On almost unbiased ridge logistic estimator for the logistic regression model

Jibo Wu∗† and Yasin Asar‡

Abstract

Schaefer et al. [15] proposed a ridge logistic estimator in logistic regression model. In this paper a new estimator based on the ridge logistic estimator is introduced in logistic regression model and we call it as almost unbiased ridge logistic estimator. The performance of the new estimator over the ridge logistic estimator and the maximum likelihood estimator in scalar mean squared error criterion is investigated. We also present a numerical example and a simulation study to illustrate the theoretical results.

Keywords: Almost unbiased ridge logistic estimator, Ridge logistic estimator, Logistic regression model, Scalar mean squared error

2000 AMS Classification: 62J07

1. Introduction

In this paper we consider the estimation of Euclidean parameters \( \beta \in \mathbb{R}^p \) in logistic regression model based on the dependent variable \( y_i \) is \( Be(\pi_i) \). The parameters \( \pi_i \) relate to \( \beta \) and \( x_1, x_2, \ldots, x_n \) with the following value:

\[
(1.1) \quad \pi_i = \frac{\exp(x_i\beta)}{1 + \exp(x_i\beta)}, \quad i = 1, 2, \ldots, n
\]

Usually the parameters of the model should be estimated using the maximum likelihood (ML) way by applying the following iterative weighted least square (IWLS) algorithm:

\[
(1.2) \quad \hat{\beta}_{ML} = (X'\hat{W}X)^{-1}X'\hat{W}\hat{Z}
\]

where \( \hat{Z} \) is a vector with ith element equals \( \log(\hat{\pi}_i) + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1 - \hat{\pi}_i)} \) and \( W = \text{diag}(\frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1 - \hat{\pi}_i)}) \).

Since ML estimation does not require any restriction on the characteristics of the independent variables, Maximum likelihood (ML) is the preferred estimation way in logistic regression. However, the ML estimator can be affected seriously by the presence of collinearity. It is known that ML parameter estimates have large variances in cases of multicollinearity. Many methods have proposed to combat this problem in linear regression model, such as the ridge estimator by Hoerl and Kennard [5], Liu estimator by Liu [10].

∗Key Laboratory of Group & Graph Theories and Applications, Chongqing University of Arts and Sciences, Chongqing 402160, China, Email: linfen52@126.com
†Corresponding Author.
‡Department of Statistics, Necmettin Erbakan University Konya, Turkey Email: yasar@konya.edu.tr
Schaefer et al. [15] use the ridge method to overcome the multicollinearity in logistic regression model and propose a ridge logistic estimator. Mansson and Shukur [13], Kibria et al. [12] proposed many methods to estimate the ridge parameter in ridge logistic estimator. Inan and Erdogan [9] proposed a Liu-type logistic estimator to overcome multicollinearity in logistic regression model.

Though the ridge logistic estimator proposed by Schaefer et al. [15] can overcome multicollinearity, however, this estimator has big bias. In this paper, we propose a new estimator which can be used not only overcome multicollinearity, but also can reduce the bias of the ridge estimator. We also discuss the statistical properties of the new estimator.

2. The almost unbiased ridge logistic estimator

The ridge logistic estimator (RLE) in the logistic regression model presented by Schaefer et al. [15] is denoted as follows:

\[
\hat{\beta}_{RLE}(k) = (X'\hat{W}X + kI)^{-1}X'\hat{W}\hat{Z}, \ k > 0
\]

It is easy to obtain that:

\[
\text{Bias}(\hat{\beta}_{RLE}(k)) = E(\hat{\beta}_{RLE}(k)) - \beta
\]

\[
= (X'\hat{W}X + kI)^{-1}X'\hat{W}X\beta - \beta
\]

\[
= [(X'\hat{W}X + kI)^{-1}X'\hat{W}X - I]\beta
\]

\[
= (X'\hat{W}X + kI)^{-1}[X'\hat{W}X - (X'\hat{W}X + kI)]\beta
\]

\[
= -k(X'\hat{W}X + kI)^{-1}\beta
\]

(2.2)

and

\[
\text{Cov}(\hat{\beta}_{RLE}(k)) = (X'\hat{W}X + kI)^{-1}X'\hat{W}X(X'\hat{W}X + kI)^{-1}
\]

(2.3)

In linear regression model, many authors have studied the almost unbiased estimator, such as Kadiyala [11], Akdeniz and Kaciranlar [1] and Xu and Yang [16, 17].

