\beta}}^*$ , the MLE of  $\boldsymbol{\beta}^*$ , for the special situation  $N = q + 1 = \text{Rank}(\mathbf{X})$  and  $y_i \neq 0$  and  $y_i \neq n_i$  for  $i = 1, \dots, (N = (q + 1))$ .

**2.2. Subsampled data when  $N \geq (q + 1)$**

Consider the general situation when  $N \geq (q + 1)$ . The class  $\mathbf{S}$  of all possible  $\mathbf{v} (= \binom{N}{q+1})$  subsamples of  $(q + 1)$  groups are considered to use (2.3) to find the closed form estimators of  $\boldsymbol{\beta}^*$ . Suppose that the closed form estimators of  $\boldsymbol{\beta}^*$  can be found for the subclass  $\mathbf{S}^*$  of  $\mathbf{S}$ , consisting of  $v^*$  subsamples out of  $v$  subsamples. The  $(v - v^*)$  subsamples in  $\mathbf{S}$  but not in  $\mathbf{S}^*$  do not provide the closed form estimators because the rank condition  $\text{Rank}(\mathbf{X}) = N = (q + 1)$  may not hold or the  $y_i$  value could be  $n_i$  or 0. For the  $j^{\text{th}}$  subsample of  $\mathbf{S}^*$ , the closed form estimator of  $\boldsymbol{\beta}^*$  is denoted by  $\widehat{\boldsymbol{\beta}}_j^*$ . The proposed alternative estimator of  $\boldsymbol{\beta}^*$  is

$$\widehat{\boldsymbol{\beta}}_{\text{AE}}^* \stackrel{\text{def}}{=} \frac{1}{v^*} \sum_{j \in \mathbf{S}^*} \widehat{\boldsymbol{\beta}}_j^*. \tag{2.4}$$

The proposed alternative estimator  $\widehat{\boldsymbol{\beta}}_{\text{AE}}^*$  in (2.4) is different from the MLE  $\widehat{\boldsymbol{\beta}}^*$  when  $N > (q + 1)$ . However, the proposed alternative estimator  $\widehat{\boldsymbol{\beta}}_{\text{AE}}^*$  is identical to the MLE  $\widehat{\boldsymbol{\beta}}^*$  in (3.9) when  $N = (q + 1)$ .

**2.3. Example 1**

The Heart Disease and Blood Pressure Data given on page 217 in Table 6.5 of Agresti (2013) consist of a sample of male residents of Framingham, Massachusetts, aged 40 to 59, having the binary response variable  $Y$  as whether the residents developed coronary heart disease (Yes = 1) during a six year follow up period and the explanatory variable (the risk factor) as the systolic blood pressure  $X_1$  obtained from the range. Thus,  $q = 1$  for this example.

Table 1 summarizes the data in the format of the presentation of this paper. Note that  $N = 8, q + 1 = 2$ , and  $N > (q + 1)$ . Moreover,  $v = \binom{8}{2} = 28 = v^*$ . The MLE of  $\boldsymbol{\beta}^* = (\alpha, \beta_1)'$  is  $\widehat{\boldsymbol{\beta}}^* = (\widehat{\alpha}, \widehat{\beta}_1)' = (-6.082, 0.024)'$ . The log-likelihood function  $\ell(\alpha, \beta_1)$  at  $\alpha = \widehat{\alpha}$  and  $\beta_1 = \widehat{\beta}_1$  is  $-19.305$ . On the other hand, the proposed alternative estimate  $\widehat{\boldsymbol{\beta}}_{\text{AE}}^*$  in (2.4) is  $(-6.514, 0.028)'$  and the log-likelihood function  $\ell(\alpha, \beta_1)$  at  $\alpha = -6.514$  and  $\beta_1 = 0.028$  is  $-19.558$ .

**2.4. Example 2**

Smoking, obesity, and snoring were investigated for hypertension in 433 men aged 40 or over and the data are given in Altman (1991). Table 2 summarizes the data in the format of this paper. Note that  $N = 8, q + 1 = 4$ , and  $N > (q + 1)$ . The MLE of  $\boldsymbol{\beta}^* = (\alpha, \beta_1, \beta_2, \beta_3)'$  is  $\widehat{\boldsymbol{\beta}}^* = (\widehat{\alpha}, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)' = (-2.378, -0.068, 0.695, 0.872)'$  and the log-likelihood function is  $\ell(\widehat{\boldsymbol{\beta}}^*) = -13.269$ .

The numerical value of  $v$  becomes  $v = \binom{8}{q+1} = \binom{8}{4} = 70$ . Excluding the 35 sub-samples including Group 4 having  $y_4 = 0$  and observing that the rank condition  $\text{Rank}(\mathbf{X}) = 4$  does not

Table 1: The Heart Disease and Blood Pressure Data (Agresti(2013))

Group ( $i$ )	Range	$x_{i1}$	$y_i$	$n_i$
1	< 117	111.5	3	156
2	117 – 126	121.5	17	252
3	127 – 136	131.5	12	284
4	137 – 146	141.5	16	271
5	147 – 156	151.5	12	139
6	157 – 166	161.5	8	85
7	167 – 186	176.5	16	99
8	> 186	191.5	8	43

hold for 6 subsamples out of the remaining  $(70 - 35) = 35$  subsamples, the numerical value of  $v^*$  is  $(35 - 6) = 29$ . The proposed alternative estimate is  $\hat{\beta}_{AE}^* = (-2.399, 0.173, 0.665, 0.796)'$  and the log-likelihood function is  $\ell(\hat{\beta}_{AE}^*) = -13.683$ .

Table 2: The Hypertension Data (Altman (1991))

Group $i$	Smoking $x_{i1}$	Obesity $x_{i2}$	Snoring $x_{i3}$	Hypertensive $y_i$	Mean $n_i$
1	0	0	0	5	60
2	1	0	0	2	17
3	0	1	0	1	8
4	1	1	0	0	2
5	0	0	1	35	187
6	1	0	1	13	85
7	0	1	1	15	51
8	1	1	1	8	23

### 3. AREs

The proposed alternative regularized estimators (AREs) depend on a non-negative integer valued tuning (regularization) parameter  $\lambda$  where  $0 \leq \lambda \leq (v^* - 1)$ . For a fixed  $\lambda$ , in the class  $\mathbf{S}^*$  of  $v^*$  subsamples of  $(q + 1)$  groups, consider  $\binom{v^*}{\lambda}$  possible subclasses, each consisting of  $(v^* - \lambda)$  subsamples. For the  $j^{\text{th}}$  subsample within the  $g(\lambda)^{\text{th}}$  subclass  $\mathbf{S}_{g(\lambda)}^*$ ,  $g(\lambda) = 1, \dots, \binom{v^*}{\lambda}$ , denote the closed form estimator of  $\beta^*$  by  $\hat{\beta}_{jg(\lambda)}^*$  using (4) and obtain the alternative estimator following

