

The Modified Logistic Two-Parameter Estimators

Thayaparan Kayathiri¹, Manickavasagar Kayanan², and Pushpakanthie Wijekoon³

¹*Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka*

²*Department of Physical Science, University of Vavuniya, Vavuniya, Sri Lanka*

³*Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka*

Received: 08 June 2024; Revised: 02 October 2024; Accepted: 04 October 2024

Abstract

In this article, we proposed a new estimator, termed the Modified Logistic Two-Parameter Estimator (MLTPE), and enhanced it by modifying its coefficients, yielding three variants: Modified Logistic Two-Parameter Estimator1 (MLTPE1), Modified Logistic Two-Parameter Estimator2 (MLTPE2), and Modified Logistic Two-Parameter Estimator3 (MLTPE3). These estimators are designed for logistic regression models in the presence of multicollinearity. Theoretically, we demonstrated the superiority of the MLTPE over existing estimators, including the Maximum Likelihood Estimator (MLE), Modified Almost Unbiased Ridge Logistic Estimator (MAURLE), and Logistic Two-Parameter Estimator (LTPE), in terms of mean square error (MSE). The superiority of the estimators is examined using a simulation study and a real-world example. In the simulation study, we varied the degree of correlation and sample size. The findings revealed that the efficacy of the estimators is significantly influenced by these factors. Furthermore, we evaluated the prediction performance of these estimators using balanced accuracy. The results suggested that the new estimators, MLTPE1, MLTPE2, and MLTPE3, outperformed the others slightly in terms of balanced accuracy, with MLTPE2 exhibiting superior performance regarding both scalar mean square error (SMSE) and balanced accuracy. Finally, we validated the simulation study using the myopia dataset, which produced satisfactory results.

Key words: Balanced accuracy; Prediction performance; Simulation and Scalar mean square error.

AMS Subject Classifications: 62J07, 62J12

1. Introduction

The logistic regression model is specified as a Bernoulli distribution since y_i is a binary response variable: $y_i \sim \text{Ber}(\pi_i)$

$$P(y_i) = \pi_i^{y_i} \cdot (1 - \pi_i)^{1-y_i}$$

The general form of the logistic regression model is given by

$$y_i = \pi_i + \epsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

where,

$$\pi_i = \frac{\exp(x'_i \beta)}{1 + \exp(x'_i \beta)}, \quad (2)$$

x_i is the i^{th} row of X , which is an $n \times (p+1)$ data matrix with p predictor variables and β is a $(p+1) \times 1$ vector of coefficients, ϵ_i are independent with mean zero and variance $\pi_i(1 - \pi_i)$ of the response y_i .

The maximum likelihood estimation technique is a commonly used method to estimate the parameter vector (β), and the Maximum Likelihood Estimator (MLE) of β is given by:

$$\hat{\beta}_{MLE} = C^{-1} X' \hat{W} Z. \quad (3)$$

where, $C = X' \hat{W} X$; Z is the column vector with i^{th} element equals $\text{logit}(\hat{\pi}_i) + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1 - \hat{\pi}_i)}$, and, $\hat{W} = \text{diag}[\hat{\pi}_i(1 - \hat{\pi}_i)]$.

Since

$$E(\hat{\beta}_{MLE}) = \beta, \quad (4)$$

$\hat{\beta}_{MLE}$ is an unbiased estimate of β . The covariance matrix of $\hat{\beta}_{MLE}$ is

$$\text{Cov}(\hat{\beta}_{MLE}) = (X' \hat{W} X)^{-1} = C^{-1}. \quad (5)$$

Hence, the mean square error (MSE) of $\hat{\beta}_{MLE}$ is

$$\begin{aligned} \text{MSE}[\hat{\beta}_{MLE}] &= \text{Cov}[\hat{\beta}_{MLE}] + B[\hat{\beta}_{MLE}]B'[\hat{\beta}_{MLE}] \\ &= C^{-1}. \end{aligned} \quad (6)$$

Therefore, the scalar mean square error (SMSE) of $\hat{\beta}_{MLE}$ is,

$$\begin{aligned} \text{SMSE}[\hat{\beta}_{MLE}] &= \text{tr}[\text{MSE}(\hat{\beta}_{MLE})] \\ &= \text{tr}[C^{-1}]. \end{aligned} \quad (7)$$

Since C is a positive definite matrix, there exists an orthogonal matrix P such that $P'CP = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, where $\lambda_1 \geq \lambda_2 \geq \dots, \lambda_p > 0$ are the ordered eigen values of C . Then,

$$\text{SMSE}[\hat{\beta}_{MLE}] = \sum_{j=1}^p \frac{1}{\lambda_j}. \quad (8)$$

The Maximum Likelihood Estimator (MLE) is susceptible to the effects of multicollinearity among explanatory variables. This susceptibility results in an inflated variance of the MLE, thereby rendering the estimates inefficient. To address this issue, numerous

alternative estimators have been proposed in the literature. In this study, we considered two such estimators based on the sample information given below to mitigate the impact of multicollinearity.

The Modified Almost Unbiased Ridge Logistic Estimator (MAURLE) is one of the alternative estimators introduced by Varathan (2022), and it is defined as

$$\hat{\beta}_{MAURLE} = F_k \hat{\beta}_{MLE}, \quad (9)$$

where $F_k = [I - k^2(C + kI)^{-2}][(C + kI)^{-1}C]$, $k > 0$.

