

# Weighted Maximum Likelihood Technique for Logistic Regression

Idriss Abdelmajid Idriss<sup>1</sup>, Weihu Cheng<sup>1</sup>, Yemane Hailu Fissuh<sup>2</sup>

<sup>1</sup>Department of Statistics, Faculty of Science, Beijing University of Technology, Beijing, China

<sup>2</sup>Department of Statistics, Aksum University, Tigray, Ethiopia

Email: Idrissabdo59@gmail.com, Chengweihu@bjut.edu.cn, yemaoroba@gmail.com

**How to cite this paper:** Idriss, I.A., Cheng, W.H. and Fissuh, H.Y. (2023) Weighted Maximum Likelihood Technique for Logistic Regression. *Open Journal of Statistics*, 13, 803-821.

<https://doi.org/10.4236/ojs.2023.136041>

**Received:** October 8, 2023

**Accepted:** December 5, 2023

**Published:** December 8, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In this paper, a weighted maximum likelihood technique (WMLT) for the logistic regression model is presented. This method depended on a weight function that is continuously adaptable using Mahalanobis distances for predictor variables. Under the model, the asymptotic consistency of the suggested estimator is demonstrated and properties of finite-sample are also investigated via simulation. In simulation studies and real data sets, it is observed that the newly proposed technique demonstrated the greatest performance among all estimators compared.

## Keywords

Logistic Regression, Clean Model, Robust Estimation, Contaminated Model, Weighted Maximum Likelihood Technique

---

## 1. Introduction

Logistic regression is a method used in statistics for modeling the relationship between a binary dependent variable and one or more explanatory variables. It is usually utilized in various fields. Often, the estimation method for parameters of a statistical model uses the maximum likelihood estimator by optimizing the likelihood function. In logistic regression models, MLE is utilized to estimate the coefficients of the predictor variables that most effectively fit the data. Unfortunately, this approach is not robust to unusual data findings. Several robust estimators have been presented as an alternative to MLE to address this issue. The MLE in logistic regression was shown to be very sensitive to outlying data by [1], who also devised a diagnostic assessment of outlying observations; see also [2]. For binary regression, [3] studied several M-estimators, their estimations depending on leverage down weight, a Mallows-type estimator. [4] created a robust

estimate for the logistic regression model based on a modified median estimator, and they also investigated a Wald-type test statistic. [5] created incredibly robust projection estimators for the GLM, but their calculation is quite difficult. [6] proposed a quasi-likelihood estimator by substituting the least absolute deviation estimator (L1 norm) for the least squares estimator (L2 norm) in the definition of quasi-likelihood. [7] suggested resilient estimators and testing techniques for Poisson and binomial models depending on the idea of a quasi-likelihood estimator created by [8]. The breakdown of the MLE in the logistic model was investigated in [9], whereas [10] offered a method for the logistic regression model. [11] created a new robust technique for Poisson regression. [12] provided reliable estimators for generalized linear models; the fundamental concept is to change the response via a variance stabilizing transformation before using an estimation. [13] indicated that was a very consistent and reliable estimator. [14] provides a reliable and efficient approach for computing the M-estimator described in [13]. [15] presented a quick technique for the generalized linear model based on breakdown points of the trimmed likelihood. Fisher-consistent estimators are another kind of robust estimator, as introduced in [13]. [16] investigated a resistant robust estimator and this estimation relied on the misclassification model. [17] presented a new family of robust methods for logistic regression. [18] compared minimal distance approaches to more robust methods discovered a unique weighted likelihood and used it in Poisson and binary regression. The ideally bounded scoring functions described by [19] for linear models were applied to the logistic model in [20].

All of these estimators possess significant disparities in their resistance to outliers and effectiveness within the framework of the model. In this paper, we have conducted a comprehensive investigation into the behavior of some of these estimators, both in terms of their asymptotic properties and their behavior in finite samples. Our findings indicate that the Mallows-type estimator proposed by [3] is very robust to outlier contamination but inefficient under the model, while Schweppe-type estimators proposed by [2] are very efficient under the model but show a poor outlier resistance. In this paper, we propose an estimator that can be as robust as Mallows-estimators under contamination but is much more efficient under the model, this is achieved by an adaptive continuous weight. This continuous weighted maximum likelihood estimator depends on the annoyance parameter estimator as a result of Kolmogorov-Smirnov statistics. The maximum likelihood estimators and a logistic regression model are covered in Section 2. Robust Methods for Logistic Regression Model discussed in section 3. We propose a robust technique for logistic regression in section 4. Section 5 displays the findings of the Monte Carlo simulation research and real data analysis. Section 6 contains the conclusions.

## 2. Logistic Regression Model and ML Estimator

The logistic regression model is a popular technique used to examine the rela-

relationship between a categorical variable and one or more predictor variables. It is often used in binary classification issues when there are only two potential outcomes for the dependent variable (e.g., true or false, yes or no, etc.). The logistic function is the foundation of the logistic regression model, which transforms a continuous input variable into a probability value between 0 and 1. Assume a random sample of observations  $(x_1, y_1), \dots, (x_n, y_n)$ , where  $X_i$  represents a  $p$  predictor variables and  $y_i \in (0, 1)$  is a binary variable and suppose the probability of positive response  $\mu_i = p(y_i = 1 | x_i)$  is associated with the covariates through the relation  $g(\mu) = X\beta$ , such that  $g^{-1}(X\beta)$  is known as the logit link function converts the covariate values in the range  $(0, 1)$ .

Using the logit link function The multiple logistic regression model may be created by:

$$p(Y = 1 | X = x_i) = F(x_i^T \beta) = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}, i = 1, \dots, n, \quad (1)$$

where  $X = (1, x_1, \dots, x_p)$  are the predictor variables values and  $\beta^T = (\beta_0, \beta_1, \dots, \beta_p)$  represents an unknown parameter vector. We may characterize the binary regression model as follows:

$$\eta_i = x_i^T \beta,$$

where  $\eta_i$  is a linear predictor that is sometimes referred to as a transformation function, where  $\eta_i = \ln\left(\frac{\mu_i}{1 - \mu_i}\right)$ . The MLE is a method for parameter estimation

of a statistical model by maximization of the likelihood function. The MLE assumes that the data are generated by a specific probability distribution (in this case, the logistic distribution) and finds the parameter values that make it most likely for that distribution to have generated those data. The MLE is often used in logistic regression because it provides unbiased estimates of model parameters and has good statistical properties. Assume that the dependent variables  $y_i$  have a Bernoulli distribution and we can derive the probability distribution for the  $i^{\text{th}}$  observation as follows:

$$f(y_i) = \mu_i^{y_i} (1 - \mu_i)^{1 - y_i}, i = 1, 2, \dots, n,$$

as well as each observation  $y_i$  takes the value 1 with probability  $\mu_i$  or the value 0 with probability  $(1 - \mu_i)$ . The probability function is defined as follows

$$\ln(\beta; y_1, y_2, \dots, y_n) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \mu_i^{y_i} (1 - \mu_i)^{1 - y_i}, \quad (2)$$

Then we compute the log-likelihood of the preceding formula:

$$\ln(\beta; y_1, y_2, \dots, y_n) = \ln \prod_{i=1}^n f(y_i) = \sum_{i=1}^n \left[ y_i \ln\left(\frac{\mu_i}{1 - \mu_i}\right) \right] + \sum_{i=1}^n \ln(1 - \mu_i),$$

where  $\eta_i = \ln\left[\frac{\mu_i}{1 - \mu_i}\right] = x_i^T \beta$  and  $1 - \mu_i = \left[1 + \exp(x_i^T \beta)\right]^{-1}$ . So, we can express the log-likelihood as:

$$\begin{aligned}\ln l(\beta; Y) &= \sum_{i=1}^n \ln y_i x_i^T \beta - \sum_{i=1}^n \ln [1 + \exp(x_i^T \beta)] \\ &= \beta^T XY - \sum_{i=1}^n \ln [1 + \exp(x_i^T \beta)].\end{aligned}\quad (3)$$

