axioms

MDPI

Article

Handling Multicollinearity and Outliers in Logistic Regression Using the Robust Kibria–Lukman Estimator

Adewale F. Lukman [1,*], Suleiman Mohammed [2], Olalekan Olaluwoye [2] and Rasha A. Farghali [3]

[1] Department of Mathematics and Statistics, University of North Dakota, Grand Forks, ND 58202, USA
[2] Department of Applied Mathematical Sciences, African Institute for Mathematical Sciences, Mbour-Thies 23000, Senegal; mohammed.suleiman@aims-senegal.org (S.M.); olalekan.t.olaluwoye@aims-senegal.org (O.O.)
[3] Department of Mathematics, Insurance and Applied Statistics, Helwan University, Cairo 11795, Egypt; rasha.farghaly@commerce.helwan.edu.eg
* Correspondence: adewale.lukman@und.edu

Abstract: Logistic regression models encounter challenges with correlated predictors and influential outliers. This study integrates robust estimators, including the Bianco–Yohai estimator (BY) and conditionally unbiased bounded influence estimator (CE), with the logistic Liu (LL), logistic ridge (LR), and logistic KL (KL) estimators. The resulting estimators (LL-BY, LL-CE, LR-BY, LR-CE, KL-BY, and KL-CE) are evaluated through simulations and real-life examples. KL-BY emerges as the preferred choice, displaying superior performance by reducing mean squared error (MSE) values and exhibiting robustness against multicollinearity and outliers. Adopting KL-BY can lead to stable and accurate predictions in logistic regression analysis.

Keywords: logistic regression; outliers; multicollinearity; robust estimators; Bianco–Yohai estimator; ridge regression estimator

MSC: 62J07; 62J12; 62J20

Academic Editor: Hari Mohan Srivastava

Received: 27 November 2024
Revised: 18 December 2024
Accepted: 26 December 2024
Published: 30 December 2024

Citation: Lukman, A.F.; Mohammed, S.; Olaluwoye, O.; Farghali, R.A. Handling Multicollinearity and Outliers in Logistic Regression Using the Robust Kibria–Lukman Estimator. *Axioms* **2025**, *14*, 19. https://doi.org/10.3390/axioms14010019

1. Introduction

Logistic regression is a powerful statistical tool widely used for modeling the relationship between a binary response variable and one or more explanatory variables. Its versatility has made it a fundamental tool in medicine, biostatistics, finance, social sciences, engineering, and many other fields. Despite their popularity, logistic regression models can be affected by challenges such as overfitting, multicollinearity, and outliers. These challenges can lead to inaccurate conclusions and poor predictive performance, highlighting the importance of developing robust predictive models for logistic regression [1]. The logistic (inverse-link) function, characterized by its S-shaped curve, is fundamental to logistic regression modeling. It transforms any independent variable value into a probability ranging between 0 and 1, where 0 indicates the lowest probability of the event occurring and 1 represents the highest probability of the event occurring.

The maximum likelihood estimator (MLE) is a widely adopted method for estimating the parameters of a logistic regression model. However, this technique is susceptible to multicollinearity and outliers in the data, which can compromise the validity and reliability of the results. Multicollinearity occurs when two or more explanatory variables in the model are correlated. This can make it difficult to distinguish the individual effects of each variable on the outcome variable and can result in unstable and unreliable parameter estimates [2]. Similarly, outliers can significantly impact the estimated coefficients, particularly

in smaller datasets. These issues can cause the model to overfit the data, leading to poor predictive performance and potentially erroneous conclusions. As a result, it is essential to identify and address multicollinearity and outliers in the data before building a logistic regression model.

Several estimators have been developed to account for multicollinearity in the logistic regression model. These include the logistic ridge estimator [3], the principal component logistic estimator [4], the Liu logistic estimator [5], the Liu-type logistic estimator [6], the linearized ridge logistic estimator [7], the modified ridge-type logistic estimator [8], and the Kibria–Lukman logistic estimator [9].