Another alternative estimator is the Logistic Two-Parameter Estimator (LTPE) proposed by Huang (2012), which is defined as

$$\hat{\beta}_{LTPE} = L_{k,d} \hat{\beta}_{MLE}, \quad (10)$$

where $L_{k,d} = (C + kI)^{-1}(C + kdI)$, $0 < d < 1$, $k > 0$.

Note that the two alternative estimators that we discussed above are a function of $\hat{\beta}_{MLE}$, and we can present them in general form as,

$$\hat{\beta}_{GLE} = J_{(i)} \hat{\beta}_{MLE}, \quad (11)$$

where $J_{(i)}$ is a positive definite matrix.

$$\hat{\beta}_{GLE} = \begin{cases} \hat{\beta}_{MLE} & \text{if } J_{(i)} = I; \\ \hat{\beta}_{MAURLE} & \text{if } J_{(i)} = F_k; \\ \hat{\beta}_{LTPE} & \text{if } J_{(i)} = L_{k,d}. \end{cases} \quad (12)$$

The asymptotic properties of the general form of estimators are

$$E[\hat{\beta}_{GLE}] = E[J_{(i)} \hat{\beta}_{MLE}] = J_{(i)} \beta. \quad (13)$$

and the dispersion matrix;

$$D[\hat{\beta}_{GLE}] = Cov[J_{(i)} \hat{\beta}_{MLE}] = J_{(i)} C^{-1} J_{(i)}'. \quad (14)$$

The bias vector and the Mean square error matrix (MSE) are

$$B[\hat{\beta}_{GLE}] = E[J_{(i)} \hat{\beta}_{MLE}] - \beta = (J_{(i)} - I) \beta. \quad (15)$$

and

$$\begin{aligned} MSE[\hat{\beta}_{GLE}] &= D[\hat{\beta}_{GLE}] + B[\hat{\beta}_{GLE}] B'[\hat{\beta}_{GLE}] \\ &= J_{(i)} C^{-1} J_{(i)}' + (J_{(i)} - I) \beta \beta' (J_{(i)} - I)', \end{aligned} \quad (16)$$

respectively. The scalar mean square error (SMSE) of the estimators can be obtained as,

$$SMSE[\hat{\beta}_{GLE}] = tr[MSE(\hat{\beta}_{GLE})]. \quad (17)$$

The superiority of the above estimators was compared in terms of mean squared error (MSE) and scalar mean square error (SMSE) in literature (Varathan, 2022; Huang, 2012). While MSE is a common metric in regression analysis, it is unsuitable for comparing logistic regression estimators. Logistic regression deals with binary classification problems, where outcomes are categorical rather than continuous. Predicted values in logistic regression represent probabilities rather than actual outcomes. Therefore, using MSE for comparison would be inappropriate as it does not consider the probabilistic nature of the predictions. Instead, evaluation metrics like balanced accuracy, the area under the receiver operating characteristic curve (ROC AUC), or log loss (cross-entropy loss) are more appropriate for assessing logistic regression model performance.

This study aimed to develop new estimators that enhance predictive performance for logistic regression models in the presence of multicollinearity. Balanced accuracy is employed as the primary metric for comparing the estimators. Balanced accuracy is an effective metric for classification problems with imbalanced datasets, where one class may predominate. It computes the average accuracy across each class, integrating sensitivity and specificity, thus providing a more comprehensive evaluation of model performance.

The paper is organized as follows: In Section 2, we propose new estimators, their properties, and the conditions for their superiority over existing estimators. Section 3 presents the simulation study for the proposed estimators, while Section 4 provides validation of the simulation study. Section 5 presents concluding remarks. Finally, the references and appendices are provided at the end of the paper.

2. Proposed estimators and their superiority

2.1. The proposed estimators

Ahmad (2020) developed the Modified New Two Parameter Estimator (MNTPE), which has shown better performance than the ordinary least squares in the linear regression model. The MNTPE is defined as:

$$\hat{\beta}_{MNTPE} = M_{k,d} \hat{\beta}_{OLSE}, \quad (18)$$

where $M_{k,d} = (X'X + I)^{-1}(X'X + dI)(X'X + kdI)^{-1}X'X$, $0 < d < 1$, $k > 0$.

Now, we propose a new estimator, named the Modified Logistic Two-Parameter Estimator (MLTPE), based on the MNTPE by replacing $\hat{\beta}_{OLSE}$ with $\hat{\beta}_{GLE}$ and $X'X$ with $X'\hat{W}X$ in equation (18) as below:

$$\begin{aligned} \hat{\beta}_{MLTPE} &= F_{k,d} \hat{\beta}_{GLE} \\ &= F_{k,d} J_{(i)} \hat{\beta}_{MLE} \\ &= L_{(i)} \hat{\beta}_{MLE}, \end{aligned} \quad (19)$$

where $F_{k,d} = (X'\hat{W}X + I)^{-1}(X'\hat{W}X + dI)(X'\hat{W}X + kdI)^{-1}X'\hat{W}X$ and $L_{(i)} = F_{k,d}J_{(i)}$. Since $C = X'\hat{W}X$, we have $F_{k,d} = (C + I)^{-1}(C + dI)(C + kdI)^{-1}C$, $0 < d < 1$, $k > 0$.