(2.4) as

$$\widehat{\beta}_{\text{AE}(\lambda)}^* \stackrel{\text{def}}{=} \frac{1}{(v^* - \lambda)} \sum_{j \in S_{\text{g}(\lambda)}^*} \widehat{\beta}_{j\text{g}(\lambda)}^*, \quad \text{g}(\lambda) = 1, \dots, \binom{v^*}{\lambda}. \quad (3.5)$$

Denote  $\text{LL}(\lambda) = \ell\left(\widehat{\beta}_{\text{AE}(\lambda)}^*\right)$ . Find a  $\lambda$  ( $= \lambda_M$ , say) maximizing the log-likelihood,  $\text{LL}(\lambda)$ , for  $\lambda = 1, \dots, (v^* - 1)$ . In other words,  $\lambda_M$  is the value of  $\lambda$  satisfying

$$\text{LL}(\lambda_M) \stackrel{\text{def}}{=} \max_{\lambda} [\text{LL}(\lambda)]. \quad (3.6)$$

Define the alternative regularized estimator (ARE) of  $\beta^*$  as

$$\widehat{\beta}_{\text{ARE}}^*(\lambda_M) \stackrel{\text{def}}{=} \widehat{\beta}_{\text{AE}(\lambda_M)}^*. \quad (3.7)$$

Notice that  $\widehat{\beta}_{\text{ARE}}^*(\lambda_M)$  depends on the tuning parameter  $\lambda$  through  $\widehat{\beta}_{\text{AE}(\lambda)}^*$ , for all values of  $\lambda$ ,  $0 \leq \lambda \leq (v^* - 1)$ . The (2.4) and (3.5) become equal for  $\lambda = 0$  or equivalently,

$$\widehat{\beta}_{\text{AE}(0)}^* = \widehat{\beta}_{\text{AE}}^*. \quad (3.8)$$

The value  $\lambda_M$  is the choice for the tuning parameter  $\lambda$  to satisfy

$$\widehat{\beta}_{\text{ARE}}^*(\lambda_M) \simeq \widehat{\beta}^*, \quad \text{LL}(\lambda_M) = \ell\left(\widehat{\beta}_{\text{ARE}}^*(\lambda_M)\right) \simeq \ell\left(\widehat{\beta}^*\right). \quad (3.9)$$

Consequently, the MLE is approximately equal to the ARE for the choice of tuning parameter  $\lambda = \lambda_M$ .

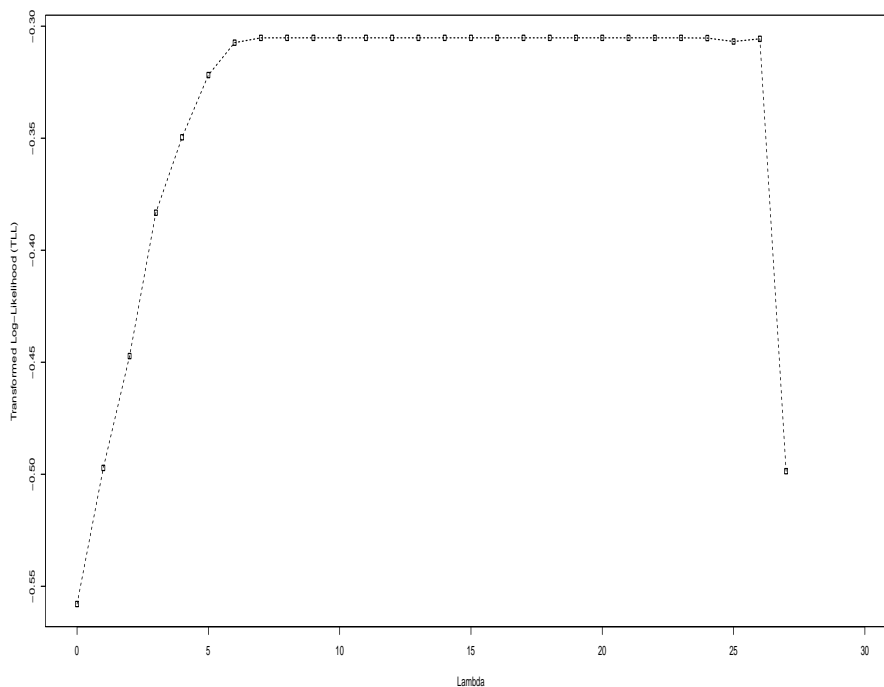
In Example 1 of Section 2.3, it can be seen that  $\lambda_M = 20$ . Define two Transformed Log-Likelihood (TLL)s from  $\text{LL}(\lambda)$  as

$$\text{TLL1}(\lambda) = (\text{LL}(\lambda) + 19), \quad \text{TLL2}(\lambda) = (\text{LL}(\lambda) + 19.305) \times 10^9.$$

Figure 1 displays the plot of  $\text{TLL1}(\lambda)$  against  $\lambda$  ( $= 0, 1, \dots, 27$ ). Figure 2 highlighting the maximum (MAX) of Figure 1 presents the plot of  $\text{TLL2}(\lambda)$  against  $\lambda$  ( $= 8, 9, \dots, 21$ ). From Table 3 and Figure 2,

$\widehat{\beta}^* = (-6.082, 0.024)'$  and  $\widehat{\beta}_{\text{ARE}}^*(20) = (-6.082, 0.024)'$  are identical up to 4 (after truncation) decimal places for the first component and 6 decimal places for the second component. Moreover,  $\ell\left(\widehat{\beta}_{\text{ARE}}^*(20)\right) = -19.30519373$  and  $\ell\left(\widehat{\beta}^*\right) = -19.30519372$  are identical up to 7 decimal places. Hence it follows from (3.9) that  $\widehat{\beta}_{\text{ARE}}^*(20) \simeq \widehat{\beta}^*$ ,  $\ell\left(\widehat{\beta}_{\text{ARE}}^*(20)\right) \simeq \ell\left(\widehat{\beta}^*\right)$ .

Figure 3 displays the plot of  $\text{LL}(\lambda)$  against  $\lambda = 0, 1, \dots, 28$  for  $N = 8$  and  $q = 3$ . On the other hand, Figure 4 displays the plot of  $\text{LL}(\lambda)$  against  $\lambda = 0, 1, \dots, 19$  for  $N = 7$  and  $q = 2$ .

Figure 1: Plot of TLL1 against  $\lambda$ .

For  $N = 8$  and  $q = 3$ , it can be seen that

$$\hat{\beta}^* = (-2.378, -0.068, 0.695, 0.872)' \text{ and } \ell(\hat{\beta}^*) = -13.269.$$

For  $\lambda = 22$ ,

$$\hat{\beta}_{\text{ARE}}^*(22) = (-2.377, 0.143, 0.651, 0.814)'$$

$$\text{and } \ell(\hat{\beta}_{\text{ARE}}^*(22)) = -13.277,$$

which is the maximum value of  $\text{LL}(\lambda)$  for  $\lambda = 0, 1, \dots, 28$ . The  $\lambda_M = 22$  is then approximating  $\hat{\beta}^*$  by  $\hat{\beta}_{\text{ARE}}^*(22)$ .