In experimental design, we execute repeated observations at each level of independent variables ( $x$ ). Let  $\eta_i$  be the trials number at each predictor level and  $y_i$  be the number of 1's observed at the  $i^{\text{th}}$  observations with  $n = n_1, n_2, \dots, n_m$ . Therefore, we may express the log-likelihood as:

$$\ln l(\beta; Y) = \beta^T XY - \sum_{i=1}^m n_i \ln [1 + \exp(x_i^T \beta)], \quad (4)$$

nevertheless, it is possible to maximize the likelihood function by differentiating it with respect to  $\beta$ :

$$\frac{\partial \ln l(\beta; Y)}{\partial \beta} = X^T Y - \sum_{i=1}^m \left[ \frac{n_i}{1 + e^{x_i^T \beta}} \right] e^{x_i^T \beta} X_i,$$

where,  $e^{x_i^T} / (1 + e^{x_i^T \beta}) = 1 / (1 + e^{-x_i^T \beta}) = \mu_i$ , we have

$$\frac{\partial \ln(\beta; Y)}{\partial \beta} = X^T Y - \sum_i n_i \mu_i X_i,$$

where  $n_i \mu_i$  denotes the average of the binomial variable, the preceding formula may be written in matrix form as  $X^T (Y - \mu)$ , where

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix}.$$

Hence, the MLE is normally computed by resolving the scoring equation:

$$X^T (Y - \mu) = 0. \quad (5)$$

Since Equation (5) is a nonlinear function of  $\beta$ , the iteratively weighted least squares (IWLS) technique may be used to find a solution.

### 3. Robust Methods for Logistic Regression Model

In logistic regression models, robust estimators are statistical methods for estimating parameters that are less sensitive to outliers and influential observations. These techniques are designed to provide reliable estimates of the regression coefficients even when the data contain extreme values or other anomalies that can distort the results. There are several robust estimators that can be used in logistic regression. [13] suggested robust estimators via the management of deviations to get theoretically unbiased estimators; however, an extra bias-correction component must be introduced, which makes the calculation of their estimator very complex and the estimator itself not straightforward. [2] developed Mallows-type estimators by separately manipulating the variables and residuals in the estimation equation. In the instance of logistic regression, they were catego-

rized as solutions of

$$\sum_{i=1}^n w(x_i; \hat{\eta}) \pi_b \left( y_i - \pi(x_i^T \hat{\beta}_i) - c\left(\pi(x_i^T \hat{\beta}_i), b\right) \right) x_i = 0, \quad (6)$$

where  $\hat{\eta}$  represents the annoyance parameters (location and scatter estimate of covariates), and  $\Phi_b$  is often considered as Huber's function

$\pi_b(t) = (-b) \vee (t \wedge b)$ ,  $c(t, b)$  is the bias-correction function expressed as

$$c(t, b) = \begin{cases} b\pi / 1 - \pi(t) - \pi(t) & \text{if } t < 0, b < 1 - \pi(t) \\ 1 - \pi(t) - b1 - \pi(t) / \pi(t) & \text{if } t > 0, b < \pi(t) \\ 0 & \text{otherwise} \end{cases}$$

the weights  $w(x_i; \hat{\eta})$  often depend only on continuous variables. Suppose we write  $x_i^T = (u_i^T, z_i^T)$ , where  $u_i \in R^{p-q}$  are qualitative variables and  $z_i \in R^q$  are the continuous variables, thereafter, the weights usually take the form

$w(x_i; \hat{\eta}) = w\left(\frac{(z_i - \hat{\mu})^T \hat{\Sigma}^{-1} (z_i - \hat{\mu})}{t}\right)$ , with  $w: R_{+1} \rightarrow R_{+1}$  a function that does not increase,  $\hat{\mu}$  is the robust estimator of the location of the  $z_i$ ,  $\hat{\Sigma}$  is the robust estimator of scatter of the  $z_i$  and  $t$  is the threshold value (usually  $t = \chi_{q,1-\alpha}$  for some  $\alpha = (0,1)$ ). The initial robust estimator of location and scale for predictor variables  $\hat{\mu}$  and  $\hat{\Sigma}$  can be calculated utilizing the minimum covariance determinant methods. MCD was one of the first multivariate location and scatter estimators that were both affine equivariant and relatively robust. It discovers the  $h(> n/2)$  observations  $x_i$  whose matrix of traditional covariance

$$v = \frac{1}{h} \sum_i (x_i - t)(x_i - t)^T$$

has the smallest possible determinant, where  $t = x$  the average of those  $h$  points. As observed, the residual weight  $\phi_b$  and covariate weight  $w$  are independent; moreover, it will decrease the effectiveness of the resultant estimators due to the fact that the estimation equation will downweight well-fitted observations with extreme variables. [21] robustly presented a family of resilient adaptive weighted maximum likelihood techniques for logistic regression models. These adaptive weights are dependent on adaptive cut-off thresholds to regulate observations with extreme covariables. They demonstrated that estimators based on adaptive thresholds are more efficient than estimators based on non-adaptive thresholds for clean models and have equivalent resilience for polluted models.

The lack of dependence between the weights assigned to covariates and the weights assigned to deviances in Equation (6) is the underlying cause for the generally lower efficiency of Mallows-type estimators compared to Schweppe-type estimators. This occurs due to the downweighting of observations with extreme covariates, even if they exhibit good fit. It is evident that enhancing the efficiency of Mallows-type estimators can be achieved by reducing the thresholding proportions, although this may compromise the robustness of the estimator. In order to simultaneously achieve both high efficiency and high robustness, it becomes necessary to employ adaptive thresholds, as detailed in the next section.

### 4. Weighted Maximum Likelihood Technique (WMLT)

In this section, we build a novel class of continuous weighted maximum likelihood estimators based on the annoyance parameter estimator as a function of the Kolmogorov-Smirnov statistics. We'll refer to these estimators as WMLT (weighted maximum likelihood technique). First build two estimators  $\hat{\mu}^{(0)}$  and  $\hat{\Sigma}^{(0)}$  that are the initial estimates of location and scatter of the predictor variables  $z_i$ 's, thereafter, calculate the Mahalanobis distances squared of  $z_i$ 's that is characterized by  $m^2 = (z_i - \hat{\mu}^{(0)})^T (\hat{\Sigma}^{(0)})^{-1} (z_i - \hat{\mu}^{(0)})$ . Furthermore, the empirical distribution function of  $m_i^2$  may be expressed as:

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n I(|m_i^2| \leq t),$$

when  $z_i$ 's have a normal distribution,  $F_n$  converge to  $F_{\chi_q^2}$  ( $\chi_q^2$  distribution function). Then we can estimate the outliers proportion in covariates by [21]

$$\alpha_n = \sup_{t \geq F_{\chi_q^2}^{-1}(1-\delta)} \left\{ F_{\chi_q^2}(t) - F_n(t) \right\}_+ = \max_{i \geq i_0} \left\{ F_{\chi_q^2}(m_i^2) - \frac{i-1}{n} \right\}_+,$$

where  $\{\cdot\}_+$  represents the positive part,  $\delta$  measures the length of the tail ( $\delta = 0.25$ ) is an acceptable option, and  $i_0 = \min \left\{ i : m_i^2 \geq F_{\chi_q^2}^{-1}(1-\delta) \right\}$ . When  $\left| F_{\chi_q^2}(t) - F_n(t) \right|$  is large for a large  $t$ , it indicates that the sample contains outliers. Hence, an adaptive threshold may be described as

$$t_n = F_n^{-1}(1 - \alpha_n) = m_{(n - \lfloor n\alpha_n \rfloor)}^2.$$

[21] proposed the adaptive threshold estimators. Specifically, they developed an estimator of the Mallows-type estimator with weights  $w(x_i; \hat{\eta}) = w(m_i^2 / t_n)$  which are basically weighted maximum likelihood estimators. The proposed weight function may be defined as

$$w(x_i; \alpha_n) = m(\alpha_n m_i^2),$$

we assume  $m$  is a completely non-increasing continuous mapping from  $\mathbb{R}^+$  to  $(0,1)$  such that  $m(0) = 1$ ,  $\sup_{x>0} \{xm(x)\} < \infty$ , and the first derivative is bounded with  $m^{(1)}(0) = 0$ . Determine an objective function

$$\psi_{\beta, \eta}(x, y) = w_\eta(x) \phi_\beta(x, y), \quad x, \beta \in \mathbb{R}^p, \quad y \in (0,1), \tag{7}$$

where  $\eta = (\alpha, \mu, \Sigma)$  is the location, scale, and goodness of fit of the explanatory variables respectively, an adaptive weight function

$w_\eta(x) = m(\alpha(z - \mu)^T \Sigma^{-1} (z - \mu))$  and  $\phi_\beta(x, y) = (y - (x^T \beta))x$ , with  $\mu \in \mathbb{R}^a$ ,  $\Sigma$  is  $q \times q$  real matrix and  $x^T = (u^T, z^T)$ . Finally, we define our adaptive estimator  $\hat{\beta}_n$  of  $\beta$  as the solution to the estimating equation

$$\sum_{i=1}^n \psi_{\beta, \hat{\eta}}(x_i, y_i) = 0,$$

where  $\hat{\eta}$  is a consistent estimator of  $\eta = (\alpha, \mu, \Sigma)$ .