The expectation, bias, covariance, MSE, and SMSE of MLTPE can be defined as:

$$E(\hat{\beta}_{MLTPE}) = L_{(i)}\beta, \quad (20)$$

$$B(\hat{\beta}_{MLTPE}) = (L_{(i)} - I)\beta, \quad (21)$$

$$Cov(\hat{\beta}_{MLTPE}) = L_{(i)}C^{-1}L'_{(i)}, \quad (22)$$

$$\begin{aligned} MSE(\hat{\beta}_{MLTPE}) &= Cov(\hat{\beta}_{MLTPE}) + B(\hat{\beta}_{MLTPE})B(\hat{\beta}_{MLTPE})' \\ &= L_{(i)}C^{-1}L'_{(i)} + (L_{(i)} - I)\beta\beta'(L_{(i)} - I)'. \end{aligned} \quad (23)$$

and

$$\begin{aligned} SMSE(\hat{\beta}_{MLTPE}) &= tr(MSE(\hat{\beta}_{MLTPE})) \\ &= tr(Cov(\hat{\beta}_{MLTPE})) + B(\hat{\beta}_{MLTPE})'B(\hat{\beta}_{MLTPE}) \\ &= tr(L_{(i)}C^{-1}L'_{(i)}) + (L_{(i)} - I)'\beta'\beta(L_{(i)} - I). \end{aligned} \quad (24)$$

By adopting $\hat{\beta}_{MLE}$, $\hat{\beta}_{MAURLE}$ and $\hat{\beta}_{LTPE}$ in place of $\hat{\beta}_{GLE}$ in equation (19), we propose three new estimators namely, Modified logistic two parameter estimator 1 (MLTPE 1), Modified logistic two parameter estimator 2 (MLTPE 2), and Modified logistic two parameter estimator 3 (MLTPE3), respectively, and defined as,

$$\hat{\beta}_{MLTPE} = \begin{cases} \hat{\beta}_{MLTPE1} & \text{if } L_{(i)} = F_{k,d}I; \\ \hat{\beta}_{MLTPE2} & \text{if } L_{(i)} = F_{k,d}F_k; \\ \hat{\beta}_{MLTPE3} & \text{if } L_{(i)} = F_{k,d}L_{k,d}. \end{cases} \quad (25)$$

2.2. Comparison among the estimators

In this section, we compared the performance of the proposed estimator MLTPE with the existing estimator GLE in terms of the mean square error matrix criterion.

Theorem 2.2.1: When $\lambda_{max}[L_{(i)}C^{-1}L'_{(i)}(J_{(i)}C^{-1}J'_{(i)})^{-1}] < 1$, the MLTPE is superior to GLE if and only if $\delta'_{new}(D_1 + \delta'_G\delta_G)^{-1}\delta_{new} \leq 1$.

Proof

Consider

$$\begin{aligned} &MSE(\hat{\beta}_{GLE}) - MSE(\hat{\beta}_{MLTPE}) \\ &= J_{(i)}C^{-1}J'_{(i)} + (J_{(i)} - I)\beta\beta'(J_{(i)} - I)' - (L_{(i)}C^{-1}L'_{(i)} + (L_{(i)} - I)\beta\beta'(L_{(i)} - I)') \\ &= [J_{(i)}C^{-1}J'_{(i)} - L_{(i)}C^{-1}L'_{(i)}] + [(J_{(i)} - I)\beta\beta'(J_{(i)} - I)' - (L_{(i)} - I)\beta\beta'(L_{(i)} - I)']. \end{aligned} \quad (26)$$

Now consider,

$$\begin{aligned} D(\hat{\beta}_{GLE}) - D(\hat{\beta}_{MLTPE}) &= [J_{(i)}C^{-1}J'_{(i)} - L_{(i)}C^{-1}L'_{(i)}] \\ &= D_1. \end{aligned} \quad (27)$$

Note that, Since C^{-1} is a positive definite matrix, $J_{(i)}C^{-1}J'_{(i)}$ and $L_{(i)}C^{-1}L'_{(i)}$ are positive definite matrices (by Lemma 1 in Appendices A.2). Consequently, by Lemma 2 (in Appendices A.2), if $\lambda_{\max}[L_{(i)}C^{-1}L'_{(i)}(J_{(i)}C^{-1}J'_{(i)})^{-1}] < 1$ then D_1 is a positive definite matrix, where $\lambda_{\max}[L_{(i)}C^{-1}L'_{(i)}(J_{(i)}C^{-1}J'_{(i)})^{-1}]$ is the largest eigenvalue of $[L_{(i)}C^{-1}L'_{(i)}(J_{(i)}C^{-1}J'_{(i)})^{-1}]$. Further by Lemma 3 (in Appendices A.2), $MSE(\hat{\beta}_{GLE}) - MSE(\hat{\beta}_{MLTPE})$ is non negative definite if $\delta'_{new}(D_1 + \delta'_G\delta_G)^{-1}\delta_{new} \leq 1$, where $\delta_{new} = (L_{(i)} - I)\beta$ and $\delta_G = (J_{(i)} - I)\beta$. Hence, the estimator MLTPE is superior to GLE if and only if $\delta'_{new}(D_1 + \delta'_G\delta_G)^{-1}\delta_{new} \leq 1$.