Some aspects of the fitting statistics for logistic regression model to the data in Table 2 are given in Table 4. The explanatory variable “smoking” is not significant in having the  $P$  value as 0.808. Deleting the explanatory variable smoking from the model,  $q + 1$  becomes 3. The numerical value of  $v$  becomes  $v = \binom{7}{q+1} = \binom{7}{3} = 35$ . The rank condition,  $\text{Rank}(\mathbf{X}) = 4$ , holds for only 20 subsamples and does not hold for 15 out of 35 subsamples. Therefore,  $v^* = 20$ .

For  $N = 7$  and  $q = 2$  excluding “smoking” as an explanatory variable, it can be seen from Figure 4 that the maximum value -13.302061150 of  $\text{LL}(\lambda)$  for  $\lambda = 0, 1, \dots, 19$  is attained at  $\lambda_M = 3$  for approximating  $\hat{\beta}^*$  by  $\hat{\beta}_{\text{ARE}}^*(3)$ .

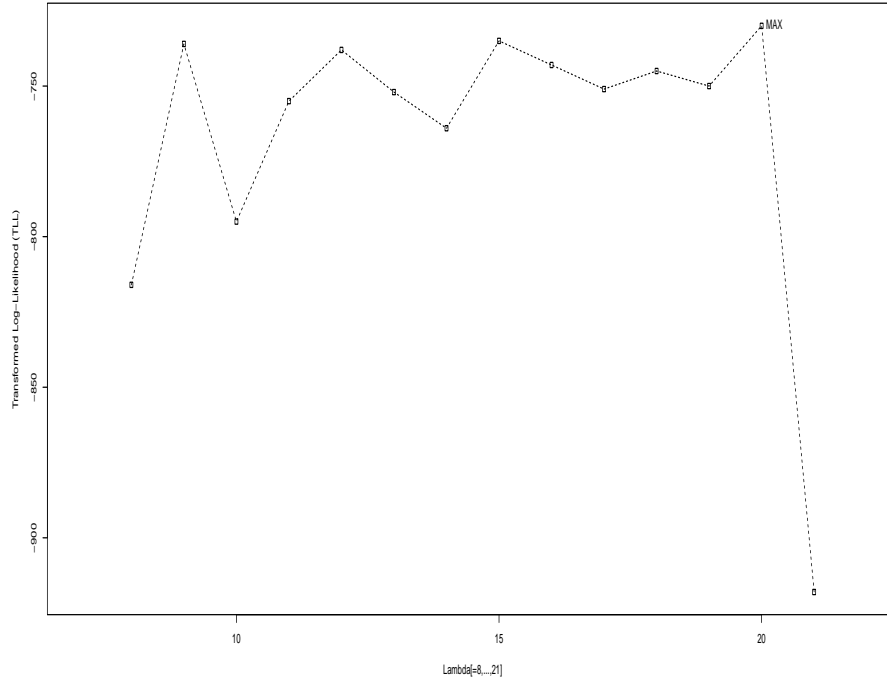


Figure 2: Plot of TLL2 against the partial values of  $\lambda[= 8, \dots, 21]$ .

#### 4. ARLE and ARRE for $N > \text{or} < \text{or} = (q+1)$

In this section,  $N$  can be equal to, greater or even less than  $(q + 1)$ . Groups having  $y_i = 0$  (Group 4 in Example 2) or  $n_i = y_i$  are ignored to reduce the value of  $N$ . Choose a number  $\gamma$ ,  $0 < \gamma < N$ , when  $N \geq (q + 1)$  or a number  $\gamma$ ,  $0 < \gamma \leq N$ , when  $N < (q + 1)$ . For the lasso and ridge estimations, the tuning parameters are denoted by  $\delta_L$  and  $\delta_R$  respectively. The optimum values of  $\delta_L$  and  $\delta_R$  are determined by minimizing the Generalized Cross Validations (GCVs) (Eq. (7.46), p. 217, Hastie et al. (2001)) for all possible  $\binom{N}{\gamma}$  subsamples of groups. Then the Lasso estimates  $\hat{\beta}_L^*(\delta_L)$  and the Ridge estimates  $\hat{\beta}_R^*(\delta_R)$  are then calculated for  $\binom{N}{\gamma}$  subsamples of groups. The average of  $\binom{N}{\gamma}$  both Lasso and Ridge estimates provide  $\hat{\beta}_{ARLE}^*$  and  $\hat{\beta}_{ARRE}^*$  similar to  $\hat{\beta}_{AE}^*$  in (5). In Example 2 excluding Group 4,  $N = 7$  and  $(q + 1) = 4$ . Taking  $\gamma = 4$ , the Lasso estimates  $\hat{\beta}_L^*(\delta_L)$  and  $\ell(\hat{\beta}_L^*(\delta_L))$  for 35 subsamples of groups are displayed in Table 5 for  $\gamma = 4$ . It can be seen from Table 5 that

$$\hat{\beta}_{ARLE}^* = (-1.977, 0.074, 0.362, 0.370)' \text{ and therefore } \ell(\hat{\beta}_{ARLE}^*) = -15.221.$$

Again taking  $\gamma = 4$ , the Ridge estimates  $\hat{\beta}_R^*(\delta_R)$  and  $\ell(\hat{\beta}_R^*(\delta_R))$  for 35 subsamples of groups are displayed in Table 6. Moreover, it can be seen from Table 6 that

$$\hat{\beta}_{ARRE}^* = (-2.013, 0.094, 0.365, 0.428)' \text{ and therefore } \ell(\hat{\beta}_{ARRE}^*) = -14.948.$$

Taking  $\gamma = 3$ , the Lasso estimates  $\hat{\beta}_L^*(\delta_L)$  and  $\ell(\hat{\beta}_L^*(\delta_L))$  for 35 subsamples of groups are displayed in Table 7. It can be seen from Table 7 that

$$\hat{\beta}_{ARLE}^* = (-1.751, 0.067, 0.191, 0.124)' \text{ and therefore } \ell(\hat{\beta}_{ARLE}^*) = -17.506.$$



Table 3: Comparison among MLE, AE, and ARE for  $\lambda = 20$

	MLE	AE $\equiv$ (ARE for $\lambda = 0$ )	ARE for $\lambda = 20$
$\beta^* :$	$\hat{\beta}^*$	$\hat{\beta}_{AE}^*$	$\hat{\beta}_{ARE}^*(20)$
$\bar{\alpha}$	- 6.082	- 6.514	- 6.082
$\bar{\beta}_1$	0.024	0.028	0.024
$\ell(\beta^*)$	$\ell(\hat{\beta}^*)$ - 19.305	$\ell(\hat{\beta}_{AE}^*)$ - 19.558	$\ell(\hat{\beta}_{ARE}^*(20))$ - 19.305

Table 4: Fitting statistics for logistic regression model to the Hypertension Data (Altman (1991))

Explanatory Variable	Parameter	MLE	Std. Error	$z$ value	$P$ value
Intercept	$\alpha$	- 2.378	0.381	- 6.254	$4 e^{-10}$
Smoking	$\beta_1$	- 0.068	0.278	- 0.244	0.808
Obesity	$\beta_2$	0.695	0.285	2.439	0.015
Snoring	$\beta_3$	0.872	0.398	2.193	0.028