If  $z_i$  are normally distributed, then  $t_n = \infty$  and the reweighted estimators are asymptotically equivalent to the sample mean and the sample covariance matrix, and therefore fully efficient. This efficient carries over to the adaptive Mallows-type estimators, as shown in the next subsection.

#### 4.1. Asymptotic of Proposed Method

The estimating equation (6) can be written as  $\sum_{i=1}^n \Psi(x_i, y_i; \hat{\beta}; \hat{\eta}) = 0$ , with  $\Psi(x, y; \beta; \eta) = w(x; \eta) \psi_b(y - F(x^T \beta) - c(x^T \beta, b))x$ . Under appropriate regularity conditions, the classical asymptotic of M-estimators hold, see [22]. Let  $\beta_0$  be the model parameter and  $E_0$  denote expectation under the model; define  $M_0(\beta) = E_0\{\Psi(x, y; \beta; \eta_0)\}$ , with  $\eta_0 = p \lim \hat{\eta}$ . If  $C_0 = DM_0(\beta)|_{\beta=\beta_0}$ , where  $D$  denote the differential and  $A_0 = E_0\{\Psi(x, y; \beta_0; \eta_0)\Psi(x, y; \beta_0; \eta_0)^T\}$ , then

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N_p(0, C_0^{-1} A_0 C_0^{-1}).$$

This result is valid for non-adaptive and adaptive weights alike, as long as  $\hat{\eta}$  converges to  $\eta_0$  in probability. For the weighted maximum likelihood estimator given by  $\psi_b(u) = u$ ,  $A_0$  and  $C_0$  have simple expressions

$$C_0 = -E\{w(x; \eta_0) F'(x^T \beta_0) x x^T\}, \quad (8)$$

$$A_0 = E\{w^2(x; \eta_0) F(x^T \beta_0)(1 - F(x^T \beta_0)) x x^T\}. \quad (9)$$

Using the asymptotic normality of  $\hat{\beta}$  it is possible to construct confidence ellipsoids for  $\beta_0$ . First, we estimate the matrices (9) and (8) with

$$\hat{C} = -\frac{1}{n} \sum_{i=1}^n w(x_i; \hat{\eta}) F'(x_i^T \hat{\beta}) x_i x_i^T,$$

$$\hat{A} = \frac{1}{n} \sum_{i=1}^n w(x_i; \hat{\beta}) F(x_i^T \hat{\beta})(1 - F(x_i^T \hat{\beta})) x_i x_i^T.$$

Then the estimated asymptotic variance of  $\sqrt{n}(\hat{\beta} - \beta_0)$  is  $\hat{V} = \hat{C}^{-1} \hat{A} \hat{C}^{-1}$ . The asymptotic confidence ellipsoid of level  $(1 - \alpha)$  for  $\beta_0$  is given

$\xi(\beta_0) = \{\beta \in R^p : n(\hat{\beta} - \beta)^T \hat{V}^{-1}(\hat{\beta} - \beta) \leq \chi_{p, 1-\alpha}^2\}$ . This can be generalized to linear transformations of  $\beta_0$ .

#### 4.2. Asymptotic Properties of WMLT

This subsection focuses on the asymptotic features of the suggested estimator  $\hat{\beta}_n$  described in the preceding section. We will show that the estimator is asymptotically consistent based on some general assumptions about the moments of predictor variables. Suppose that  $\beta_0$ ,  $\mu_0$  and  $\Sigma_0$  are the actual values of  $\beta$ ,  $\mu$  and  $\Sigma$  respectively, and the independent sample  $(x_1, y_1), \dots, (x_n, y_n)$  follows the logistic model  $P(y_i = 1) = \mu_i$ ,  $i = 1, \dots, n$ . We define functions

$$\mathbb{P}_n \psi_\beta = \Psi_n(\beta) = \frac{1}{n} \sum_{i=1}^n \psi_\beta \hat{\eta}(x_i, y_i)$$

and

$$\Psi(\beta) = P_{\psi_{\beta,0}} = P_{\phi_\beta},$$

where  $\psi_{\beta,0}$  is the calculated by Equation (7) with  $\eta$  replaced with  $\eta_0 = (0, \mu_0, \Sigma_0)$  and  $p$  indicates the joint probability distribution of the  $(x, y)$ 's. Theorem 4.2 establishes that  $\hat{\beta}_n$  is consistent, which makes use of the conclusions of Lemma 1, 2, and 3 mentioned below. To prove the Lemmas and the Theorems, the following assumptions must be made:

B1:  $\hat{\mu} \rightarrow^p \mu_0$  and  $\hat{\Sigma} \rightarrow^p \Sigma_0$ .

B2:  $E(xx^T)$  is nonsingular.

B3:  $E_{G_0}(\|x\|^4) < \infty$ .

B4: The  $m(x)$  is a continuous weight function and has a first derivative that is bounded,  $m(0) = 1$  and  $m'(0) = 0$ .

B1 is met for the vast majority of well-known initial robust estimators, such as the Minimum Covariance Determinant used in the simulation experiments. In the following lemmas and theorem, the asymptotic characteristics are assumed to be as  $n \rightarrow \infty$ . The proof for Lemma 1 is given in [21].

**Lemma 1.** *In case B1 is holds, then  $\alpha_n = 0_p(1)$ .*

**Lemma 2.** *In case B2 is holds, then  $\|\Psi(\beta_n)\| \rightarrow 0$  implies  $\|\beta_n - \beta_0\| \rightarrow 0$  for any sequence  $\{\beta_n\} \in \Theta$ .*

*Proof of Lemma 2.* Consider the following

$$\begin{aligned} \|\Psi(\beta_n)\| &= \|P\psi_{\beta_n}, 0\| = \|P\phi_{\beta_n}\| = \|P(y - (x^T \beta_n))x\| \\ &= \|P(y - (x^T \beta_0) + (x^T \beta_0) - (x^T \beta_n))x\| \\ &= \|P((x^T \beta_0))x + P((x^T \beta_0) - (x^T \beta_n))x\| \end{aligned}$$

If  $\|\Psi(\beta_n)\| \rightarrow 0$ , then  $\Psi(\beta_n) \rightarrow 0$ . Since  $P(y - (x^T \beta_n))x = 0$ , we may see from the equality above that  $P((x^T \beta_0) - (x^T \beta_n))x \rightarrow 0$ . Note

$$\begin{aligned} P((x^T \beta_0) - (x^T \beta_n))x &= P^{(1)}(c)(x^T \beta_n - x^T \beta_0)x \\ &\leq \frac{1}{4}P(x^T(\beta_n - \beta_0)x), \end{aligned}$$

where  $c \in (x^T \beta_0, x^T \beta_n)$  and  $h'(\mu)$  is the initial derivation for logistic link,

$$h'(\mu) = \frac{e^x}{(e^x + 1)^2} \in (0, 1/4). \text{ We also have } X = (x_1, x_2, \dots, x_p)^T,$$

$$\beta_0 - \beta_n = \beta_d = (\beta_{d1}, \beta_{d2}, \dots, \beta_{dp})^T.$$

$$P(x^T(\beta_0 - \beta_n)x) = P\left(\sum_{i=1}^p \beta_{di} \begin{bmatrix} x_1 x_i \\ x_2 x_i \\ \vdots \\ x_p x_i \end{bmatrix}\right) = \sum_{i=1}^p \beta_{di} P\left(\begin{bmatrix} x_1 x_i \\ x_2 x_i \\ \vdots \\ x_p x_i \end{bmatrix}\right) \rightarrow 0.$$



Since  $P(xx^T) = E(xx^T)$  is nonsingular, we have  $\beta_0 - \beta_1 \rightarrow 0$ , is shown by the fact that  $\|\beta_n - \beta_0\| \rightarrow 0$ .