Note that the Theorem 2.2.1 outlines the necessary and sufficient conditions for the superiority of the proposed estimator (MLTPE) over the general existing estimator (GLE). By substituting $J_{(i)}$ with an appropriate matrix, we can derive the following conditions for the superiority of MLTPE over the existing estimators MLE, MAURLE, and LTPE with respect to mean square error (MSE).

- If $J_{(i)} = I$; MLTPE is superior than MLE if $\lambda_{\max}[L_{(i)}C^{-1}L'_{(i)}C^{-1}] < 1$ and $\delta'_{new}[C^{-1} - L_{(i)}C^{-1}L'_{(i)}]\delta_{new} \leq 1$.
- If $J_{(i)} = F_k$; MLTPE is superior than MAURLE if $\lambda_{\max}[L_{(i)}C^{-1}L'_{(i)}(F_kC^{-1}F'_k)^{-1}] < 1$ and $\delta'_{new}[(F_kC^{-1}F'_k - L_{(i)}C^{-1}L'_{(i)} + \beta'(F_k - I)'(F_k - I)\beta)^{-1}]\delta_{new} \leq 1$.
- If $J_{(i)} = L_{k,d}$; MLTPE is superior than LTPE if $\lambda_{\max}[L_{(i)}C^{-1}L'_{(i)}(L_{k,d}C^{-1}L'_{k,d})^{-1}] < 1$ and $\delta'_{new}[(L_{k,d}C^{-1}L'_{k,d} - L_{(i)}C^{-1}L'_{(i)} + \beta'(L_{k,d} - I)'(L_{k,d} - I)\beta)^{-1}]\delta_{new} \leq 1$.

3. Simulation study

Following McDonald and Galarneau (1975), and Alheety *et al.* (2021), we generate the explanatory variables using the following equation.

$$x_{i,j} = \sqrt{(1 - \rho^2)}z_{i,j} + \rho z_{i,p+1} \quad ; i = 1, 2, \dots, n. \quad j = 1, 2, \dots, p. \quad (28)$$

where z_{ij} are pseudo-random numbers from a standard normal distribution, and ρ represents the correlation between any two explanatory variables. Four explanatory variables are generated using equation (28), and we choose $\rho = (0.80, 0.85, 0.90)$. Further, we considered three different sample sizes: 80, 100, and 200. The dependent variable y_i is obtained from the Bernoulli distribution with $\pi_i = \frac{\exp(x'_i\beta)}{1 + \exp(x'_i\beta)}$. The parameter values of $\beta_1, \beta_2, \dots, \beta_p$ are chosen so that $\beta'\beta = 1$. In the simulation study, we implemented a 5-fold cross-validation approach. However, due to the difficulty of achieving balanced accuracy with small sample sizes, we initiated our study with a sample size of 80. The simulation was repeated 1000 times by generating new pseudo-random numbers, and we calculated the SMSE values of the estimators using equation (29).

$$SMSE(\hat{\beta}) = \frac{1}{1000} \sum_{r=1}^{1000} (\hat{\beta}_r - \beta)'(\hat{\beta}_r - \beta), \quad (29)$$

where $\hat{\beta}_r$ denotes any estimator considered in the r^{th} simulation.

The confusion matrix is utilized to determine the balanced accuracy, a widely used measure for solving classification problems. Table 1 details the components of this matrix.

Table 1: Confusion Matrix

		Predicted Outcome	
		P	N
Actual	P	True Positive (TP)	False Negative (FN)
	N	False Positive (FP)	True Negative (TN)

The evaluation metrics are,

$$Sensitivity = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{TN}{(TN + FP)}$$

$$Balanced\ accuracy = \frac{(Sensitivity + Specificity)}{2}$$

The steps for K-fold cross-validation are:

Step 1: Split the data set into K groups.

Step 2: For each individual group $i=1,2,\dots, K$; take the one group as a test data set and the remaining $(K-1)$ groups as a training data set.

Step 3: Estimate the respective estimator $(\hat{\beta}_{k,d})$ with shrinkage parameter (k, d) using training data sets and compute its cross-validated balanced accuracy in predicting the test data set. Use initial values of k and d as 0.01, then increase the values of k and d by a small increment of 0.1. For $k > 0, 0 < d < 1$.

Step 4: Calculate the average cross-validated balanced accuracy.

Step 5: Choose the maximum value of the average cross-validated balanced accuracy of the estimators and corresponding shrinkage parameters.