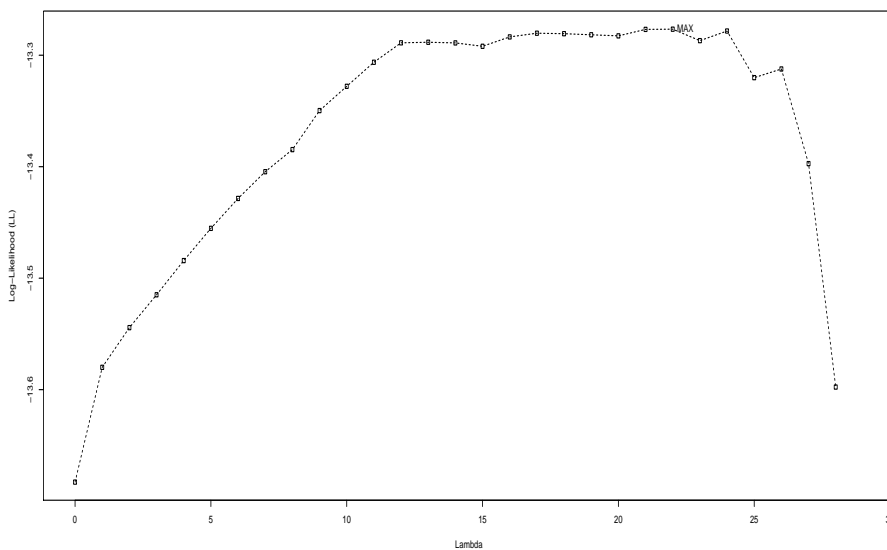


Figure 3: Plot of  $LL(\lambda)$  against  $\lambda$  for  $N = 8$  and  $q = 3$ .

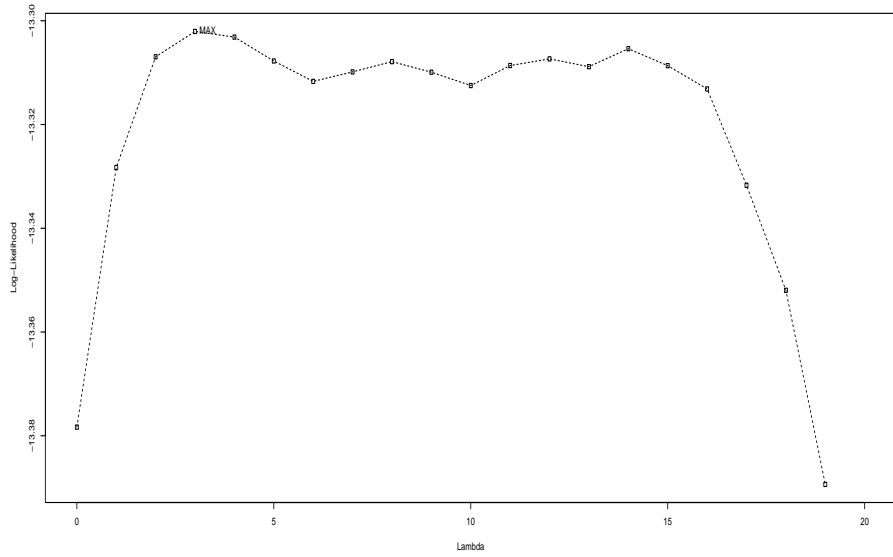


Figure 4: Plot of  $LL(\lambda)$  against  $\lambda$  for  $N = 8$  and  $q = 2$ .

Again taking  $\gamma = 3$ , the Ridge estimates  $\hat{\beta}_R^*(\delta_R)$  and  $\ell(\hat{\beta}_R^*(\delta_R))$  for 35 subsamples of groups are displayed in Table 6. Moreover, it can be seen from Table 8 that  $\hat{\beta}_{ARRE}^* = (-1.600, 0.005, 0.020, 0.021)'$  and therefore  $\ell(\hat{\beta}_{ARRE}^*) = -19.382$ . The results in Table 9 demonstrate that ARE ( $\lambda = 22$ ) is closest to MLE. Lasso(1,5,7), Lasso(1,5,6,7), AE, and Ridge(1,5,6,7) are also closer to MLE but behind ARE ( $\lambda = 22$ ). ARLE( $\gamma = 4$ ), ARLE( $\gamma = 3$ ), ARRE( $\gamma = 4$ ), and ARRE( $\gamma = 3$ ) are a bit away from MLE.

### 5. A Simulation Study

The 1,000 data sets are generated on  $y_i, i = 1, \dots, 8$  under the logistic regression model in (1) for  $q = 1$  having the assumed values of  $\alpha = -6.08$  and  $\beta_1 = 0.025$  and keeping the  $(x_{i1}, n_i)$  values the same as in Table 1. For the  $i^{th}$  data set the numerical values of MLE  $\hat{\beta}^*$ , AE  $\hat{\beta}_{AE}^*$ , and ARE  $\hat{\beta}_{ARE}^*(\lambda), 0 \leq \lambda \leq 27$  are obtained and denote them by  $\hat{\beta}^{*(i)} = (\hat{\alpha}^{(i)}, \hat{\beta}_1^{(i)})', \hat{\beta}_{AE}^{*(i)} = (\hat{\alpha}_{AE}^{(i)}, \hat{\beta}_{1AE}^{(i)})'$ , and  $\hat{\beta}_{ARE}^{*(i)}(\lambda) = (\hat{\alpha}_{ARE}^{(i)}(\lambda), \hat{\beta}_{1ARE}^{(i)}(\lambda))'$ . Define

$$\begin{aligned} \bar{\alpha} &= \frac{1}{1000} \sum_{i=1}^{1000} \hat{\alpha}^{(i)}, & \bar{\alpha}_{AE} &= \frac{1}{1000} \sum_{i=1}^{1000} \hat{\alpha}_{AE}^{(i)}, & \bar{\alpha}_{ARE}(\lambda) &= \frac{1}{1000} \sum_{i=1}^{1000} \hat{\alpha}_{ARE}^{(i)}(\lambda), \\ \bar{\beta}_1 &= \frac{1}{1000} \sum_{i=1}^{1000} \hat{\beta}_1^{(i)}, & \bar{\beta}_{1AE} &= \frac{1}{1000} \sum_{i=1}^{1000} \hat{\beta}_{1AE}^{(i)}, & \bar{\beta}_{1ARE}(\lambda) &= \frac{1}{1000} \sum_{i=1}^{1000} \hat{\beta}_{1ARE}^{(i)}(\lambda). \end{aligned} \tag{5.10}$$

The log likelihoods  $\ell(\hat{\beta}^*), \ell(\hat{\beta}_{AE}^*),$  and  $\ell(\hat{\beta}_{ARE}^*(\lambda)), 0 \leq \lambda \leq 27$  are also obtained. Then their averages are found over 1,000 data sets.

$$\bar{\ell}(\widehat{\boldsymbol{\beta}}^*) = \frac{1}{1000} \sum_{i=1}^{1000} \ell(\widehat{\boldsymbol{\beta}}^{*(i)}), \quad \bar{\ell}(\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*) = \frac{1}{1000} \sum_{i=1}^{1000} \ell(\widehat{\boldsymbol{\beta}}_{\text{ARE}}^{*(i)}), \quad \bar{\ell}(\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(\lambda)) = \frac{1}{1000} \sum_{i=1}^{1000} \ell(\widehat{\boldsymbol{\beta}}_{\text{ARE}}^{*(i)}(\lambda)). \quad (5.11)$$

Table 10 presents the averages in (5.10) and (5.11) with the exceptions on  $\widehat{\boldsymbol{\beta}}_{\text{1ARE}}^*(\lambda)$  and  $\bar{\ell}(\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(\lambda))$  only for the  $\lambda$  values: 6, 21, and 25 to save the space.