Outliers are observations that deviate significantly from the rest of the data and can substantially impact statistical analysis [10]. Outliers can arise due to measurement errors, data entry errors, or genuine extreme values. They can distort the relationship between variables and lead to inaccurate estimates of statistical parameters. In linear regression analysis, outliers can pull the regression line toward themselves, leading to biased estimates of the coefficients and affecting prediction accuracy. Outliers can also increase the variance of the regression estimates, thus reducing the precision of the results. One common approach to address the influence of outliers is to use robust estimation techniques. These methods dampen the effect of the outliers on the estimation process and provide more reliable estimates of the regression coefficients. Robust estimation techniques are essential when the data distribution is non-normal or when there is an outlier. Robust estimation methods for linear regression models include the M-estimator, MM-estimator, S-estimator, least absolute deviation, and the least trimmed squares estimator, among others [11–13].

Research on linear regression has demonstrated that models can be significantly affected by both multicollinearity and outliers [14–16]. To address these challenges, various methods have been developed, such as shrinkage techniques (e.g., ridge regression) to mitigate multicollinearity and robust estimators (e.g., the M-estimator) to handle outliers. In some cases, these techniques have been combined to simultaneously address both problems in linear regression [14–16]. However, similar advancements have not been adequately explored in the context of logistic regression. Existing studies on logistic regression have primarily focused on addressing multicollinearity or outliers individually, leaving a gap in the research for methods capable of handling both issues simultaneously.

This study proposes a novel approach for logistic regression that integrates shrinkage estimators like ridge regression with robust methods such as the Bianco–Yohai estimator. By combining these two techniques, our work offers a unified solution to simultaneously handle multicollinearity and outliers. This dual-focus approach distinguishes our study from previous research, which has only tackled these problems separately. Our contribution, therefore, provides a more robust and accurate method for modeling logistic regression under the simultaneous presence of multicollinearity and outliers.

The article is structured as follows. In Section 2, we provide a comprehensive review of some of the existing methods and introduce the novel estimator. We delve into the discussion of the properties of this innovative estimator and the existing ones. See Section 3 for the theoretical comparison. To empirically validate the performance of the new estimator, we conducted a rigorous Monte Carlo simulation study, described in Section 4. Through meticulous experimentation and analysis, we uncover valuable insights into its effectiveness and efficiency compared to existing estimators. Furthermore, to illustrate the practical relevance of the proposed estimator, we present a compelling numerical example in Section 5. Finally, in Section 6, we summarize the main findings of our research, highlighting the contributions of the new estimator and discussing its implications for future advancements in estimation techniques.

2. Review

Logistic regression is a powerful statistical tool for modeling the relationship between a binary response variable and one or more explanatory variables. In logistic regression, the binary response variable y_i is modeled using a Bernoulli distribution: $y_i \sim Be(\pi_i)$.

$$p(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \tag{1}$$

where $\pi_i = \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}} = \frac{1}{1+e^{-(x_i^T\beta)}}$, i = 1,2,...,n, x_i is the ith row of X, which is an $n \times (p+1)$ matrix of independent variables, and β is a $(p+1) \times 1$ vector of regression coefficients. The logistic regression model is popularly estimated using the MLE when there is no multicollinearity. The next section presents a concise overview of this method.

2.1. The Maximum Likelihood Estimator (MLE)

The MLE is frequently adopted for parameter estimation in the logistic regression model. However, this technique is susceptible to multicollinearity and outliers in the data, which can compromise the validity and reliability of the estimates. The MLE of β for the logistic regression model is as follows:

$$\hat{\beta}^{LMLE} = \left(X^T\hat{G}_nX\right)^{-1}X^T\hat{G}_n\hat{z} \tag{2}$$

where $\hat{G}_n = diag(\hat{\pi}_i(1-\hat{\pi}_i))$ and $\hat{z}_i = \log(\hat{\pi}_i) + \frac{y_i-\hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)}$.

The asymptotic covariance matrix is calculated as follows:

$$Cov\left(\hat{\beta}^{LMLE}\right) \approx \left(X^T\hat{G}_nX\right)^{-1} \tag{3}$$

The matrix mean squared error $(MMSE)$ of $\hat{\beta}^{LMLE}$ is obtained by

$$MMSE\left(\hat{\beta}^{LMLE}\right) = Cov\left(\hat{\beta}^{LMLE}\right) = \left(X^T\hat{G}_nX\right)^{-1} \tag{4}$$

The scalar mean squared error $(SMSE)$ of $\hat{\beta}^{LMLE}$ is obtained by

$$SMSE\left(\hat{\beta}^{LMLE}\right) = \text{trace}\left(X^T\hat{G}_nX\right)^{-1} = \sum_{j=1}^{p+1}\frac{1}{\lambda_j} \tag{5}$$

where λ_j is the jth eigenvalue of the $(X^T\hat{G}_nX)$ matrix, and $j = 1, 2, \ldots, p+1$.