To obtain the almost unbiased ridge logistic estimator, we firstly list the following definitions.

\textbf{Definition 2.1.} [16, 17] Suppose \( \hat{\beta} \) is a biased estimator of parameter vector \( \beta \), and if the bias vector of \( \hat{\beta} \) is given by \( \text{Bias}(\hat{\beta}) = E(\hat{\beta}) - \beta = R\beta \), which shows that \( E(\hat{\beta} - R\beta) = \beta \), then we call the estimator \( \tilde{\beta} = \hat{\beta} - R\hat{\beta} = (I - R)\hat{\beta} \) is the almost unbiased estimator based on the biased estimator \( \hat{\beta} \).
In logistic regression model, we have the new estimator superior to the maximum likelihood (ML) estimator in the scalar mean squared error (MSE) sense. Firstly, we give its definition. Let \(\hat{\beta}\) be an estimator of \(\beta\), then the scalar mean squared error is defined as follows:

\[
\text{Bias}(\hat{\beta}) = E(\hat{\beta} - \beta) = \text{tr}(\text{Cov}(\hat{\beta})) + \text{Bias}(\hat{\beta})
\]

Next, we discuss the superiority of the new estimator in the scalar MSE sense. Let \(\beta_i\) be an estimator of \(\beta_i\), then the scalar mean squared error is defined as follows:

\[
\text{Bias}(\hat{\beta}_i) = E(\hat{\beta}_i - \beta_i) = \text{tr}(\text{Cov}(\hat{\beta}_i)) + \text{Bias}(\hat{\beta}_i)
\]

For the convenience of the following discussions, let \(\alpha = Q'\beta\), \(\Lambda = \text{diag}(\lambda_1, ..., \lambda_p) = Q'(X'WX)Q\), where \(\lambda_1 \geq ..., \lambda_p > 0\) are the ordered eigenvalues of \(X'WX\).

3. The performance of the new estimator

The new estimator is proposed to reduce the bias of the ridge logistic estimator (RLE). So now we compare the new estimator with the RLE.

3.1. Theorem. In logistic regression model, we have \(\|\text{Bias}(\hat{\beta}_{AURLE}(k))\|^2 < \|\text{Bias}(\hat{\beta}_{RLE}(k))\|^2\) for \(k > 0\).

Next, we discuss the superiority of the new estimator in the scalar MSE sense. Firstly, we give its definition. Let \(\hat{\beta}\) be an estimator of \(\beta\), then the scalar mean squared error is defined as follows:

\[
\text{MSE}(\hat{\beta}) = E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) = \text{tr}(\text{Cov}(\hat{\beta})) + \text{Bias}(\hat{\beta})
\]

3.2. Theorem. A sufficient of the new estimator superior to the RLE by the MSE criterion in logistic regression model is

\[
k > 3 - \lambda_i \alpha_i^2 + \sqrt{(3 + \lambda_i \alpha_i^2)^2 + 4 \lambda_i \alpha_i^2}
\]

for all \(i = 1, ..., p\).

3.3. Theorem. The new estimator is superior to the maximum likelihood (ML) estimator in logistic regression model for \(k > 0\) if \(1 - \lambda_i \alpha_i^2 > 0\) for all \(i = 1, ..., p\) and for \(k < \frac{2\lambda_i + \lambda_1 \sqrt{2(1 + \lambda_1 \alpha_1^2)}}{\alpha_i^2 \lambda_{i-1}}\) if \(1 - \lambda_i \alpha_i^2 < 0\) for some \(i\).

4. The selection of ridge parameter \(k\)

In this section, we consider that the ridge parameter which is obtained by using the ridge parameter introduced in the previous section and the ridge parameters proposed by Hoerl and Kennard [5], Hoerl et al. [6], Batah et al. [3], Lawless and Wang [7] and Khurana et al. [4].