The value of  $\lambda_{\text{M}}$  satisfying (3.9) is determined from Table 10 as  $\lambda_{\text{M}} = 21$  giving  $\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(21) = (-6.083899, 0.025019)'$  and  $\bar{\ell}(\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(21)) = -19.714012$  in contrast to  $\widehat{\boldsymbol{\beta}}^* = (-6.083941, 0.025019)'$  and  $\bar{\ell}(\widehat{\boldsymbol{\beta}}^*) = -19.714009$ . The true value of  $\boldsymbol{\beta}^* = (\alpha, \beta_1)' = (-6.08, 0.025)'$ . The value of  $\lambda_{\text{T}}$  satisfying

$$\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(\lambda_{\text{T}}) \simeq \boldsymbol{\beta}^*, \quad (5.12)$$

is also obtained from Table 10. A choice for  $\lambda_{\text{T}}$  is 25 having  $\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(25) = (-6.083519, 0.025016)'$  closest to  $\boldsymbol{\beta}^* = (-6.080, 0.025)'$  but giving  $\bar{\ell}(\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(25)) = -19.716031$ . Denote the mean square error by MSE and the mean absolute difference by MAD. Define

$$\begin{aligned} \text{MSE}(\widehat{\alpha}) &= \frac{1}{1000} \sum_{i=1}^{1000} (\widehat{\alpha}^{(i)} + 6.080)^2, & \text{MAD}(\widehat{\alpha}) &= \frac{1}{1000} \sum_{i=1}^{1000} |\widehat{\alpha}^{(i)} + 6.080|, \\ \text{MSE}(\widehat{\beta}_1) &= \frac{1}{1000} \sum_{i=1}^{1000} (\widehat{\beta}_1^{(i)} - 0.025)^2, & \text{MAD}(\widehat{\beta}_1) &= \frac{1}{1000} \sum_{i=1}^{1000} |\widehat{\beta}_1^{(i)} - 0.025|. \end{aligned} \quad (5.13)$$

Table 11 presents the numerical values of  $\text{MSE}(\widehat{\alpha})$ ,  $\text{MSE}(\widehat{\beta}_1)$ ,  $\text{MAD}(\widehat{\alpha})$ , and  $\text{MAD}(\widehat{\beta}_1)$  for  $\widehat{\boldsymbol{\beta}}^*$ ,  $\widehat{\boldsymbol{\beta}}_{\text{AE}}^*$ ,  $\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(\lambda)$  when  $\lambda = 6, 21$ , and 25. It is clear from Table 11 that the best choice for  $\lambda_{\text{T}}$  is 21 with respect to the criterion functions MSE and MAD. Note that

$$\begin{aligned} -6.080000 &< \bar{\alpha} = -6.083941 < -6.083899 = \bar{\alpha}_{\text{ARE}}(21), \\ 0.025000 &< 0.025019 = \bar{\beta}_1 = \bar{\beta}_{\text{1ARE}}(21), \\ \bar{\ell}(\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(21)) &= -19.714012 < -19.714009 = \bar{\ell}(\widehat{\boldsymbol{\beta}}^*), \\ \frac{1}{1000} \sum_{i=1}^{1000} |\widehat{\alpha}_{\text{ARE}}^{(i)}(21) - \widehat{\alpha}^{(i)}| &= 0.000812, \\ \frac{1}{1000} \sum_{i=1}^{1000} |\widehat{\beta}_{\text{1ARE}}^{(i)}(21) - \widehat{\beta}_1^{(i)}| &= 0.000005, \end{aligned} \quad (5.14)$$

demonstrating the closeness of the true value of  $\boldsymbol{\beta}^* = (\alpha, \beta_1)' = (-6.08, 0.025)'$  to the averaged MLEs  $\widehat{\boldsymbol{\beta}}^* = (-6.083941, 0.025019)'$  as well as to the averaged AREs  $\widehat{\boldsymbol{\beta}}_{\text{ARE}}^*(21) = (-6.083899, 0.025019)'$  from the 1000 simulated data sets. In other words, the performance of ARE for the choice of the tuning parameter  $\lambda = 21$  is almost equivalent to MLE.

## 6. Conclusions

Subsampling methods are implemented to obtain the AE, AREs, ARLEs, ARREs. By using these estimators, new insights are gained for fitting the logistic models to the data. The proposed AREs depend on the tuning parameter  $\lambda$  and AEs. The tuning parameter value can be obtained to make the ARE and its likelihood approximately equal to the corresponding values of MLE. Two real data one with an explanatory variable and the other with three explanatory variables are presented to illustrate the proposed methods. Example 2 demonstrates the closeness of ARE for a choice of the tuning parameter  $\lambda$  relative to AE, ARLE, ARRE, Lasso and Ridge for specific sub-sampled data. A simulated data example is also presented to demonstrate the almost equivalent performance of ARE and MLE for a choice of tuning parameter. The AREs have closed form expressions helpful to study their statistical properties even for small samples and are good closed form approximate representations for MLEs particularly when they do not exist.

The future investigation of this work will include the detailed theoretical properties of the AE, AREs, ARLEs and ARREs in comparison with the MLEs. More detailed simulation studies will be beneficial. The investigation for the high dimensional data will be another important direction. For the longitudinal data (Chan, 2014; Ghosh & Chakravartty, 2009), the investigation remains to be performed. In the area of quantitatively estimate the accuracy of artificial intelligence in machine learning (Hastie, T. *et al.*, 2001; James, G. *et al.*, 2013), IRT could be very useful and our proposed average estimators could be applicable.

## Acknowledgements

The authors are grateful to the Editor and the entire review team for their constructive comments on the earlier version of this paper.

## References

- [1] Agresti, A. (2012). *Categorical Data Analysis*, Wiley: New York, 3rd Edition.
- [2] Altman, D. G. (1991). *Practical Statistics for Medical Research*, Chapman & Hall: London, 1st Edition.
- [3] Baker, F. B. and Kim, S-H (2017). *The Basics of Item Response Theory Using R*, Springer : New York.
- [4] Bishop, Y.M., Fienberg, S.E., and Holland, P.W. (1975). *Discrete Multivariate Analysis*, MIT Press : Cambridge, MA.
- [5] Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference.* **21**,191–208.
- [6] Chan, J. S. K. (2014). Analysis of correlation structures using generalized estimating equation approach for longitudinal binary data. *Journal of Data Science.* **12**, 293–305.
- [7] Cox, D. R. (1958). Regression analysis of binary sequences. *Journal of Royal Statistical Society.Ser. B.* **20**, 215–242.
- [8] Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association.* **70**, 892–898.