**Lemma 3.** *Suppose that B3 and B4 are true. Then the class  $\{\psi_{\eta, \beta} : \beta \in \Theta, \|\eta - \eta_0\| < \delta\}$  is P-Glivenko-Cantelli for some  $\delta > 0$ , where  $\eta_0 = (0, \mu_0, \Sigma_0)$ .*

*Proof of Lemma 3.* To demonstrate that a class  $F$  of vector-valued functions  $\psi : (x, y) \rightarrow \mathbb{R}^p$  to be Glivenko-Cantelli, we must display each of the coordinate classes  $\phi^i : (x, y) \rightarrow \mathbb{R}$  with  $\psi = (\psi^1, \dots, \psi^p)^T$  ranging over  $F (i = 1, 2, \dots, p)$  is Glivenko-Cantelli.

The class  $F = \{\psi_\gamma : \gamma = (\beta, \eta) = (\beta, \alpha, \mu, \Sigma), \alpha \in [0, 1], \mu \in \mathbb{R}^q, \Sigma \in \mathbb{S}_+^q, \beta \in \Theta, \|\eta - \eta_0\| < \delta\}$  is a set of measurable functions that are indexed by a bounded subset in  $\Gamma \subset \Theta \times \mathbb{R}^q \times \mathbb{R} \times \mathbb{R}^{q \times q}$ , and  $\mathbb{S}_+^q$  represents a collection of positive semidefinite matrices described in  $\mathbb{R}^{q \times q}$ . This is because  $\Sigma$  is basically a variance-covariance matrix of continuous predictor variables, hence it is positive semidefinite and symmetric. For the norm, we use

$\|\eta - \eta_n\| = (\alpha^2 + \|\mu - \mu_0\|^2 + \|\Sigma - \Sigma_0\| + \|\beta - \beta_0\|^2)^{\frac{1}{2}}$ , with  $\|\cdot\|$  indicates the Euclidean norm of vectors  $\mu, \beta$  and for matrix  $\Sigma$ ,  $\|\cdot\|$  represents the induced norm in general. For two values  $\gamma_i = (\beta_i, \eta_i)$ , ( $i = 1, 2$ ) of  $\gamma$ , we have

$$\begin{aligned} & |\psi_{\gamma_1}^i(x, y) - \psi_{\gamma_2}^i(x, y)| \\ &= |w_{\eta_1}(x)\phi_{\beta_1}^i(x, y) - w_{\eta_2}(x)\phi_{\beta_2}^i(x, y)| \\ &= |w_{\eta_1}(x)\phi_{\beta_1}^i(x, y) - w_{\eta_1}(x)\phi_{\beta_1}^i(x, y) + w_{\eta_1}(x, y)\phi_{\beta_2}^i(x, y) - w_{\eta_2}(x)\phi_{\beta_2}^i(x, y)| \\ &\leq |w_{\eta_1}(x)\phi_{\beta_1}^i(x, y) - w_{\eta_1}(x)\phi_{\beta_2}^i(x, y)| + |w_{\eta_1}(x)\phi_{\beta_2}^i(x, y) - w_{\eta_2}(x)\phi_{\beta_2}^i(x, y)|. \end{aligned} \tag{10}$$

But

$$\begin{aligned} & |w_{\eta_1}(x)\phi_{\beta_1}^i(x, y) - w_{\eta_1}(x)\phi_{\beta_2}^i(x, y)| \\ &= w_{\eta_1}(x) \left| (y - (x^T \beta_1))x_i - (y - (x^T \beta_2))x_i \right| \leq \left| (x^T \beta_1) - (x^T \beta_2) \right| |x_i| \\ &\leq K_0 |x_i| |x^T \beta_2 - x^T \beta_1| \leq K_0 |x_i| \|c\| \|\beta_2 - \beta_1\|, \end{aligned} \tag{11}$$

where  $K_0$  represents the upper limit of the first derivation of the link function  $h'(\mu)$ .

Making use of the Mean Value Theorem, for each  $x_1$  and  $x_2$ , there exists  $c \in (x_1, x_2)$  such that

$$\|\mu(x_1) - \mu(x_2)\| = h'(c)|x_1 - x_2| < K_0 |x_1 - x_2|.$$

In addition, we have

$$\begin{aligned} & |w_{\eta_1}(x)\phi_{\beta_2}^i(x, y) - w_{\eta_2}(x)\phi_{\beta_2}^i(x, y)| = |w_{\eta_1}(x) - w_{\eta_2}(x)| |y - (x^T \beta_2)| |x_i| \\ &= |w_{\alpha_1, \mu_1, \Sigma_1}(x) - w_{\alpha_1, \mu_2, \Sigma_2}(x) + w_{\alpha_1, \mu_2, \Sigma_2}(x) - w_{\alpha_2, \mu_2, \Sigma_2}(x)| |y - (x^T \beta_2)| |x_i| \\ &\leq \left( |w_{\alpha_1, \mu_1, \Sigma_1}(x) - w_{\alpha_1, \mu_2, \Sigma_2}(x)| + |w_{\alpha_1, \mu_2, \Sigma_2}(x) - w_{\alpha_2, \mu_2, \Sigma_2}(x)| \right) |x_i|, \end{aligned} \tag{12}$$

then, we get

$$\begin{aligned} & \left| w_{\alpha_1, \mu_2, \Sigma_2}(x) - w_{\alpha_2, \mu_2, \Sigma_2}(x) \right| \\ &= \left| m \left( \alpha_1 (Z - \mu_2)^\top \Sigma_2^{-1} (Z - \mu_2) \right) - m \left( \alpha_2 (Z - \mu_2)^\top \Sigma_2^{-1} (Z - \mu_2) \right) \right| \quad (13) \\ &\leq K_1 |\alpha_2 - \alpha_1| (Z - \mu_2)^\top \Sigma_2^{-1} (Z - \mu_2), \end{aligned}$$

and

$$\begin{aligned} & w_{\alpha_1, \mu_1, \Sigma_1}(x) - w_{\alpha_1, \mu_2, \Sigma_2}(x) \\ &= \left| m \left( \alpha_1 (Z - \mu_1)^\top \Sigma_1^{-1} (Z - \mu_1) \right) - m \left( \alpha_1 (Z - \mu_2)^\top \Sigma_2^{-1} (Z - \mu_2) \right) \right| \\ &\leq K_1 \alpha_1 \left| (Z - \mu_1)^\top \Sigma_1^{-1} (Z - \mu_1) - (Z - \mu_2)^\top \Sigma_2^{-1} (Z - \mu_2) \right| \quad (14) \\ &\leq K_1 \left| (Z - \mu_1)^\top \Sigma_1^{-1} (Z - \mu_1) - (Z - \mu_1)^\top \Sigma_2^{-1} (Z - \mu_1) \right| \\ &\quad + K_1 \left| (Z - \mu_1)^\top \Sigma_2^{-1} (Z - \mu_1) - (Z - \mu_2)^\top \Sigma_2^{-1} (Z - \mu_2) \right|. \end{aligned}$$

Given that  $\Sigma$  is a positive semidefinite matrix, as well  $\Sigma^{-1}$ . Indicate what the eigenvalues of  $\Sigma_2^{-1}$  are as  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Consequently,  $\Sigma_2^{-1}$  has a set of orthogonal eigenvectors, say  $p_1, \dots, p_q$ , s.t.  $\Sigma_2^{-1} p_i = \lambda_i p_i$ . There exists an orthogonal matrix in matrix format  $Q$  s.t.