The SMSE values and cross-validated balanced accuracy values of the estimators MLE, MAURLE, LTPE, MLTPE1, MLTPE2, and MLTPE3 are obtained for different shrinkage parameters (k, d) . We summarize the results of the simulation study in Table 2 - 4,

Table 2: The results of the simulation study when $n=80$

	Estimators	SMSE	Balanced Accuracy
$\rho = 0.80$	MLE	45.036111	0.517190
	MAURLE	0.574975 ($k=0.61$)	0.528781 ($k=0.01$)
	LTPE	0.704999 ($k=1.71, d=0.01$)	0.527550 ($k=1.41, d=0.01$)
	MLTPE1	0.629874 ($k=2.91, d=0.11$)	0.529005 ($k=4.21, d=0.11$)
	MLTPE2	0.566116 ($k=0.31, d=0.21$)	0.529138 ($k=0.61, d=0.11$)
	MLTPE3	0.610834 ($k=0.11, d=0.01$)	0.528975 ($k=1.51, d=0.11$)
$\rho = 0.85$	MLE	58.195287	0.520008
	MAURLE	0.542518 ($k=0.61$)	0.530163 ($k=0.01$)
	LTPE	0.671765 ($k=1.61, d=0.01$)	0.529779 ($k=2.11, d=0.01$)
	MLTPE1	0.594680 ($k=4.91, d=0.11$)	0.530442 ($k=4.31, d=0.01$)
	MLTPE2	0.536631 ($k=0.31, d=0.21$)	0.530542 ($k=0.41, d=0.99$)
	MLTPE3	0.573781 ($k=0.11, d=0.01$)	0.530594 ($k=2.11, d=0.21$)
$\rho = 0.90$	MLE	84.328874	0.520109
	MAURLE	0.522461 ($k=0.61$)	0.530471 ($k=0.01$)
	LTPE	0.643522 ($k=1.51, d=0.01$)	0.530473 ($k=2.01, d=0.01$)
	MLTPE1	0.557923 ($k=4.91, d=0.11$)	0.530814 ($k=1.81, d=0.21$)
	MLTPE2	0.520367 ($k=0.31, d=0.21$)	0.530821 ($k=1.21, d=0.01$)
	MLTPE3	0.547428 ($k=0.11, d=0.01$)	0.530921 ($k=1.71, d=0.11$)

presenting the minimum SMSE with the corresponding shrinkage value and the maximum balanced accuracy with their respective shrinkage values for each estimator. These results are also depicted in Figure 1 - 3 in Appendices A.3.

According to Table 2 - 4, it is observed that the proposed estimators MLTPE1, MLTPE2, and MLTPE3 demonstrate slightly better prediction performance (with the highest balanced accuracy) than the existing estimators MLE, MAURLE, and LTPE. Moreover, we noticed that the new estimator MLTPE2 exhibits better performance (with the lowest SMSE and highest balanced accuracy) than MLE, MAURLE, LTPE, MLTPE1, and MLTPE3 in almost all situations. An increase in the sample size positively impacts the proposed estimators in terms of SMSE but has a negative effect on balanced accuracy. Furthermore, increasing the degrees of correlation positively affects the proposed estimators in terms of both SMSE and balanced accuracy.

Figures 1 - 3 show the results of the simulation study regarding SMSE and balanced accuracy. When comparing the estimators MLE, MAURLE, LTPE, and MLTPE1, we observe that MAURLE has the smallest SMSE; however, the proposed estimator MLTPE1 exhibits slightly higher balanced accuracy. The second new estimator, MLTPE2, demonstrates the best performance in terms of both SMSE and balanced accuracy. Considering all the estimators, their performance in terms of SMSE can be ordered as follows: MLTPE2, MAURLE, MLTPE3, MLTPE1, and LTPE. In terms of balanced accuracy, MLTPE2 shows the best performance in most cases, while MLTPE3 and MLTPE1 also demonstrate comparatively good performance compared to MAURLE and LTPE.

Table 3: The results of the simulation study when $n = 100$

	Estimators	SMSE	Balanced Accuracy
$\rho = 0.80$	MLE	42.868378	0.513073
	MAURLE	0.617397 ($k = 0.61$)	0.519054 ($k = 0.01$)
	LTPE	0.741937 ($k = 1.91, d = 0.01$)	0.518463 ($k = 2.21, d = 0.01$)
	MLTPE1	0.666742 ($k = 3.21, d = 0.11$)	0.519411 ($k = 2.41, d = 0.01$)
	MLTPE2	0.607717 ($k = 0.31, d = 0.11$)	0.519457 ($k = 2.61, d = 0.21$)
	MLTPE3	0.646961 ($k = 0.21, d = 0.01$)	0.519447 ($k = 0.51, d = 0.01$)
$\rho = 0.85$	MLE	54.337142	0.513864
	MAURLE	0.589061 ($k = 0.61$)	0.520987 ($k = 0.01$)
	LTPE	0.712068 ($k = 1.71, d = 0.01$)	0.521142 ($k = 2.31, d = 0.01$)
	MLTPE1	0.635751 ($k = 2.91, d = 0.11$)	0.521016 ($k = 0.51, d = 0.99$)
	MLTPE2	0.582717 ($k = 0.31, d = 0.21$)	0.521657 ($k = 0.41, d = 0.11$)
	MLTPE3	0.617451 ($k = 0.11, d = 0.01$)	0.521297 ($k = 2.71, d = 0.11$)
$\rho = 0.90$	MLE	79.150907	0.515745
	MAURLE	0.556903 ($k = 0.61$)	0.523131 ($k = 0.01$)
	LTPE	0.676103 ($k = 1.61, d = 0.01$)	0.522414 ($k = 1.41, d = 0.01$)
	MLTPE1	0.594124 ($k = 4.91, d = 0.01$)	0.523567 ($k = 4.81, d = 0.11$)
	MLTPE2	0.554349 ($k = 0.31, d = 0.21$)	0.523503 ($k = 1.51, d = 0.01$)
	MLTPE3	0.579814 ($k = 0.11, d = 0.01$)	0.523573 ($k = 0.51, d = 0.11$)