- [9] Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*, Cambridge University Press: University of Cambridge,UK.
- [10] Engelhardt, M. (1975). *Linear estimation of the parameters of the logistic distribution from a complete or censored sample*. *Journal of the American Statistical Association*. **70**, 899–902.
- [11] Ghosh, S. and Chakravartty, A. (2009). Some observations in likelihood based fitting of longitudinal models for binary data. *Journal of Data Science*. **7**, 409–421.
- [12] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer: New York.
- [13] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning : with Applications in R*, Springer: New York.
- [14] Khan, Md. H. R. and Shaw, J. E. H. (2011). Multilevel logistic regression analysis applied to binary contraceptive prevalence data. *Journal of Data Science*. **9**, 93–110.
- [15] Lord, F. M. (1983a). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*. **48**, 233–245.
- [16] Lord, F. M. (1983b). Statistical bias in maximum likelihood estimators of item parameters. *Psychometrika*. **48**, 425–435.
- [17] Lord, F. M. (1983c). Maximum likelihood estimation of item response parameters when some responses are omitted. *Psychometrika*. **48**, 477–482.
- [18] Lord, F. M. (1986). Maximum Likelihood and Bayesian Parameter Estimation in Item Response Theory. *Journal of Educational Measurement*. **23**, 157–162.
- [19] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley: New York,2nd Edition.
- [20] Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*,Danish Institute for Educational Research: Copenhagen. (Expanded edition (1980) with foreword and afterword by B.D. Wright, The University of Chicago Press: Chicago.
- [21] Stone C.A. (1992). Recovery of Marginal Maximum Likelihood Estimates in the Two-Parameter Logistic Response Model: An Evaluation of MULTILOG. *Applied Psychological Measurement*. **16**, 1–16.
- [22] Webb, M. C., Wilson, J. R. and Chong, J. (2004). An analysis of quasi-complete binary data with logistic models: Applications to alcohol abuse data. *Journal of Data Science*. **2**, 273–285.

Table 5: Lasso estimates  $\hat{\beta}_L^*(\delta_L)$  and  $\ell(\hat{\beta}_L^*(\delta_L))$  for 35 subsamples of groups

Subset	$\delta_L$	$\hat{\beta}_L^*(\delta_L)$	$\ell(\hat{\beta}_L^*(\delta_L))$
1, 2, 3, 5	0.058	-2.042, 0.000, 0.000, 0.342	-19.524
1, 2, 3, 6	0.077	-2.018, 0.000, 0.000, 0.000	-27.259
1, 2, 3, 7	0.213	-1.836, 0.000, 0.000, 0.111	-20.588
1, 2, 3, 8	0.010	-2.359, 0.344, 0.413, 0.934	-15.479
1, 2, 5, 6	0.128	-1.951, 0.000, 0.000, 0.105	-22.763
1, 2, 5, 7	0.044	-2.118, 0.000, 0.417, 0.650	-13.976
1, 2, 5, 8	0.010	-2.357, 0.342, 0.457, 0.888	-15.096
1, 2, 6, 7	0.228	-1.750, 0.000, 0.000, 0.000	-21.437
1, 2, 6, 8	0.010	-2.357, 0.342, 1.042, 0.303	-17.707
1, 2, 7, 8	0.011	-2.322, 0.273, 1.412, 0.000	-23.117
1, 3, 5, 6	0.121	-1.930, 0.000, 0.000, 0.099	-22.485
1, 3, 5, 7	0.008	-2.401, 0.000, 0.491, 0.968	-13.600
1, 3, 5, 8	0.010	-2.358, 0.389, 0.412, 0.890	-15.549
1, 3, 6, 7	0.008	-2.398,-0.260, 0.483, 0.978	-13.833
1, 3, 6, 8	0.010	-2.517, 0.963, 0.729, 0.000	-29.954
1, 3, 7, 8	0.010	-2.357, 0.206, 0.411, 1.070	-15.822
1, 5, 6, 7	0.008	-2.365,-0.177, 0.593, 0.863	-13.500
1, 5, 6, 8	0.231	-1.552, 0.000, 0.000, 0.000	-19.608
1, 5, 7, 8	0.010	-2.356, 0.205, 0.593, 0.888	-14.107
1, 6, 7, 8	0.010	-2.356, 0.206, 1.042, 0.439	-15.864
2, 3, 5, 6	0.003	-1.809,-0.206, 0.124, 0.316	-19.400
2, 3, 5, 7	0.202	-1.576, 0.000, 0.000, 0.000	-19.708
2, 3, 5, 8	0.233	-1.515, 0.000, 0.000, 0.000	-19.529
2, 3, 6, 7	0.172	-1.637, 0.000, 0.000, 0.000	-20.112
2, 3, 6, 8	0.203	-1.575, 0.000, 0.000, 0.000	-19.703
2, 3, 7, 8	0.008	-2.132, 0.149, 0.186, 1.103	-20.541
2, 5, 6, 7	0.006	-1.747,-0.243, 0.569, 0.279	-15.019
2, 5, 6, 8	0.008	-1.804,-0.179, 1.019, 0.303	-15.727
2, 5, 7, 8	0.081	-1.691, 0.000, 0.555, 0.222	-15.420
2, 6, 7, 8	0.027	-1.906, 0.000, 0.905, 0.195	-15.128
3, 5, 6, 7	0.111	-1.500, 0.000, 0.000, 0.000	-19.522
3, 5, 6, 8	0.134	-1.439, 0.000, 0.000, 0.000	-19.645
3, 5, 7, 8	0.008	-2.448, 0.247, 0.532, 1.010	-14.537
3, 6, 7, 8	0.164	-1.290, 0.000, 0.000, 0.000	-21.002
5, 6, 7, 8	0.075	-1.440, 0.000, 0.537, 0.000	-17.480

Table 6: Ridge estimates  $\widehat{\beta}_R^*(\delta_R)$  and  $\ell\left(\widehat{\beta}_R^*(\delta_R)\right)$  for 35 subsamples of groups