$$Q^{-1} \Sigma_2^{-1} Q = Q^\top \Sigma_2^{-1} Q = \Lambda.$$

Then we get

$$\begin{aligned} & \left| (Z - \mu_1)^\top \Sigma_2^{-1} (Z - \mu_1) - (Z - \mu_2)^\top \Sigma_2^{-1} (-\mu_2) \right| \\ &= \left| \sum_{i=1}^q \lambda_i (p_i^\top (-\mu_1))^2 - \sum_{i=1}^q \lambda_i (p_i^\top (Z - \mu_2))^2 \right| \\ &= \left| \sum_{i=1}^q \lambda_i p_i^\top (2Z - \mu_1 - \mu_2) p_i^\top (\mu_2 - \mu_1) \right| \\ &\leq \sum_{i=1}^q \lambda_i \|p_i\|^2 \|2Z - \mu_1 - \mu_2\| \|\mu_2 - \mu_1\| \\ &\leq \lambda_1 \|\mu_2 - \mu_1\| \|2Z - \mu_1 - \mu_2\| \sum_{i=1}^q \|p_i\|^2 \\ &= \rho(\Sigma_2^{-1}) \|\mu_2 - \mu_1\| \|2Z - \mu_1 - \mu_2\| \sum_{i=1}^q \|p_i\|^2 \quad (15) \\ &\leq \|\Sigma_2^{-1}\| \|\mu_2 - \mu_1\| \|2Z - \mu_1 - \mu_2\|, \end{aligned}$$

where  $\rho(\Sigma_2^{-1})$  is the spectral radius of  $\Sigma_2^{-1}$  and  $\rho(A) \leq \|A\|$  applicable to any induced norm. Similarly  $\Sigma_2^{-1} - \Sigma_1^{-1}$  is also symmetric, thus indicate the eigenvalues of  $\Sigma_2^{-1} - \Sigma_1^{-1}$  as  $p_1^*, \dots, p_q^*$  s.t.  $(\Sigma_2^{-1} - \Sigma_1^{-1}) p_i^* = \lambda_i^* p_i^*$ . There exists an orthogonal matrix in matrix format  $Q^*$  s.t.

$$(Q^*)^{-1} (\Sigma_2^{-1} - \Sigma_1^{-1}) Q^* = (Q^*)^\top (\Sigma_2^{-1} - \Sigma_1^{-1}) Q^* = \Lambda^*.$$

Next, we have

$$\begin{aligned}
 & \left| (Z - \mu_1)^T \Sigma_2^{-1} (Z - \mu_1) - (Z - \mu_1)^T \Sigma_1^{-1} (Z - \mu_1) \right| \\
 &= \left| (Z - \mu_1)^T (\Sigma_2^{-1} - \Sigma_1^{-1}) (Z - \mu_1) \right| \\
 &= \left| (Z - \mu_1)^T \mathcal{Q}^* \Lambda^* (\mathcal{Q}^*)^T (Z - \mu_1) \right| \\
 &= \left| \left[ (\mathcal{Q}^*)^T (Z - \mu_1) \right]^T \Lambda^* \left[ (\mathcal{Q}^*)^T (Z - \mu_1) \right] \right| \\
 &= \left| \sum_{i=1}^q \lambda_i \left( (p_i^*)^T (Z - \mu_1) \right)^2 \right| \\
 &\leq \max_{i=1, \dots, q} |\lambda_i^*| \sum_{i=1}^q \left( (p_i^*)^T (Z - \mu_1) \right)^2 \\
 &\leq \rho(\Sigma_2^{-1} - \Sigma_1^{-1}) \|Z - \mu_1\|^2 \sum_{i=1}^q \left( (p_i^*)^2 \right) \\
 &\leq \|Z - \mu_1\|^2 \|\Sigma_2^{-1} - \Sigma_1^{-1}\|.
 \end{aligned} \tag{16}$$

[23] presented a simple and popular M-estimator that minimizes a “bounded” version of the sum of residuals. The estimating equation is

$$\hat{\theta}_M = \sum_{i=1}^n \psi \left( \frac{r_i(\hat{\theta})}{\sigma} \right) x_i = 0 \tag{17}$$

Then, using (12), (13), (14), (15) and (17) we get

$$\begin{aligned}
 & \left\| w_{n_1}(x) \phi_{\beta_2}^i(x, y) - w_{n_2}(x) \phi_{\beta_2}(x, y) \right\| \\
 &< K_1 |\alpha_2 - \alpha_1| (Z - \mu_2)^T \Sigma_2^{-1} (Z - \mu_2) |x_i| \\
 &+ K_1 \|\Sigma_2^{-1}\| \|\mu_2 - \mu_1\| \|2Z - \mu_1 - \mu_2\| \|x_i\| \\
 &+ K_1 \|Z - \mu_1\|^2 \|\Sigma_2^{-1} - \Sigma_1^{-1}\| \|x_i|.
 \end{aligned} \tag{18}$$

Since  $|\alpha_2 - \alpha_1|, \|\mu_2 - \mu_1\|, \|\beta_2 - \beta_1\| < \|\gamma_2 - \gamma_1\|$ , from (11), (10) and (16) now, we can provide a bound for  $\psi$  :

$$\begin{aligned}
 & \left| \psi_{\gamma_1}^i(x, y) - \psi_{\gamma_2}(x, y) \right| \\
 &< K_0 \|\beta_2 - \beta_1\| \|x_i\| \|x\| + K_1 |\alpha_2 - \alpha_1| (Z - \mu_2)^T \Sigma_2^{-1} (Z - \mu_2) |x_i| \\
 &+ K_1 \|\Sigma_2^{-1}\| \|\mu_2 - \mu_1\| \|2Z - \mu_1 - \mu_2\| \|x_i\| + K_1 \|Z - \mu_1\|^2 \|\Sigma_2^{-1} - \Sigma_1^{-1}\| \|x_i\| \\
 &\leq \left( K_0 \|x_i\| \|x\| + K_1 (Z - \mu_2)^T \Sigma_2^{-1} (Z - \mu_2) |x_i| \right. \\
 &\quad \left. + K_1 \|\Sigma_2^{-1}\| \|2Z - \mu_1 - \mu_2\| \|x_i\| + K_1 \|Z - \mu_1\|^2 \|x_i\| \right) \|\gamma_2 - \gamma_1\| \\
 &= L^i(x) \|\gamma_2 - \gamma_1\|,
 \end{aligned}$$

for every  $\gamma_1, \gamma_2$ . We have constructed a Lipschitz criterion for each  $\psi^i, i = 1, \dots, p$ , so for  $\psi$  we have  $|\psi_{\gamma_1}(x, y) - \psi_{\gamma_2}(x, y)| < L(x) \|\gamma_2 - \gamma_1\|$ , every  $\gamma_1, \gamma_2$ , where

$$\begin{aligned}
 L(x) &= K_0 \|x\| \|x\| + K_1 (Z - \mu_2)^T \Sigma_2^{-1} (Z - \mu_2) x \\
 &+ K_1 \|\Sigma_2^{-1}\| \|2Z - \mu_1 - \mu_2\| \|x_i\| + K_1 \|Z - \mu_1\|^2 x.
 \end{aligned}$$

Now investigate the bracketing entropy in relative to the  $L_r(P)$ -norm

$$\|\psi_i\|_{P_r} = \left( P|\psi|^r \right)^{\frac{1}{r}}.$$

Use brackets of the type  $[\psi_\gamma - \varepsilon L, \psi_\gamma + \varepsilon L]$  for  $\gamma$  spanning over a sufficient chosen subset of  $\Gamma$  and these brackets have  $L_r(P)$ -size  $2\varepsilon\|L\|_{P_r}$ . If  $\gamma$  ranges over a grid of mesh width  $\varepsilon$  over  $\Gamma$  then the brackets  $[\psi_\gamma - \varepsilon L, \psi_\gamma + \varepsilon L]$  range over  $F$ . Based on the Lipschitz condition we get

$$\psi_{\gamma_1} - \varepsilon L \leq \psi_{\gamma_2} \leq \psi_{\gamma_1} + \varepsilon L, \text{ if } \|\gamma_2 - \gamma_1\| \leq \varepsilon,$$

Hence, many brackets are required as in radial balls  $\frac{\varepsilon}{2}$  to cover  $\Gamma$ , alternatively we require less than  $(\text{diam}\Gamma/\varepsilon)^{2p+2}$  cubes with size  $\varepsilon$  to cover parameter space  $\Gamma$ . If  $P|L_i|^r < \infty$ , then a constant  $J$  exists, depends only on  $\Gamma$  and  $P$ , in such a way that the bracketing numbers fulfill

$$N_{[]}(\varepsilon, F, L_r(P)) \leq J \left( \frac{\text{diam}\Gamma}{\varepsilon} \right)^p, \text{ every } 0 < \varepsilon < \text{diam}\Gamma.$$

Given that all  $\psi \in F$  are continuous functions, they can be measured. If B3 is fulfilled, then  $P|L| < \infty$ , and as a result, class  $F$  is P-Glivenko-Cantelli from the Theorem (19.4) (Glivenko-Cantelli) in [24].  $\square$

If B1, B2, B3 and B4 are hold, then estimators  $\hat{\beta}_n$  are used as the solving to the equation  $\Psi(\hat{\beta}_n) = 0$  converges to  $\beta_0$  in probability.