The ridge parameter corresponding to Eq. (7.2) is

\[
k_{NEW} = \sum_{i=1}^{p} \frac{\alpha_i^2}{[1 + (1 + \lambda_i \alpha_i^2)^{1/2}]}\]
Table 1: Estimated quadratic bias with $\gamma = 0.9$

<table>
<thead>
<tr>
<th>$k$</th>
<th>NEW</th>
<th>HK</th>
<th>HKB</th>
<th>LW</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLE $\times 10^{-2}$</td>
<td>3.8401</td>
<td>0.5730</td>
<td>2.9811</td>
<td>0.0000</td>
<td>5.0532</td>
</tr>
<tr>
<td>AURLE $\times 10^{-2}$</td>
<td>0.0023</td>
<td>0.0000</td>
<td>0.0014</td>
<td>0.0000</td>
<td>0.0039</td>
</tr>
</tbody>
</table>

Table 2: Estimated quadratic bias with $\gamma = 0.95$

<table>
<thead>
<tr>
<th>$k$</th>
<th>NEW</th>
<th>HK</th>
<th>HKB</th>
<th>LW</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLE</td>
<td>0.1093</td>
<td>0.0020</td>
<td>0.0100</td>
<td>0.0000</td>
<td>0.1627</td>
</tr>
<tr>
<td>AURLE</td>
<td>0.0002</td>
<td>0.0006</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Second, the Hoerl and Kennard [5] ridge parameter is defined as

$$k_{HK} = \frac{\hat{\sigma}^2}{\max_{iML} \hat{\beta}^2}$$

Third, the Hoerl et al. [6] ridge parameter is defined as

$$k_{HKB} = \frac{p \hat{\sigma}^2}{\hat{\beta}_{ML}^p \hat{\beta}_{ML}}$$

Fourth, the Lawless and Wang [7] ridge parameter is defined as

$$k_{LW} = \frac{p \hat{\sigma}^2}{\hat{\beta}_{ML}^p X' WX \hat{\beta}_{ML}}$$

Fifth, the Lindley and Smith [8] ridge parameter is defined as

$$k_{LS} = \frac{(n - p)(p + 2)}{n + 2} \frac{\hat{\sigma}^2}{\hat{\beta}_{ML}^p \hat{\beta}_{ML}}$$

5. Monte Carlo simulation

The main purpose of this article is to compare the MSE properties and bias of the ML, RLE and AURLE when the regressors are highly intercorrelated. Hence, the core factor varied in the design of the experiment is the degree of correlation $\gamma$ between the regressors. Therefore, the following formula which enables us to vary the strength of the correlation is used to generate the explanatory variables:

$$x_{ij} = (1 - \gamma^2)^{1/2} z_{ij} + \gamma z_{ij}, \quad i = 1, ..., n, \quad j = 1, ..., p$$

where $z_{ij}$ are independent standard normal pseudo-random numbers, and $\gamma$ is specified so that the correlation between any two explanatory variables is given by $\gamma^2$.

Four different values of $\gamma$ corresponding to 0.9, 0.95, 0.99 are considered and the sample size is equal to 50.

All simulation results are given in Tables 1-6.

From Tables 1-3, we can see that the new estimator has smaller quadratic bias than the RLE. When we see the estimated MSE of the new estimator and the RLE,
Table 3: Estimated quadratic bias with $\gamma = 0.99$

<table>
<thead>
<tr>
<th>$k$</th>
<th>NEW</th>
<th>HK</th>
<th>HKB</th>
<th>LW</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLE</td>
<td>0.9889</td>
<td>0.2645</td>
<td>0.9696</td>
<td>0.0000</td>
<td>1.3761</td>
</tr>
<tr>
<td>AURLE</td>
<td>0.1230</td>
<td>0.0090</td>
<td>0.0118</td>
<td>0.0000</td>
<td>0.2368</td>
</tr>
</tbody>
</table>

Table 4: Estimated MSE with $\gamma = 0.9$

<table>
<thead>
<tr>
<th>$k$</th>
<th>0</th>
<th>NEW</th>
<th>HK</th>
<th>HKB</th>
<th>LW</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>0.0376</td>
<td>0.0376</td>
<td>0.0376</td>
<td>0.0376</td>
<td>0.0376</td>
<td>0.0376</td>
</tr>
<tr>
<td>RLE</td>
<td>0.0376</td>
<td>0.0709</td>
<td>0.0413</td>
<td>0.0629</td>
<td>0.0376</td>
<td>0.0823</td>
</tr>
<tr>
<td>AURLE</td>
<td>0.0376</td>
<td>0.0374</td>
<td>0.0376</td>
<td>0.0374</td>
<td>0.0376</td>
<td>0.0375</td>
</tr>
</tbody>
</table>