Subset	$\delta_R$	$\widehat{\beta}_R^*(\delta_R)$	$\ell\left(\widehat{\beta}_R^*(\delta_R)\right)$
1, 2, 3, 5	57.993	-1.957,-0.000, 0.000, 0.001	-25.615
1, 2, 3, 6	53.175	-2.018, 0.000, 0.000, 0.0004	-27.253
1, 2, 3, 7	0.031	-2.265, 0.0205, 0.406, 0.807	-14.427
1, 2, 3, 8	0.037	-2.360, 0.409, 0.466, 0.701	-15.248
1, 2, 5, 6	0.131	-2.035, 0.028, 0.000, 0.246	-21.131
1, 2, 5, 7	0.031	-2.211, 0.161, 0.520, 0.703	-13.951
1, 2, 5, 8	0.038	-2.270, 0.299, 0.488, 0.743	-14.574
1, 2, 6, 7	0.256	-1.941,-0.031, 0.280, 0.273	-16.687
1, 2, 6, 8	0.038	-2.323, 0.322, 0.800, 0.385	-16.406
1, 2, 7, 8	0.040	-2.266, 0.259, 0.657, 0.656	-14.114
1, 3, 5, 6	0.163	-1.991, 0.028, 0.016, 0.198	-21.121
1, 3, 5, 7	0.030	-2.300, 0.000, 0.431, 0.824	-13.742
1, 3, 5, 8	0.037	-2.259, 0.458, 0.340, 0.729	-16.143
1, 3, 6, 7	0.029	-2.327,-0.139, 0.475, 0.783	-13.984
1, 3, 6, 8	0.036	-2.435, 0.451, 0.627, 0.450	-17.850
1, 3, 7, 8	0.039	-2.303, 0.302, 0.446, 0.863	-15.020
1, 5, 6, 7	0.029	-2.249,-0.147, 0.545, 0.714	-13.714
1, 5, 6, 8	0.085	-2.136, 0.155, 0.562, 0.488	-14.302
1, 5, 7, 8	0.039	-2.232, 0.258, 0.561, 0.726	-14.363
1, 6, 7, 8	0.039	-2.290, 0.185, 0.868, 0.480	-14.789
2, 3, 5, 6	0.012	-1.805,-0.176,-0.116, 0.273	-19.670
2, 3, 5, 7	0.074	-1.874,-0.085, 0.175, 0.464	-15.649
2, 3, 5, 8	232.959	-1.515, 0.000, 0.000, 0.001	-19.523
2, 3, 6, 7	0.337	-1.724,-0.091, 0.091, 0.173	-17.873
2, 3, 6, 8	0.362	-1.847, 0.105, 0.175, 0.211	-17.256
2, 3, 7, 8	0.031	-2.161, 0.211, 0.326, 0.889	-16.289
2, 5, 6, 7	0.023	-1.683,-0.277, 0.455, 0.253	-15.577
2, 5, 6, 8	0.030	-1.837,-0.097, 0.779, 0.346	-14.604
2, 5, 7, 8	0.030	-1.993, 0.0970, 0.591, 0.537	-14.313
2, 6, 7, 8	0.031	-1.998, 0.054, 0.801, 0.332	-14.393
3, 5, 6, 7	111.376	-1.501,-0.000, 0.000, 0.000	-19.518
3, 5, 6, 8	134.265	-1.439, 0.001, 0.000, 0.001	-19.638
3, 5, 7, 8	0.029	-2.131, 0.338, 0.349, 0.740	-16.132
3, 6, 7, 8	163.871	-1.291, 0.000, 0.000, 0.001	-20.991
5, 6, 7, 8	0.030	-1.500, 0.001, 0.656, 0.000	-16.912

Table 7: Lasso estimates  $\widehat{\beta}_L^*(\delta_L)$  and  $\ell\left(\widehat{\beta}_L^*(\delta_L)\right)$  for 35 subsamples of groups

Subset	$\delta_L$	$\widehat{\beta}_L^*(\delta_L)$	$\ell\left(\widehat{\beta}_L^*(\delta_L)\right)$
1, 2, 3	0.058	-2.120, 0.000, 0.000, 0.000	-30.360
1, 2, 5	0.164	-1.960, 0.000, 0.000, 0.000	-25.723
1, 2, 6	0.006	-2.382, 0.367, 0.000, 0.287	-26.833
1, 2, 7	0.296	-1.763, 0.000, 0.000, 0.000	-21.632
1, 2, 8	0.014	-2.357, 0.343, 1.346, 0.000	-23.613
1, 3, 5	0.156	-1.938, 0.000, 0.000, 0.000	-25.148
1, 3, 6	0.103	-2.019, 0.000, 0.000, 0.000	-27.285
1, 3, 7	0.012	-2.361, 0.000, 0.416, 1.034	-14.215
1, 3, 8	0.015	-2.354, 1.274, 0.409, 0.000	-31.109
1, 5, 6	0.180	-1.859, 0.000, 0.000, 0.000	-23.385
1, 5, 7	0.013	-2.360, 0.0000, 0.555, 0.891	-13.429
1, 5, 8	0.014	-2.356, 0.798, 0.000, 0.888	-22.293
1, 6, 7	0.262	-1.662, 0.000, 0.000, 0.000	-20.340
1, 6, 8	0.015	-2.354, 0.642, 1.039, 0.000	-23.747
1, 7, 8	0.333	-1.398, 0.000, 0.146, 0.000	-18.935
2, 3, 5	0.114	-1.810, 0.000, 0.000, 0.000	-22.424
2, 3, 6	0.060	-1.891, 0.000, 0.000, 0.000	-24.059
2, 3, 7	0.246	-1.612, 0.000, 0.000, 0.000	-19.920
2, 3, 8	0.300	-1.530, 0.000, 0.000, 0.000	-19.551
2, 5, 6	0.004	-1.772,-0.230, 0.0000, 0.290	-18.225
2, 5, 7	0.009	-1.469,-0.520, 0.566, 0.000	-17.955
2, 5, 8	0.247	-1.371, 0.000, 0.000, 0.000	-20.076
2, 6, 7	0.009	-1.179,-0.808, 0.000, 0.275	-25.545
2, 6, 8	0.012	-1.980, 0.000, 1.048, 0.268	-15.196
2, 7, 8	0.281	-1.173, 0.000, 0.000, 0.000	-23.213
3, 5, 6	0.079	-1.709, 0.000, 0.000, 0.000	-20.870
3, 5, 7	0.172	-1.430, 0.000, 0.000, 0.000	-19.682
3, 5, 8	0.125	-1.571, 0.465, 0.000, 0.103	-21.271
3, 6, 7	0.145	-1.511, 0.000, 0.000, 0.000	-19.526
3, 6, 8	0.173	-1.429, 0.000, 0.000, 0.000	-19.688
3, 7, 8	0.201	-1.344, 0.000, 0.000, 0.291	-23.039
5, 6, 7	0.007	-1.469,-0.223, 0.573, 0.000	-16.908
5, 6, 8	0.214	-1.270, 0.000, 0.000, 0.000	-21.318
5, 7, 8	0.007	-1.448, 0.227, 0.573, 0.000	-18.803
6, 7, 8	0.213	-1.072, 0.000, 0.000, 0.000	-25.977



Table 8: Ridge estimates  $\widehat{\beta}_R^*(\delta_R)$  and  $\ell\left(\widehat{\beta}_R^*(\delta_R)\right)$  for 35 subsamples of groups