Multicollinearity impacts the MLE in the regression model, and its presence can influence the variance in the regression parameters. In the subsequent section, we will discuss three alternative methods to the MLE with a single shrinkage parameter when dealing with multicollinearity.

2.2. Logistic Ridge Regression (LR)

Ridge regression is a powerful approach for addressing multicollinearity in regression analysis. It offers a reliable solution for estimating and interpreting regression parameters when multicollinearity exists in the regression models. The ridge regression estimator (RRE) provides an excellent alternative to the widely used MLE in linear and logistic regression models [3,17]. The logistic ridge estimator (LR) is defined as follows:

$$\hat{\beta}^{LR} = \left(X^T\hat{G}_nX + kI_p\right)^{-1}\left(X^T\hat{G}_nX\right)\hat{\beta}^{LMLE} \tag{6}$$

where I_p is the identity matrix, and k is the ridge parameter with $k \geq 0$, defined by Hoerl et al. [18] for the linear regression as

$$k = \frac{(p+1)\sigma^2}{\sum_{j=1}^{p+1} \alpha_j^2} \tag{7}$$

and $\alpha = Q^T \hat{\beta}^{LMLE}$, where Q is the eigenvector of $\left(X^T \hat{G}_n X\right)$. However, the logistic version is defined by Schaefer et al. [3] as follows:

$$k = \frac{(p+1)}{\sum_{j=1}^{p+1} \alpha_j^2} \tag{8}$$

The estimated covariance matrix is calculated as follows:

$$Cov\left(\hat{\beta}^{LR}\right) = \left(X^T \hat{G}_n X + kI_p\right)^{-1} \left(X^T \hat{G}_n X\right) \left(X^T \hat{G}_n X + kI_p\right)^{-1} \tag{9}$$

The bias of $\hat{\beta}^{LR}$ is obtained by

$$Bias\left(\hat{\beta}^{LR}\right) = E\left(\hat{\beta}^{LR}\right) - \beta = \left[\left(X^T \hat{G}_n X + kI_p\right)^{-1} \left(X^T \hat{G}_n X\right) - I_p\right]\beta \tag{10}$$

Thus, the matrix mean squared error $(MMSE)$ of $\hat{\beta}^{LR}$ can be written as

$$MMSE\left(\hat{\beta}^{LR}\right) = Cov\left(\hat{\beta}^{LR}\right) + bias\left(\hat{\beta}^{LR}\right) bias\left(\hat{\beta}^{LR}\right)' \tag{11}$$

$$\begin{aligned} MMSE\left(\hat{\beta}^{LR}\right) &= \left(X^T \hat{G}_n X + kI_p\right)^{-1} \left(X^T \hat{G}_n X\right) \left(X^T \hat{G}_n X + kI_p\right)^{-1} \\ &+ \left[\left(X^T \hat{G}_n X + kI_p\right)^{-1} \left(X^T \hat{G}_n X\right) - I_p\right] \beta\beta' \left[\left(X^T \hat{G}_n X + kI_p\right)^{-1} \left(X^T \hat{G}_n X\right) - I_p\right]' \end{aligned} \tag{12}$$

2.3. Logistic Liu Estimator (LL)

The Liu estimator is an alternative to the ridge estimator in the linear regression model [19]. The logistic Liu estimator (LL) [5] is expressed as follows:

$$\hat{\beta}^{LL} = \left(X^T \hat{G}_n X + I_p\right)^{-1} \left(X^T \hat{G}_n X + dI_p\right) \hat{\beta}^{LMLE} \tag{13}$$