Table 5: Estimated MSE with $\gamma = 0.95$

<table>
<thead>
<tr>
<th>$k$</th>
<th>0</th>
<th>NEW</th>
<th>HK</th>
<th>HKB</th>
<th>LW</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>0.0717</td>
<td>0.0717</td>
<td>0.0717</td>
<td>0.0717</td>
<td>0.0717</td>
<td>0.0717</td>
</tr>
<tr>
<td>RLE</td>
<td>0.0717</td>
<td>0.1645</td>
<td>0.0843</td>
<td>0.1558</td>
<td>0.0717</td>
<td>0.2146</td>
</tr>
<tr>
<td>AURLE</td>
<td>0.0717</td>
<td>0.0711</td>
<td>0.0714</td>
<td>0.0711</td>
<td>0.0717</td>
<td>0.0722</td>
</tr>
</tbody>
</table>

Table 6: Estimated MSE with $\gamma = 0.99$

<table>
<thead>
<tr>
<th>$k$</th>
<th>0</th>
<th>NEW</th>
<th>HK</th>
<th>HKB</th>
<th>LW</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>0.3111</td>
<td>0.3111</td>
<td>0.3111</td>
<td>0.3111</td>
<td>0.3111</td>
<td>0.3111</td>
</tr>
<tr>
<td>RLE</td>
<td>0.3111</td>
<td>1.1128</td>
<td>0.4654</td>
<td>1.0948</td>
<td>0.3111</td>
<td>1.4769</td>
</tr>
<tr>
<td>AURLE</td>
<td>0.3111</td>
<td>0.2943</td>
<td>0.2955</td>
<td>0.3487</td>
<td>0.3111</td>
<td>0.4392</td>
</tr>
</tbody>
</table>

we see that the new estimator is always superior to the RLE. The new estimator is superior to the RLE in the MSE criterion.

From the Tables, we also conclude that the new ridge parameter perform well.

6. Numerical example

In this section, we present a real data application in order to illustrate the benefits of the new estimator AURLE and satisfy the theoretical results. The data set is obtained from the official website of the Statistics Sweden (http://www.scb.se/) and it was also used in Asar and Genc [2] and a similar data set was used in Mansson et al. [4]. There are 271 observations which are the municipalities of Sweden in the data set. We fit a logistic regression model where the followings are the independent variables: the population (x1), the number of unemployed people
Table 7: The estimated theoretical MSEs and coefficients of estimators for different estimators of $k$

<table>
<thead>
<tr>
<th></th>
<th>$k$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>SMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NEW</td>
<td>2.3757</td>
<td>0.2549</td>
<td>1.1752</td>
<td>-2.7641</td>
<td>973.7117</td>
</tr>
<tr>
<td></td>
<td>HK</td>
<td>18.1501</td>
<td>-11.8274</td>
<td>3.6423</td>
<td>-9.0784</td>
<td>1157.0656</td>
</tr>
<tr>
<td>RLE</td>
<td>HKB</td>
<td>4.1873</td>
<td>-0.9625</td>
<td>1.9584</td>
<td>-4.1488</td>
<td>904.7912</td>
</tr>
<tr>
<td></td>
<td>KLW</td>
<td>0.1674</td>
<td>0.1576</td>
<td>0.1342</td>
<td>0.0622</td>
<td>1076.8163</td>
</tr>
<tr>
<td></td>
<td>LS</td>
<td>8.3886</td>
<td>-4.2012</td>
<td>2.8882</td>
<td>-6.0829</td>
<td>826.1210</td>
</tr>
<tr>
<td>AURLE</td>
<td>HKB</td>
<td>3.9688</td>
<td>-0.6164</td>
<td>2.0167</td>
<td>-4.3217</td>
<td>931.4176</td>
</tr>
<tr>
<td></td>
<td>KLW</td>
<td>23.2398</td>
<td>-15.7957</td>
<td>3.8558</td>
<td>-10.4687</td>
<td>1642.0958</td>
</tr>
<tr>
<td></td>
<td>LS</td>
<td>7.1375</td>
<td>-3.1442</td>
<td>2.9676</td>
<td>-5.9501</td>
<td>865.1787</td>
</tr>
<tr>
<td></td>
<td>25.3151</td>
<td>-17.4071</td>
<td>3.6388</td>
<td>-8.0437</td>
<td>936.7677</td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td></td>
<td>25.3151</td>
<td>-17.4071</td>
<td>3.8669</td>
<td>-10.9661</td>
<td>1894.3979</td>
</tr>
</tbody>
</table>

(x2), the number of newly constructed buildings (x3) and the number of bankrupt firms (x4). We consider the net population change as the dependent variable such that it is coded as 1 if there is an increase in the population and 0 vice versa. We computed the bivariate correlations and observed that they are all greater than 0.90. The condition number being a measure of the degree of multicollinearity is computed as 38.3274 showing that there is severe multicollinearity problem with this data.