4. Real data application

Myopia data was used to check the performance of the proposed estimators MLTPE1, MLTPE2, and MLTPE3 with existing estimators, such as MLE, MAURLE, and LTPE. This dataset is about a study of myopia taken from Hosmer Jr *et al.* (2013) and also studied by Asar *et al.* (2017). In this data, 618 subjects who were not myopic when they entered the study were followed up for at least five years, and observations were made on 17 parameters. However, following Asar *et al.* (2017), we focused our analysis on four explanatory variables: spherical equivalent refraction (SPHEQ), axial length (AL), anterior chamber depth (ACD), and vitreous chamber depth (VCD). These variables are continuous and measured on the same scale (mm). We limited our analysis to the first 100 observations. The dependent variable indicates whether a subject has myopia (coded as 1) or not (coded as 0).

The condition number, a measure of multicollinearity, is obtained as 10.4647. This result provides evidence of moderate multicollinearity among the explanatory variables in the dataset. We examined the explanatory variables with positive and negative correlations to one another and discovered that the correlation between the variables AL and VCD is very strong (0.9322). The correlation matrix of the explanatory variables, SPHEQ, AL, ACD, and VCD, is presented in Table 6 in Appendices A.1. The corresponding Variance Inflation Factor (VIF) values for these variables are 1.1322, 24.9815, 3.0186, and 22.0122. The VIF measures the extent to which the variance of the coefficient estimates is inflated due to multicollinearity among the predictor variables. As a result, variables SPHEQ and ACD have $1 < \text{VIF} < 5$, indicating moderate collinearity, while variables AL and VCD have

Table 4: The results of the simulation study when $n=200$

	Estimators	SMSE	Balanced Accuracy
$\rho = 0.80$	MLE	38.785726	0.507707
	MAURLE	0.600115 ($k=0.71$)	0.512734 ($k=0.01$)
	LTPE	0.727127 ($k=1.91, d=0.01$)	0.511634 ($k=3.91, d=0.01$)
	MLTPE1	0.648386 ($k=3.41, d=0.11$)	0.512373 ($k=3.51, d=0.11$)
	MLTPE2	0.580982 ($k=0.21, d=0.11$)	0.513346 ($k=0.41, d=0.11$)
	MLTPE3	0.626576 ($k=0.21, d=0.01$)	0.512861 ($k=1.01, d=0.01$)
$\rho = 0.85$	MLE	50.044391	0.508435
	MAURLE	0.565265 ($k=0.61$)	0.513212 ($k=0.01$)
	LTPE	0.693498 ($k=1.81, d=0.01$)	0.513712 ($k=1.51, d=0.01$)
	MLTPE1	0.612667 ($k=3.01, d=0.11$)	0.513775 ($k=1.91, d=0.01$)
	MLTPE2	0.557344 ($k=0.31, d=0.21$)	0.514450 ($k=0.61, d=0.01$)
	MLTPE3	0.593446 ($k=0.21, d=0.01$)	0.513737 ($k=1.11, d=0.91$)
$\rho = 0.90$	MLE	72.871783	0.510181
	MAURLE	0.539267 ($k=0.61$)	0.514214 ($k=0.01$)
	LTPE	0.661072 ($k=1.61, d=0.01$)	0.514231 ($k=0.41, d=0.01$)
	MLTPE1	0.579958 ($k=4.91, d=0.01$)	0.514413 ($k=3.71, d=0.11$)
	MLTPE2	0.535675 ($k=0.31, d=0.21$)	0.515615 ($k=0.41, d=0.01$)
	MLTPE3	0.561105 ($k=0.11, d=0.01$)	0.514474 ($k=1.11, d=0.11$)

VIF values greater than 5, suggesting severe collinearity exists. The results of the real data application are presented in Table (5).

Table 5: The results of the real data application

Estimators	SMSE	Balanced Accuracy
MLE	195.6843	0.746096
MAURLE	28.3958 ($k=0.11$)	0.757534 ($k=0.01$)
LTPE	25.2972 ($k=0.11, d=0.01$)	0.757534 ($k=0.01, d=0.01$)
MLTPE1	25.4710 ($k=0.11, d=0.99$)	0.751978 ($k=0.01, d=0.01$)
MLTPE2	25.1819 ($k=0.01, d=0.61$)	0.763382 ($k=0.01, d=0.01$)
MLTPE3	25.5134 ($k=0.11, d=0.99$)	0.751978 ($k=0.01, d=0.01$)

From Table 5, it is evident that the proposed estimator MLTPE2 outperforms other estimators in terms of both SMSE and balanced accuracy. Note that similar results were also observed in the simulation study.

5. Concluding remarks

Based on the findings of this study, the Modified Logistic Two-Parameter Estimators: MLTPE1, MLTPE2, and MLTPE3 demonstrated slightly better performance in terms of balanced accuracy compared to MLE, MAURLE, and LTPE. Meanwhile, the MLTPE2

significantly outperformed others in terms of SMSE and prediction accuracy. Additionally, estimators performing better in terms of SMSE do not necessarily provide superior prediction results. Therefore, when comparing the performance of logistic regression, metrics that focus on classification accuracy are more appropriate.