We adopted the Liu parameter (d) suggested by Ozkale and Kaciranlar [20] and adapted to the logistic regression as follows:

$$d = \min\left(\frac{\alpha_j^2}{1/\lambda_j} + \alpha_j^2\right) \tag{14}$$

where min represents the minimum operator, and λ_j and α_j $(j = 1, 2, \ldots, p+1)$ are as defined before. The estimated covariance matrix is calculated as follows:

$$Cov\left(\hat{\beta}^{LL}\right) = \left(X^T \hat{G}_n X + I_p\right)^{-1} \left(X^T \hat{G}_n X + dI_p\right) \left(X^T \hat{G}_n X\right)^{-1} \left(X^T \hat{G}_n X + dI_p\right) \left(X^T \hat{G}_n X + I_p\right)^{-1} \tag{15}$$

The bias of $\hat{\beta}_{BLE}$ is obtained by

$$Bias\left(\hat{\beta}^{LL}\right) = E\left(\hat{\beta}^{LL}\right) - \beta = \left[\left(X^T \hat{G}_n X + I_p\right)^{-1} \left(X^T \hat{G}_n X + dI_p\right) - I_p\right]\beta \tag{16}$$

Thus, the matrix mean squared error $(MMSE)$ of $\hat{\beta}^{LL}$ can be written as

$$MMSE\left(\hat{\beta}^{LL}\right) = Cov\left(\hat{\beta}^{LL}\right) + Bias\left(\hat{\beta}^{LL}\right)Bias\left(\hat{\beta}^{LL}\right)' \tag{17}$$

$$\begin{aligned} MMSE\left(\hat{\beta}^{LL}\right) = {} & \left(X^T\hat{G}_nX + I_p\right)^{-1}\left(X^T\hat{G}_nX + dI_p\right)\left(X^T\hat{G}_nX\right)^{-1}\left(X^T\hat{G}_nX + dI_p\right)\left(X^T\hat{G}_nX + I_p\right)^{-1} \\ & + \left[\left(X^T\hat{G}_nX + I_p\right)^{-1}\left(X^T\hat{G}_nX + dI_p\right) - I_p\right]\beta\beta'\left[\left(X^T\hat{G}_nX + I_p\right)^{-1}\left(X^T\hat{G}_nX + dI_p\right) - I_p\right]' \end{aligned} \tag{18}$$

2.4. Logistic Kibria–Lukman Estimator (KL)

The Kibria–Lukman estimator (KL), introduced by Kibria and Lukman [21], stands out among the class of single-parameter estimators, particularly when compared to established methods like the ridge and Liu estimators. Its competitive performance makes it a noteworthy contribution to linear regression modeling. Building upon the success of the KL estimator for linear regression, Lukman et al. [9] recently introduced the logistic Kibria–Lukman estimator (LKL). The expression is as follows:

$$\hat{\beta}^{LKL} = \left(X^T\hat{G}_nX + kI_p\right)^{-1}\left(X^T\hat{G}_nX - kI_p\right)\hat{\beta}^{LMLE} \tag{19}$$

We adopted the ridge parameter (k) in Equation (8). The estimated covariance matrix is calculated as follows:

$$Cov\left(\hat{\beta}^{LKL}\right) = \left(X^T\hat{G}_nX + kI_p\right)^{-1}\left(X^T\hat{G}_nX - kI_p\right)\left(X^T\hat{G}_nX\right)^{-1}\left(X^T\hat{G}_nX - kI_p\right)\left(X^T\hat{G}_nX + kI_p\right)^{-1} \tag{20}$$

The bias of $\hat{\beta}^{LKL}$ is obtained by

$$Bias\left(\hat{\beta}^{LKL}\right) = E\left(\hat{\beta}^{LKL}\right) - \beta = \left[\left(X^T\hat{G}_nX + kI_p\right)^{-1}\left(X^T\hat{G}_nX - kI_p\right) - I_p\right]\beta \tag{21}$$

Thus, the matrix mean squared error $(MMSE)$ of $\hat{\beta}^{LKL}$ can be written as

$$MMSE\left(\hat{\beta}^{LKL}\right) = Cov\left(\hat{\beta}^{LKL}\right) + Bias\left(\hat{\beta}^{LKL}\right)Bias\left(\hat{\beta}^{LKL}\right)' \tag{22}$$

$$\begin{aligned} MMSE\left(\hat{\beta}^{LKL}\right) = {} & \left(X^T\hat{G}_nX + kI_p\right)^{-1}\left(X^T\hat{G}_nX - kI_p\right)\left(X^T\hat{G}_nX\right)^{-1}\left(X^T\hat{G}_nX - kI_p\right)\left(X^T\hat{G}_nX + kI_p\right)^{-1} \\ & + \left[\left(X^T\hat{G}_nX + kI_p\right)^{-1}\left(X^T\hat{G}_nX - kI_p\right) - I_p\right]\beta\beta'\left[\left(X^T\hat{G}_nX + kI_p\right)^{-1}\left(X^T\hat{G}_nX - kI_p\right) - I_p\right]' \end{aligned} \tag{23}$$