We provide the estimated theoretical MSE and coefficients of ML, RLE and AURLE for $k_{NEW}, k_{HK}, k_{HKB}, k_{LW}$ and $k_{LS}$ in Table 7.

It is observed from Table 7 that MSE of ML is the largest among all possible situations. The new estimator NEW works well with the estimator AURLE such that AURLE has a less MSE than RLE when NEW is used. Moreover, AURLE has better performance when HKB and LS are used. In Figure 1, we plot the MSE values of RLE and AURLE for changing values of the parameter $k$ between zero and 1. It is seen from Figure 1 that AURLE has less MSE values in this interval. According to Theorem 3.2, for $k > 4.0608$, AURLE should have a less MSE than that of RLE. This result can be seen from Figure 2.

Moreover, we plot the biases of the estimators to illustrate Theorem 3.1 in Figure 3. According to Figure 3, it is observed that the squared bias of AURLE is always less than that of RLE which coincides with Theorem 3.1.

Finally, Theorem 3.3 is also satisfied. Since $1 - \lambda_i \alpha_i^2 > 0$, AURLE has less MSE value than that of ML.

7. Conclusion

In this paper we propose an almost unbiased ridge logistic estimator based on the ridge logistic estimator and we also discuss the properties of the new estimator. The comparison results show that the new estimator has smaller quadratic bias than the RLE, and under certain conditions the new estimator is superior to the ML.
Figure 1: The estimated MSE of RLE and AURLE when $0 < k < 1$

Figure 2: The estimated MSE of RLE and AURLE for satisfying Theorem 3.2

and RLE in the MSE sense.

Acknowledgements
Figure 3: The biases of the estimator RLE and AURLE when $0 < k < 1$

The authors are thankful to the editor and the reviewers for their valuable comments and suggestions which have led to substantial improvement in the paper.

This work was supported by Yong Chuan National Natural Science Foundation of China (No: Ycstc2014nc8001), the National Natural Science Foundation of China (No: 11426054), and Program for Innovation Team Building at Institutions of Higher Education in Chongqing (grant no. KJTD201321).

References


By (2.4), we can compute that:

\[ \text{MSE}(\hat{\beta}_{RLE}(k)) = \text{tr} \{ \text{Cov}(\hat{\beta}_{RLE}(k)) \} + \text{Bias}(\hat{\beta}_{RLE}(k))^\top \text{Bias}(\hat{\beta}_{RLE}(k)) \]

\[ = \text{tr}\{ (X'\hat{W}X + kI)^{-1}X'\hat{W}X(X'\hat{W}X + kI)^{-1} \} + \alpha'k^2(\Lambda + kI)^{-2}\alpha \]

\[ = \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^2} + \sum_{i=1}^{p} \frac{k^2\alpha_i^2}{(\lambda_i + k)^2} \]

\[ = \sum_{i=1}^{p} \frac{\lambda_i + k^2\alpha_i^2}{(\lambda_i + k)^2} \]  

(7.3)

By (2.4), we can compute that:

\[ \text{Cov}(\hat{\beta}_{AURLE}(k)) = [I - k^2(X'\hat{W}X + kI)^{-1}](X'\hat{W}X)^{-1}[I - k^2(X'\hat{W}X + kI)^{-2}] \]

(7.4)
Then we get

\[
MSE(\hat{\beta}_{AURLE}(k)) = \text{tr}\{\text{Cov}(\hat{\beta}_{AURLE}(k))\} + \text{Bias}(\hat{\beta}_{AURLE}(k))'\text{Bias}(\hat{\beta}_{AURLE}(k))
\]

\[
= \text{tr}[[I - k^2(X'WX + kI)^{-2}](X'WX)^{-1}[I - k^2(X'WX + kI)^{-2}]] + \alpha'k^4(\Lambda + kI)^{-4}\alpha
\]

\[
= \sum_{i=1}^{p}(1 - \frac{k^2}{(\lambda_i + k)^2})^2 \frac{1}{\lambda_i} + \sum_{i=1}^{p} \frac{k^4\alpha_i^2}{(\lambda_i + k)^4}
\]