Acknowledgements

We express our sincere gratitude to the editorial team, especially the Chair Editor, for their valuable insights and constructive feedback, and to the anonymous reviewers for their thorough evaluation and thoughtful comments, all of which greatly improved the quality of this manuscript.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

- Ahmad, S. (2020). Another proposal about the new two-parameter estimator for linear regression model with correlated regressors. *Communications in Statistics - Simulation and Computation*, **51**, 1–19.
- Alheety, M., Månsson, K., and Golam Kibria, B. (2021). A new kind of stochastic restricted biased estimator for logistic regression model. *Journal of Applied Statistics*, **48**, 1559–1578.
- Asar, Y., Arashi, M., and Wu, J. (2017). Restricted ridge estimator in the logistic regression model. *Communications in Statistics-Simulation and Computation*, **46**, 6538–6544.
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.
- Huang, J. (2012). A simulation research on a biased estimator in logistic regression model. In *International Symposium on Intelligence Computation and Applications*, pages 389–395. Springer.
- McDonald, G. C. and Galarneau, D. I. (1975). A monte carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, **70**, 407–416.
- Rao, C. R., Shalabh, Toutenburg, H., and Heumann, C. (2008). *Linear Model and Generalizations*. Springer, Berlin.
- Rao, C. R. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives, Second Edition*. Springer, New York.
- Trenkler, G. and Toutenburg, H. (1990). Mean squared error matrix comparisons between biased estimators—an overview of recent results. *Statistical Papers*, **31**, 165–179.
- Varathan, N. (2022). An improved ridge type estimator for logistic regression. *Statistics in Transition. New Series*, **23**, 113–126.

A. Appendices

A.1. Correlation matrix

Table 6: The correlation matrix of the explanatory variables of real data

Variables	SPHEQ	AL	ACD	VCD
SPHEQ	1.0000	-0.2865	-0.2143	-0.2643
AL	-0.2865	1.0000	0.4766	0.9322
ACD	-0.2143	0.4766	1.0000	0.2007
VCD	-0.2643	0.9322	0.2007	1.0000

A.2. Lemmas

Lemma 1: (Rao and Toutenburg, 1995) Let A be positive definite and B be a regular matrix, then $B'AB > 0$.

Lemma 2: (Rao *et al.*, 2008) Let the two $n \times n$ matrices $M > 0$, $N \geq 0$, then $M > N$ if and only if $\lambda_{\max}(NM^{-1}) < 1$.

Lemma 3: (Trenkler and Toutenburg, 1990) Let $\hat{\beta}_j = A_j y, j = 1, 2$ be two competing homogenous linear estimators of β . Suppose that $D = Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2) > 0$; where $Cov(\hat{\beta}_j), j = 1, 2$ denotes the covaraince matrix of $\hat{\beta}_j$. Then $\Delta(\hat{\beta}_1, \hat{\beta}_2) = MSEM(\hat{\beta}_1) - MSEM(\hat{\beta}_2) \geq 0$ if and only if $d_2'(D + d_1'd_1)d_2 \leq 1$, where $MSEM(\hat{\beta}_j), d_j; j = 1, 2$ denote the Mean Square Error Matrix and bias vector of $\hat{\beta}_j$, respectively.