2.5. Outliers

Outliers are observations that significantly deviate from the rest of the data and can substantially impact statistical analysis. They can arise due to measurement errors, data entry errors, or genuine extreme values. When dealing with logistic regression, it is important to differentiate between different types of outliers that can occur in the Y-space (response variable), X-space (predictor variables), or both. In the case of binary logistic regression, where the response variable is binary, outliers in the Y-space can only occur as errors in the classification of the response variable. These outliers are also known as residual outliers or misclassification-type errors. For example, a misclassification-type outlier can happen when an observation with a true value of 0 is mistakenly classified as 1 or vice versa [22].

On the other hand, outliers in the X-space, known as leverage points, refer to extreme observations in the design space of predictor variables. A leverage point can be considered either good or bad. A good leverage point occurs when the response variable (Y) equals 1 with a high probability (large value of $p\ (Y = 1 \mid x_i)$) or when Y equals 0 with a low

probability (small value of $p\ (Y = 1 \mid x_i)$). Conversely, a bad leverage point occurs when the response variable has the opposite characteristics. The influence of extreme values in the design space (X) on maximum likelihood estimates has been studied by Victoria-Feser [23]. Additionally, the impact of misclassification errors in logistic regression has been explored [22,24]. Croux et al. [25] found that the most problematic outliers, called bad leverage points, are misclassified observations and exhibit outlying behavior in the design space of predictor variables. The consequences of an outlier in a model are as follows:

Influence on Parameter Estimates: Outliers can substantially affect the estimation of logistic regression coefficients. Since logistic regression models the relationship between independent variables and the probability of an event occurring, outliers with extreme values can pull the estimated coefficients toward them, leading to biased estimates. Consequently, the estimated effects of independent variables may be distorted, and the interpretation of their impact on the outcome variable may be misleading.

Impact on Model Fit: Outliers can also influence the goodness-of-fit statistics of the logistic regression model. Influential outliers can inflate model fit statistics such as deviance, chi-square, or likelihood ratio tests, making the model appear a better fit than it is. This can lead to overconfidence in the model's performance and inaccurate assessments of its predictive power.

Inaccurate Inference: Outliers might substantially impact the hypothesis testing and confidence intervals. Since outliers may affect parameter estimates and standard errors, statistical tests may yield incorrect results, leading to flawed inferences. Confidence intervals can also be biased, potentially giving misleading conclusions about the significance and precision of the estimated coefficients.

Model Assumptions: Outliers can violate the assumptions underlying logistic regression models. Logistic regression assumes that the relationship between independent variables and the log odds of the outcome variable is correctly specified.

2.6. Robust Regression

Robust regression is employed when the distribution of residuals deviates from normality or when outliers significantly impact the model. Robust regression is a valuable method for analyzing data affected by outliers, ensuring the resulting models remain reliable and unaffected by these influential observations. When researchers construct regression models and encounter violations of common regression assumptions, traditional transformations often prove inadequate in eliminating or mitigating the influence of outliers, leading to biased predictions. In such situations, robust regression emerges as the preferred method, capable of providing reliable results resilient to the influence of outliers [26].

The subsequent section will cover the application of robust regression in both the linear and the logistic regression models.

2.6.1. M-Estimator

An M-estimator is a robust statistical method for estimating unknown parameters in a linear regression model. It is adopted when dealing with outliers, as it is less influenced by them compared to the maximum likelihood estimates. The estimation process involves solving an optimization problem by minimizing a criterion function, typically represented as $\rho(u)$, which measures the discrepancy between the observed data and the model's predicted values [27]. The M-estimator selects parameter values that minimize the sum of the criterion function evaluated at each data point. Popular choices for the criterion function include Huber loss, Tukey's biweight, etc. This approach has found applications in

various fields, including statistics, econometrics, and machine learning, providing reliable estimates even in the presence of outliers.