*Proof of Theorem 1.* denote

$$\Psi(\beta) = P\psi_{\beta,0} = P\phi_\beta \Psi(\beta, \eta) = P\psi_{\beta,\eta} = P\phi_\beta W_\eta$$

$$\Psi(\beta) = \Psi_n(\beta, \eta) = \frac{1}{n} \sum_{i=1}^n \psi_{\beta,\eta} \Psi_n(\beta, \hat{\eta}) = \frac{1}{n} \sum_{i=1}^n \psi_{\beta,\hat{\eta}}$$

Note that  $\Psi(\hat{\beta}_n) = (\Psi(\hat{\beta}_n) - \Psi_n(\hat{\beta}_n)) + \Psi_n(\hat{\beta}_n)$ . Then we can establish that  $\sup_{\beta \in \Theta} \|\Psi_n(\beta) - \Psi(\beta)\| = 0_p(1)$ , then  $\|\Psi(\hat{\beta}_n)\| = 0_p(1)$  since  $\Psi_n(\hat{\beta}_n) = 0_p(1)$ . Thus, based upon Lemma 2, we have  $\|\hat{\beta}_n - \beta_0\| = 0_p(1)$ . To show  $\sup_{\beta \in \Theta} \|\Psi_n(\beta) - \Psi(\beta)\| = 0_p(1)$ , we consider it as follows:

$$\begin{aligned} & \sup_{\beta \in \Theta} \|\Psi_n(\beta, \hat{\eta}) - \Psi(\beta)\| \\ &= \sup_{\beta \in \Theta} \|\Psi_n(\beta, \hat{\eta}) + \Psi(\beta, \eta) + \Psi_n(\beta, \eta) - \Psi(\beta, \eta) + \Psi(\beta, \eta) - \Psi(\beta, 0)\| \\ &\leq J_1 + J_2 + J_3, \end{aligned}$$

where

$$J_1 = \sup_{\beta \in \Theta} \|\Psi_n(\beta, \hat{\eta}) - \Psi_n(\beta, \eta)\|$$

$$J_2 = \sup_{\beta \in \Theta} \|\Psi_n(\beta, \eta) - \Psi(\beta, \eta)\|$$

$$J_3 = \sup_{\beta \in \Theta} \|\Psi(\beta, \eta) - \Psi(\beta, 0)\|.$$

Lemma 3 informs us that  $F$  is P-Glivenko-Cantelli class, so  $J_2 \rightarrow 0$ . For  $J_1$ ,

$$\begin{aligned} J_1 &= \sup_{\beta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n \phi_{\beta}(x_i, y_i) (w_{\hat{\eta}}(x_i) - w_{\eta}(x_i)) \right\| \\ &\leq \frac{1}{n} \sum_{i=1}^n \sup_{\beta \in \Theta} \left\| \phi_{\beta}(x_i, y_i) (w_{\hat{\eta}}(x_i) - w_{\eta}(x_i)) \right\| \\ &= \frac{1}{n} \sum_{i=1}^n |w_{\hat{\eta}}(x_i) - w_{\eta}(x_i)| \sup_{\beta \in \Theta} \left\| \phi_{\beta}(x_i, y_i) \right\|. \end{aligned}$$

Using (16), we get

$$\begin{aligned} |w_{\hat{\eta}}(x_i) - w_{\eta}(x_i)| &< K_1 |\hat{\alpha} - \alpha| (Z - \mu)^{\top} \Sigma^{-1} (Z - \mu) \\ &\quad + K_1 \left\| \Sigma^{-1} \right\| \left\| \hat{\mu} - \mu \right\| \left\| 2Z - \hat{\mu} - \mu \right\| \\ &\quad + K_1 \left\| Z - \hat{\mu} \right\|^2 \left\| \hat{\Sigma}^{-1} - \Sigma^{-1} \right\|. \end{aligned} \quad (19)$$

Since  $\hat{\eta} - \eta = (\hat{\alpha} - \alpha, \hat{\mu} - \mu, \hat{\Sigma} - \Sigma) = 0_p(1)$ , we have  $\hat{\alpha} - \alpha = 0_p(1)$ ,  $\hat{\mu} - \mu = 0_p(1)$  and  $\hat{\Sigma} - \Sigma = 0_p(1)$ . Also,  $\hat{\Sigma}^{-1} - \Sigma^{-1} = \hat{\Sigma}^{-1} (\Sigma - \hat{\Sigma}) \Sigma^{-1}$ , as a result  $\hat{\Sigma}^{-1} - \Sigma^{-1} = 0_p(1)$ . Therefore, (19) follows that  $|w_{\hat{\eta}}(X_i)| = 0_p(1)$ . Because B2 and B3 are fulfilled, we also have  $\sup_{\beta \in \Theta} \left\| \phi_{\beta}(x, y) \right\|^2 < \infty$ , then we get  $J_1 \rightarrow^p 0$ . For  $J_3$

$$\begin{aligned} J_3 &= \sup_{\beta \in \Theta} \left\| P \phi_{\beta}(x, y) (w_{\eta}(x) - w_{\eta_0}(x)) \right\| \\ &\leq \sup_{\beta \in \Theta} P \left\| \phi_{\beta}(x, y) (w_{\eta}(x) - w_{\eta_0}(x)) \right\| \\ &\leq \sup_{\beta \in \Theta} P^{\frac{1}{2}} \left\| \phi_{\beta}(x, y) \right\|^2 P^{\frac{1}{2}} |w_{\eta}(x) - w_{\eta_0}(x)|^2 \\ &= P^{\frac{1}{2}} |w_{\eta}(x) - w_{\eta_0}(x)|^2 \sup_{\beta \in \Theta} P^{\frac{1}{2}} \left\| \phi_{\beta}(x, y) \right\|^2, \end{aligned}$$

and the Cauchy-Schwarz inequality was employed in the final inequality. We already know

$$w_{\eta}(x) - w_{\eta_0}(x) = 1 - m \left( \eta(Z - \mu)^{\top} \Sigma^{-1} (Z - \mu) \right) \rightarrow 0.$$

Then  $\hat{\eta} = 0_p(1)$   $m(0) = 1$  and  $|w_{\hat{\eta}}(x) - w_{\eta_0}(x)|^2$  is bounded by 1, using the dominated convergence theorem we get

$$\lim_{n \rightarrow \infty} P^{\frac{1}{2}} |w_{\eta}(x) - w_{\eta_0}(x)|^2 \rightarrow 0.$$

We also have

$$\sup_{\beta \in \Theta} P \left\| \phi_{\beta}(x, y) \right\|^2 = \sup_{\beta \in \Theta} P \left\| (y - x^{\top} \beta) x \right\|^2 \leq P \|x\|^2.$$

and from assumption B3, we obtain  $\sup_{\beta \in \Theta} P \left\| \phi_{\beta}(x, y) \right\|^2 < \infty$ . It then follows that  $J_3 \rightarrow 0$ .  $\square$

## 5. Assess Robustness of the Estimators

To evaluate the robustness of the methods, two methods have been employed. The first includes simulated models for contrasting the novel technique with the traditional MLE, Mallows-type estimator (Mallows) and (CUBI). In the second, we used the actual leukemia data set and The Erythrocyte Sedimentation Rate Data.