(7.5)  \[= \sum_{i=1}^{p} \frac{(\lambda_i + 2k)^2\lambda_i + k^4\alpha_i^2}{(\lambda_i + k)^4}\]

Now we consider the difference:

\[
\Delta_1 = MSE(\hat{\beta}_{RLE}(k)) - MSE(\hat{\beta}_{AURLE}(k))
\]

\[
= \sum_{i=1}^{p} \frac{\lambda_i + k^2\alpha_i^2}{(\lambda_i + k)^2} - \sum_{i=1}^{p} \frac{(\lambda_i + 2k)^2\lambda_i + k^4\alpha_i^2}{(\lambda_i + k)^4}
\]

(7.6)  \[= \sum_{i=1}^{p} \frac{\lambda_i k[2k^2\alpha_i^2 + (\lambda_i\alpha_i^2 - 3)k - 2\lambda_i]}{(\lambda_i + k)^4}\]

\[\Delta_1\] will be positive for \(k > 0\) if and only if

\[
2k^2\alpha_i^2 + (\lambda_i\alpha_i^2 - 3)k - 2\lambda_i > 0
\]

for all \(i = 1, ..., p\). The expression in (7.7) is a quadratic function of \(k\) which has two distinct roots

\[
k_{1,2} = \frac{3 - \lambda_i\alpha_i^2 \pm \sqrt{(3 + \lambda_i\alpha_i^2)^2 + 4\lambda_i\alpha_i^2}}{4\alpha_i^2}
\]

(7.8) Though the root \(\frac{3 - \lambda_i\alpha_i^2 - \sqrt{(3 + \lambda_i\alpha_i^2)^2 + 4\lambda_i\alpha_i^2}}{4\alpha_i^2}\) is negative. Thus when \(k > 0\) and

\[
k > \frac{3 - \lambda_i\alpha_i^2 + \sqrt{(3 + \lambda_i\alpha_i^2)^2 + 4\lambda_i\alpha_i^2}}{4\alpha_i^2}
\]

for all \(i = 1, ..., p\), the new estimator is superior to the RLE by the MSE criterion in logistic regression model. \(\square\)

3.3 Theorem

Proof. It is easy to obtain that

\[
\text{MSE}(\hat{\beta}_{ML}) = \sum_{i=1}^{p} \frac{1}{\lambda_i}
\]

(7.9)
Now we study the following difference:
\[
\Delta_2 = \text{MSE}(\hat{\beta}_{ML}) - \text{MSE}(\hat{\beta}_{AURLE}(k))
\]
\[
= \sum_{i=1}^{p} \frac{1}{\lambda_i} - \sum_{i=1}^{p} \frac{(\lambda_i + 2k)^2 \lambda_i + k^4 \alpha_i^2}{(\lambda_i + k)^4}
\]
\[
= k^2 \sum_{i=1}^{p} \frac{(1 - \alpha_i^2 \lambda_i)k^2 + 4 \lambda_i k + 2 \lambda_i^2}{\lambda_i (\lambda_i + k)^4}
\]
(7.10)

\[\Delta_2\] will be positive if and only if
\[
(1 - \alpha_i^2 \lambda_i)k^2 + 4 \lambda_i k + 2 \lambda_i^2 > 0
\]
(7.11)

Now we discuss \((1 - \alpha_i^2 \lambda_i)k^2 + 4 \lambda_i k + 2 \lambda_i^2\).

(1) If \(1 - \lambda_i \alpha_i^2 > 0\) for all \(i = 1, ..., p\), then \((1 - \alpha_i^2 \lambda_i)k^2 + 4 \lambda_i k + 2 \lambda_i^2 > 0\).

(2) If \(1 - \lambda_i \alpha_i^2 < 0\) for some \(i = 1, ..., p\), then using the method in Theorem 3.2, we have if \(k < \frac{2 \lambda_i + \lambda_i \sqrt{2(1 + \alpha_i^2 \lambda_i)}}{\alpha_i^2 \lambda_i - 1}\), \((1 - \alpha_i^2 \lambda_i)k^2 + 4 \lambda_i k + 2 \lambda_i^2 > 0\).

This completes the proof of Theorem. \(\square\)