A.3. Figures

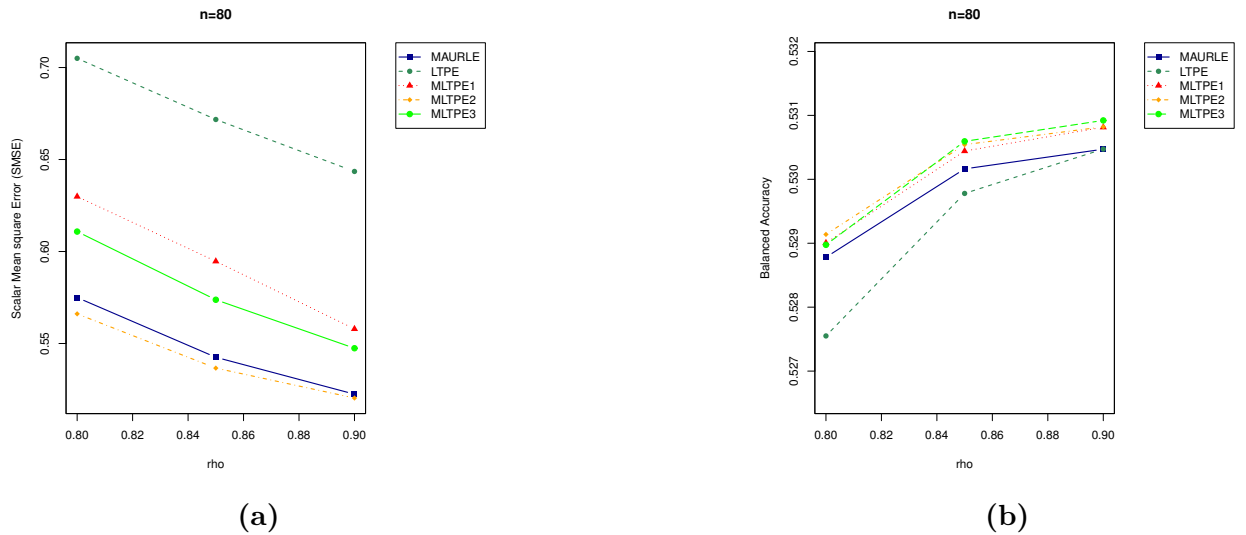


Figure 1: The result of the simulation study when $n = 80$

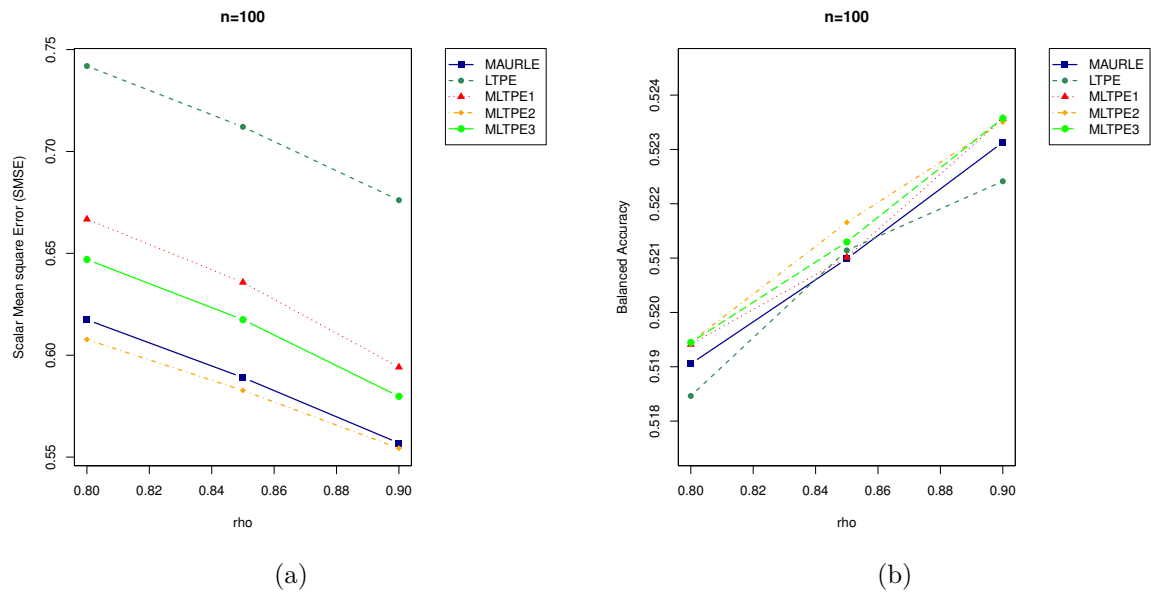


Figure 2: The result of the simulation study when $n = 100$

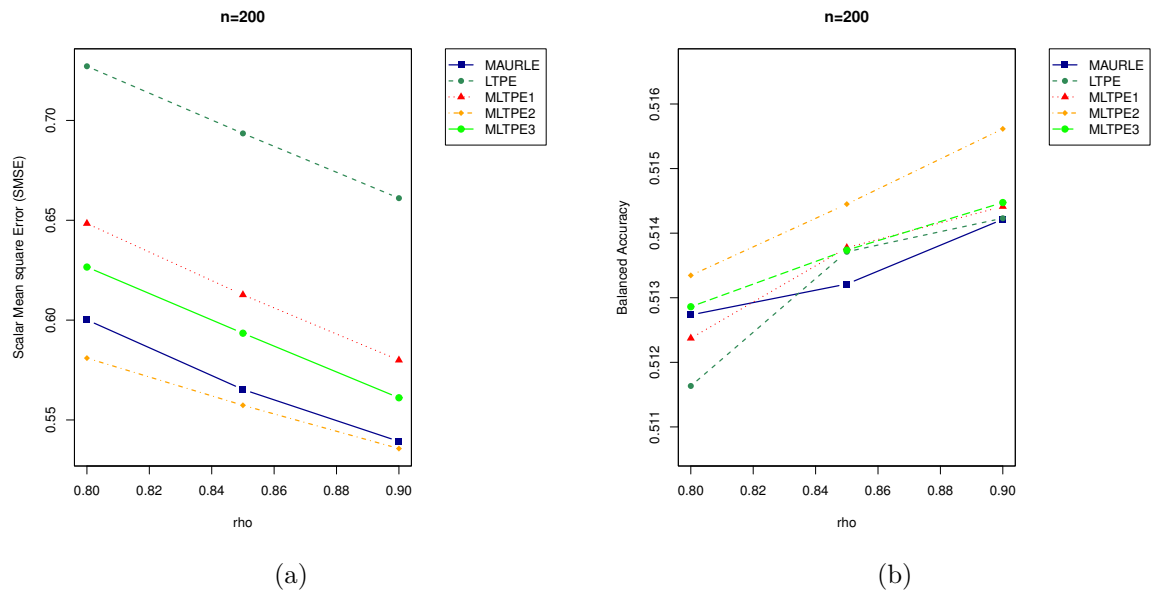


Figure 3: The result of the simulation study when $n = 200$