### 5.1. Simulation study

A Monte Carlo simulation analysis was performed for this subsection to assess the efficiency and robustness of the suggested estimator  $\hat{\beta}_n$ . For the initial robust estimators of the scatter and location  $\hat{\Sigma}$  and  $\hat{\mu}$  utilized the minimum covariance determinant (MCD). We calculated the following estimators for comparison: MLE, conditionally unbiased bounded influence (CUBI) of [2], and the Mallows-type estimator (Mallows) of [3]. In the simulation, the following weight function  $m(x)$  was applied:

$$m(x) = a(1 - (x-1)^6) + b,$$

where  $a = 0.8$  and  $b = 0.2$ .

Three models are involved in the simulation study: a clean logistic regression model, a contaminated model with a 10% contamination rate, and a contaminated model with a 20% contamination rate. First, clean model, the standard normal distribution was used to generate two predictor variables,  $x_1 \sim N(0,1)$  and  $x_2 \sim N(0,1)$ . Three sample sizes were used: ( $n = 100, 300, 500$ ) and  $p = 2$ . We generate the response variable according to the Bernoulli distribution with a parameter equal to  $\pi_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2) / (1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2))$ . The values of the true parameters  $\beta$  are taken  $(0, 1.6, 1.2)$  for three models. Second, the percentage of contamination in data equal 10%, and their predicted variables are generated from a normal distribution with  $(\mu = 0)$  and  $(\sigma = 1)$ . Third, the percentage of contamination in data equals 20% in a similar manner to the above.

The performance of these estimators is evaluated using Bias and mean squared error for various models. Nevertheless, the estimator with the smallest Bias and MSE is the best. Each scenario simulation included over 1000 repetitions. Consequently, for each parameter, the following are the calculations for bias and mean squared error:

$$\text{Bias} = \left| \frac{1}{1000} \sum_{i=1}^{1000} \beta_i - \beta \right|,$$

and

$$\text{MSE} = \frac{1}{1000} \sum_{i=1}^n |\hat{\beta}_i - \beta|^2.$$

### 5.2. Numerical Results

The numerical results, displayed in this paper, are based on simulation studies,

and two real data applications. This numerical result is expected to evaluate, the performance of a proposed model. **Table 1** shows the bias and mean squared errors of the four estimation techniques for the clean model. The findings show that the bias and MSE of the MLE, Mallows, and CUBIF estimators are relatively similar, while the WMLT estimator performs worse than the other estimators. When the sample size increases, the bias and mean squared errors are observed to decrease. As shown in **Table 2**, under 10% of the data were contaminated, so the new robust approach WMLT has the greatest overall performance among all comparable estimators for varied sample sizes. **Table 3** demonstrates that when 20% of the data are contaminated, our proposed technique (WMLT) outperforms other estimators in terms of bias and mean squared errors. Due to the sensitivity of anomalies, conventional maximum likelihood estimates perform inadequately in the contaminated model. In conclusion, the proposed method outperforms all other methods compared with contaminated data. Furthermore, the new estimator performs reasonably well with clear data.

**Table 1.** Bias and mean squared errors of estimators for clean model.

Methods	$n = 100$		$n = 300$		$n = 500$	
	Bias	MSE	Bias	MSE	Bias	MSE
MLE	0.340	0.279	0.170	0.071	0.129	0.040
Mallows	0.335	0.276	0.179	0.074	0.131	0.041
CUBIF	0.340	0.279	0.170	0.070	0.129	0.040
WMLT	0.967	2.543	0.850	1.091	0.808	0.563

**Table 2.** Bias and (MSE) of estimators for second model (10% of data are contaminated).

Methods	$n = 100$		$n = 300$		$n = 500$	
	Bias	MSE	Bias	MSE	Bias	MSE
MLE	1.565	4.457	1.321	2.806	1.282	2.590
Mallows	1.538	4.400	1.323	2.819	1.280	2.578
CUBIF	1.551	4.435	1.312	2.772	1.275	2.562
WMLT	0.330	0.386	0.252	0.174	0.214	0.071

**Table 3.** Bias and (MSE) of estimators for second model (20% of data are contaminated).

Methods	$n = 100$		$n = 300$		$n = 500$	
	Bias	MSE	Bias	MSE	Bias	MSE
MLE	2.522	3.168	1.392	3.011	1.339	2.857
Mallows	1.739	3.118	1.377	2.809	1.329	2.137
CUBIF	2.250	3.885	1.383	3.145	1.328	2.809
WMLT	0.242	0.104	0.222	0.077	0.069	0.050

### 5.3. Leukemia Data

This study uses data from [25] which includes information from 33 people who died from acute myeloid leukemia. Each patient was measured for three variables: AG, WBC, and time. The response variable represents the survival time in weeks of the patient; it was converted into a binary variable with ( $Y = 1$ ) indicating patients whose survival time exceeded 52 weeks and ( $Y = 0$ ) indicating those who did not. WBC represents the white blood cell concentration of the patient. Whereas AG (present = 1, absent = 0) measured the presence or absence of a morphologic characteristic of white blood cells. The observation number 17 appears to be atypical. Using AG and WBC as predictor variables and binary survival time  $y$  as the response variable, a logistic regression model was constructed. The estimators analyzed here are weighted maximum likelihood technique (WMLT), MLE, MLE17 (MLE17 is the maximum likelihood estimator for clean data when observation 17 is excluded.), Mallows (estimator of the Mallows type), and CUBIF (conditionally unbiased bounded-influence function estimator).

**Table 4** demonstrates that the MLE is extremely sensitive to influential observations. Furthermore, eliminating observation 17 lowered the impact of WBC to near nil. For the leukemia data, the new estimator (WMLT) demonstrated the greatest performance among all other estimators. However, Mallows's estimators are reasonably close to the MLE17.

### 5.4. The Erythrocyte Sedimentation Rate Data

The Erythrocyte Sedimentation Rate (ESR) data. In this data, the primary objective was to determine if the levels of two plasma proteins (Fibrinogen and  $\gamma$ -Globulin) were responsible for the increase in ESR in healthy individuals. The research was conducted by the Institute of Medical Research in Kuala Lumpur, Malaysia, on 32 patients, and the original data was collected by [26]. The response of zero indicates a healthy person, whereas the response of one indicates an unwell person. Here, the continuous variables (FIB and  $\gamma$ GLO) are compared to the binary response (ESR). In the original ESR data, [27] identified two outliers (cases 13 and 29) in  $X$ -space. Cases 14 and 15 are influential observations. Therefore, removing instances 14 and 15 would result in cases with no overlap.

**Table 4.** The estimated parameters and standard errors for the leukemia data.

Intercept	WBC		AG			
	Est.	S.E.	Est.	S.E.	Est.	S.E.
MLE	-1.3073	0.8145	0.3717	0.0186	2.2617	0.9522
MLE <sub>37</sub>	0.2119	1.0830	-0.0002	0.0001	2.5580	1.2341
Mallows	0.1602	1.0697	-0.2245	0.0129	2.5252	1.2159
CUBIF	-1.4503	1.8504	-0.0527	0.0337	0.2202	0.9756
WMLT	-0.404	0.285	-1.710	0.052	1.302	0.341



Case 13 was modified so that  $(y, x_1, x_2) = (3, 6, 37)$  in order to execute uncontaminated data analysis. From the uncontaminated data, the ESR data were contaminated where the occurrences ( $Y = 1$ ) and non-occurrences ( $Y = 0$ ) were replaced with each other for cases 14 and 15, and this might only leave one out of the three overlapping cases for the ESR data.

Under contaminated ESR data,  $\beta_0$  and  $se(\beta_0)$  of all estimators are impacted by outliers as compared to the other parameters (see **Table 5**). The results shown in **Table 5** also indicate that the MLE is primarily influenced by outliers. Following the modification of the tainted data, only one overlapping observation, case 13, remains. This is the reason why the coefficients and standard errors of the WMLT that downweights this observation are so large. Even though the WMLT has the smallest  $\chi^2$  value, the WMLT estimator should also be taken into consideration. The results shown in **Table 5** indicate that the WMLT is a suitable estimator for the ESR data because its estimates are relatively closer to the MLE for uncontaminated data.