Subset	$\delta_L$	$\widehat{\beta}_L^*(\delta_L)$	$\ell\left(\widehat{\beta}_L^*(\delta_L)\right)$
1, 2, 3	57.886	-2.120, 0.000, 0.000, 0.000	-30.359
1, 2, 5	102.978	-1.961,-0.000, 0.000, 0.001	-25.711
1, 2, 6	1.806	-2.059, 0.018, 0.000, 0.016	-27.794
1, 2, 7	4.096	-1.788,-0.011, 0.043, 0.043	-20.707
1, 2, 8	4.424	-1.742, 0.038, 0.055, 0.055	-19.624
1, 3, 5	118.239	-1.938, 0.000,-0.000, 0.001	-25.138
1, 3, 6	102.264	-2.019, 0.000, 0.000, 0.000	-27.279
1, 3, 7	3.990	-1.777, 0.000, 0.033, 0.044	-20.601
1, 3, 8	4.750	-1.715, 0.050, 0.036, 0.050	-19.596
1, 5, 6	2.486	-1.880, 0.007, 0.000, 0.027	-23.106
1, 5, 7	4.141	-1.618, 0.000, 0.034, 0.039	-19.139
1, 5, 8	4.558	-1.556, 0.042, 0.042, 0.045	-18.848
1, 6, 7	3.983	-1.698,-0.003, 0.039, 0.037	-19.734
1, 6, 8	4.817	-1.645, 0.038, 0.045, 0.038	-19.153
1, 7, 8	4.615	-1.388, 0.035, 0.057, 0.057	-19.703
2, 3, 5	113.750	-1.810,-0.000,-0.000, 0.000	-22.421
2, 3, 6	59.709	-1.891, 0.000,-0.000, 0.000	-24.058
2, 3, 7	3.098	-1.632,-0.020, 0.020, 0.039	-19.380
2, 3, 8	300.400	-1.531, 0.000, 0.000, 0.001	-19.543
2, 5, 6	1.435	-1.732,-0.013, 0.000, 0.014	-20.972
2, 5, 7	2.926	-1.472,-0.027, 0.029, 0.027	-19.141
2, 5, 8	3.760	-1.406, 0.005, 0.036, 0.031	-19.514
2, 6, 7	3.040	-1.539,-0.032, 0.032, 0.024	-19.086
2, 6, 8	3.800	-1.484, 0.000, 0.041, 0.028	-19.031
2, 7, 8	3.541	-1.222,-0.015, 0.044, 0.044	-22.287
3, 5, 6	79.064	-1.709, 0.000,-0.000, 0.000	-20.869
3, 5, 7	171.980	-1.430, 0.000, 0.000, 0.000	-19.679
3, 5, 8	3.320	-1.385, 0.037, 0.006, 0.031	-20.016
3, 6, 7	144.959	-1.511,-0.000, 0.000, 0.001	-19.521
3, 6, 8	172.388	-1.430, 0.001, 0.000, 0.001	-19.680
3, 7, 8	3.348	-1.188, 0.028, 0.000, 0.043	-23.572
5, 6, 7	2.199	-1.354,-0.018, 0.024, 0.000	-20.018
5, 6, 8	2.959	-1.287, 0.010, 0.032, 0.000	-20.921
5, 7, 8	2.205	-1.013, 0.018, 0.024, 0.000	-28.138
6, 7, 8	2.953	-1.087,-0.009, 0.032, 0.000	-25.341

Table 9: Comparison of MLE, AE, ARE, ARLE ( $\gamma = 4$ ), ARRE ( $\gamma = 4$ ), ARLE ( $\gamma = 3$ ) and ARRE ( $\gamma = 3$ ) for Example 2

Method ( $u$ )	$\hat{\beta}_u^*$	$\ell(\hat{\beta}_u^*)$
MLE	-2.378, -0.068, 0.695, 0.872	-13.269
AE	-2.399, 0.173, 0.665, 0.796	-13.683
ARE( $\lambda = 22$ )	-2.377, 0.143, 0.651, 0.814	-13.277
ARLE( $\gamma = 4$ )	-1.977, 0.074, 0.362, 0.370	-15.221
ARRE( $\gamma = 4$ )	-2.013, 0.094, 0.365, 0.428	-14.948
ARLE( $\gamma = 3$ )	-1.751, 0.067, 0.191, 0.124	-17.506
ARRE( $\gamma = 3$ )	-1.600, 0.005, 0.020, 0.021	-19.382
Lasso(1, 5, 6, 7) $\delta_L = 0.00830$	-2.364, -0.177, 0.593, 0.863	-13.500
Ridge(1, 5, 6, 7) $\delta_R = 0.02845$	-2.249, -0.147, 0.545, 0.714	-13.714
Lasso(1, 5, 7) $\delta_L = 0.01264$	-2.360, 0.000, 0.555, 0.891	-13.429
Ridge(1, 5, 7) $\delta_R = 4.14056$	-1.618, 0.000, 0.034, 0.039	-19.139
Ridge (1,5,8) $\delta_R = 4.55745$	-1.556, 0.042, 0.042, 0.045	-18.848

Table 10: The average values  $\overline{\beta}^*$ ,  $\overline{\beta}_{AE}^*$ ,  $\overline{\beta}_{ARE}^*(\lambda)$ ,  $\bar{\ell}(\hat{\beta}^*)$ ,  $\bar{\ell}(\hat{\beta}_{AE}^*)$ , and  $\bar{\ell}(\hat{\beta}_{ARE}^*(\lambda))$  calculated over 1,000 simulated data for  $\lambda = 6, 21$ , and 25

$\beta^*$ :	$\overline{\beta}^*$	$\overline{\beta}_{AE}^*$	$\overline{\beta}_{ARE}^*(6)$	$\overline{\beta}_{ARE}^*(21)$	$\overline{\beta}_{ARE}^*(25)$
$\alpha$	- 6.084	- 6.183	- 6.084	- 6.084	- 6.084
$\beta_1$	0.025	0.026	0.025	0.025	0.025
$l(\beta^*)$	$\bar{l}(\hat{\beta}^*)$	$\bar{l}(\hat{\beta}_{AE}^*)$	$\bar{l}(\hat{\beta}_{ARE}^*(6))$	$\bar{l}(\hat{\beta}_{ARE}^*(21))$	$\bar{l}(\hat{\beta}_{ARE}^*(25))$
	- 19.714	- 20.458	- 19.725	- 19.714	- 19.716

Table 11: The  $\text{MSE}(\hat{\alpha})$ ,  $\text{MSE}(\hat{\beta}_1)$ ,  $\text{MAD}(\hat{\alpha})$ , and  $\text{MAD}(\hat{\beta}_1)$  for  $\hat{\beta}^*$ ,  $\hat{\beta}_{\text{AE}}^*$ ,  $\hat{\beta}_{\text{ARE}}^*(\lambda)$  when  $\lambda = 6, 21$ , and  $25$

	$\text{MSE}(\hat{\alpha})$	$\text{MSE}(\hat{\beta}_1)$	$\text{MAD}(\hat{\alpha})$	$\text{MAD}(\hat{\beta}_1)$
$\hat{\beta}^*$	0.444	0.000	0.530	0.004
$\hat{\beta}_{\text{AE}}^*$	0.712	0.000	0.654	0.005
$\hat{\beta}_{\text{ARE}}^*(6)$	0.447	0.000	0.532	0.004
$\hat{\beta}_{\text{ARE}}^*(21)$	0.444	0.000	0.530	0.004
$\hat{\beta}_{\text{ARE}}^*(25)$	0.446	0.000	0.531	0.004