## 6. Conclusions

In this paper, we develop a new robust technique for logistic regression, also known as the weighted maximum likelihood technique (WMLT). The asymptotic consistency of the proposed estimator was demonstrated.

In order to evaluate the robustness of a new technique, we conducted simulation experiments using a variety of scenarios and data sets. Classical maximum likelihood estimates show the lack of robustness in the presence of outliers. Our simulation study for the clean model illustrated that the MLE, Mallows, and CUBIF estimators perform similarly, while the new weighted technique performs less effectively than the other estimators. The simulation study also shows that the WMLT technique outperforms other estimators when dealing with contaminated data and demonstrates the greatest performance among all estimators compared to various scenarios and real data sets. The proposed method (WMLT) can be applied to other generalized linear models (GLMs) and is expected to be superior to existing methods in practical applications. The findings

**Table 5.** Estimated coefficients, standard errors, and  $\chi^2$  for ESR.

	MLE <sub>un</sub>	MLE	MALLOWS	CUBIF	WMLL
Int. $\beta_0$	12.263	19.882	20.449	20.579	19.435
$se(\beta_0)$	5.839	9.417	9.809	10.031	9.371
FIB $\beta_1$	1.830	2.597	2.648	3.053	2.357
$se(\beta_1)$	1.062	1.543	1.611	1.681	1.787
$\gamma$ .GLO $\beta_2$	0.153	0.278	0.286	0.256	0.253
$se(\beta_2)$	0.116	0.165	0.170	0.170	0.123
$\chi^2$	42.237	27.050	26.628	26.047	25.274

of this chapter provide researchers and practitioners with a new approach to developing robust estimators for logistic regression and potentially other generalized linear models (GLMs).

### Acknowledgments

We thank the Editor and the referee for their comments. Research of M. Baron is funded by the National Science Foundation grant DMS 1322353. This support is greatly appreciated.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Pregibon, D. (1981) Logistic Regression Diagnostics. *Annals of Statistics*, **4**, 705-724. <https://doi.org/10.1214/aos/1176345513>
- [2] Kunsch, H.R., Stefonski, L.A. and Carroll, R.J. (1989) Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models. *Journal of the American Statistical Association*, **84**, 460-466. <https://doi.org/10.1080/01621459.1989.10478791>
- [3] Carroll, R.J. and Pederson, S. (1993) On Robustness in the Logistic Regression Model. *Journal of the Royal Statistical Society*, **55**, 693-706. <https://doi.org/10.1111/j.2517-6161.1993.tb01934.x>
- [4] Hobza, T., Mart, N. and Pardo, L. (2017) A Wald-Type Test Statistic Based on Robust Modified Median Estimator in Logistic Regression Models. *Journal of Statistical Computation and Simulation*, **87**, 2309-2333. <https://doi.org/10.1080/00949655.2017.1330414>
- [5] Bergesio, A. and Yohai, V.J. (2011) Projection Estimators for Generalized Linear Models. *Journal of the American Statistical Association*, **106**, 661-671. <https://doi.org/10.1198/jasa.2011.tm09774>
- [6] Myers, M.D.C.R.H. and Vining, G.G. (1992) Least-Absolute-Deviations Fits for Generalized Linear Models. *Biometrika*, **79**, 747-754. <https://doi.org/10.1093/biomet/79.4.747>
- [7] Cantoni, E. and Ronchetti, E. (2001) Robust Inference for Generalized Linear Models. *Journal of the American Statistical Association*, **96**, 1022-1030. <https://doi.org/10.1198/016214501753209004>
- [8] Wedderburn, R.W.M. (1974) Quasi-Likelihood Functions, Generalized Linear Models, and the Gaussian Newton Method. *Biometrika*, **61**, 439-447. <https://doi.org/10.1093/biomet/61.3.439>
- [9] Croux, C., Flandre, C. and Haesbroeck, G. (2002) The Breakdown Behavior of the Maximum Likelihood Estimator in the Logistic Regression Model. *Statistics and Probability Letters*, **60**, 377-386. [https://doi.org/10.1016/S0167-7152\(02\)00292-4](https://doi.org/10.1016/S0167-7152(02)00292-4)
- [10] Rousseeuw, P.J. and Christmann, A. (2005) Robustness against Separation and Outliers in Logistic Regression. *Quality Control and Applied Statistics*, **50**, 451-452.
- [11] Idriss, I.A. and Cheng, W. (2023) Robust Estimators for Poisson Regression. *Open Journal of Statistics*, **13**, 112-118. <https://doi.org/10.4236/ojs.2023.131007>

- [12] Valdora, M. and Yohai, V.J. (2014) Robust Estimators for Generalized Linear Models. *Journal of Statistical Planning and Inference*, **146**, 31-48. <https://doi.org/10.1016/j.jspi.2013.09.016>
- [13] Bianco, A.M. and Yohai, V.J. (1996) Robust Estimation in the Logistic Regression Model. Springer, New York, 17-34. [https://doi.org/10.1007/978-1-4612-2380-1\\_2](https://doi.org/10.1007/978-1-4612-2380-1_2)
- [14] Croux, C. and Haesbroeck, G. (2014) Implementing the Bianco and Yohai Estimator for Logistic Regression. *Computational Statistics and Data Analysis Journal*, **146**, 31-48.
- [15] Muller, C.H. and Neykov, N. (2003) Breakdown Points of Trimmed Likelihood Estimators and Related Estimators in Generalized Linear Models. *Journal of Statistical Planning and Inference*, **44**, 273-295. [https://doi.org/10.1007/978-3-642-57338-5\\_24](https://doi.org/10.1007/978-3-642-57338-5_24)
- [16] Copas, J.B. (1988) Binary Regression Models for Contaminated Data. *Journal of the Royal Statistical Society*, **50**, 225-265. <https://doi.org/10.1111/j.2517-6161.1988.tb01723.x>
- [17] Ahmed, I.A.I. and Cheng, W.H. (2020) The Performance of Robust Methods in Logistic Regression Model. *Open Journal of Statistics*, **10**, 127-138. <https://doi.org/10.4236/ojs.2020.101010>
- [18] Ruckstuhl, A.F. and Welsh, A.H. (2001) Robust Fitting of the Binomial Model. *The Annals of Statistics*, **29**, 1117-1136. <https://doi.org/10.1214/aos/1013699996>
- [19] Krasker, W.S. and Welsch, R.E. (1982) Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association*, **77**, 595-604. <https://doi.org/10.1080/01621459.1982.10477855>
- [20] Stefanski, L.A., Carroll, R.J. and Ruppert, D. (1986) Optimally Bounded Score Functions for Generalized Linear Models with Applications to Logistic Regression. *Biometrika*, **73**, 413-424. <https://doi.org/10.1093/biomet/73.2.413>
- [21] Gervini, D. (2005) Robust Adaptive Estimators for Binary Regression Models. *Journal of Statistical Planning and Inference*, **131**, 297-311. <https://doi.org/10.1016/j.jspi.2004.02.006>
- [22] Vandev, D.L. and Neykov, N.M. (1998) About Regression Estimators with High Breakdown Point. *Statistics*, **32**, 111-129. <https://doi.org/10.1080/02331889808802657>
- [23] Huber, P.J. (1964) Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, **35**, 73-101. <https://doi.org/10.1214/aoms/1177703732>
- [24] Vaart, A.W. (2000) Asymptotic Statistics. Cambridge University Press, Cambridge.
- [25] Cox, D. and Oakes, D. (1984) Analysis of Survival Data. Chapman and Hall, New York.
- [26] Collett, D. and Jemain, A. (1985) Residuals, Outliers and Influential Observations in Regression Analysis. *Sains Malaysiana*, **14**, 493-511.
- [27] Syaiba, B. and Habshah, M. (2010) Robust Logistic Diagnostic for the Identification of High Leverage Points in Logistic Regression Model. *Journal of Applied Sciences*, **10**, 3042-3050. <https://doi.org/10.3923/jas.2010.3042.3050>