2.6.2. MM-Estimator

The MM-estimator [12] is a distinct form of M-estimation that combines the desirable qualities of M-estimators, which possess high asymptotic relative efficiency, with the robustness exhibited by S-estimators. MM-estimation refers to the utilization of multiple M-estimation procedures during the computation of the estimator. Yohai [12] outlined three distinct stages that define an MM-estimator:

Stage 1 To obtain an initial estimate, a high breakdown estimator, denoted as $\widetilde{\beta}$, is employed. This estimator is chosen for its ability to handle outliers effectively while maintaining efficiency. Using this initial estimate, the residuals can be computed as follows: $r_i(\beta) = y_i - X_i^T\hat{\beta}$.

Stage 2 Using these residuals from the robust fit, where $\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i}{s}\right) = k$. k is a constant, and with the objective function ρ, an M-estimate of scale with a 50% breakdown point (BDP) is computed. This s($r_1\widetilde{\beta}$, $\ldots$, $r_n\widetilde{\beta}$) is denoted by s_n. The objective function used in this stage is labeled ρ_0.

Stage 3 The MM-estimator, $\widetilde{\beta}$, is defined as an M-estimator of β using a redescending score function, $\varphi_1(u) = \frac{\partial\rho_1(u)}{\partial u}$, and the scale estimate s_n obtained from stage 2. In other words, $\widetilde{\beta}$ is a solution to the following equation:

$$\sum_{i=1}^{n} x_{ij}\varphi_1\left(\frac{y_i - X_i^T\hat{\beta}}{s_n}\right) = 0, \quad j = 1, \ldots, p+1 \tag{24}$$

2.7. Robust Logistic Regression

Various robust estimators are available for the logistic regression model, with some readily available in statistical software packages [28,29]. Two of the estimators are discussed in the next subsection.

2.7.1. The Bianco and Yohai Estimator (BY)

Pregibon [24] introduced robust M-estimates as an alternative to the total deviance function by minimizing the weighted total deviance:

$$M(\beta) = \sum_{i=1}^{n}\rho\left(d^2(\pi_i(\beta), y_i)\right) \tag{25}$$

where $\rho(u)$ is an increasing Huber loss function. Deviance residuals represent the discrepancies between the predicted probabilities, obtained using regression coefficients β, and the observed values. They identify how well the logistic regression model fits the data, with positive residuals indicating overestimation and negative residuals indicating underestimation. Later, Bianco and Yohai [30] addressed the limitations of the existing estimator, which lacked the down-weighting of high leverage points and exhibited inconsistency. They proposed an improvement by minimizing the estimator, as follows:

$$M(\beta) = \sum_{i=1}^{n}\rho\left(d^2(\pi_i(\beta), y_i)\right) + q(\pi_i(\beta)) \tag{26}$$

The function $\rho(u)$ mentioned is a bounded, differentiable, and non-decreasing function, which is defined by

$$\rho(u) = \begin{cases} u - \left(\frac{x^2}{2k}\right), & for\ x \le k \\ s/2 & otherwise \end{cases} \tag{27}$$

where s is a positive number. The researchers defined $q(u) = v(u) + v(1-u)$ with $v(u) = 2\int_0^u \psi(-2\log t)dt$ and $\psi = \rho^T$.

2.7.2. Conditionally Unbiased Bounded Influence Estimator

Künsch et al. [31] improved the resilience of the maximum likelihood estimator by integrating the M-estimator with a distinct mathematical expression:

$$\sum_{i=1}^{n} \Psi(y_i, x_i, \beta) \tag{28}$$

where $\Psi : \mathcal{R}^{1+p+p} \rightarrow \mathcal{R}^p$ Rp such that

$$\mathrm{E}(\Psi(y_i, x, \beta)|x_i) = 0 \tag{29}$$

These estimators, abbreviated as CE, possess the property of Fisher consistency. The optimal score function Φ can be expressed in the following manner:

$$\Psi(y_i, x, \beta, b, B) = W(\beta, y, x, b, B)\left\{y - g\left(\beta^T x\right) - c\left(\beta^T x, \frac{b}{h(x,\beta)}\right)\right\} \tag{30}$$

They introduce a parameter b, representing the maximum permissible value for the measure of infinitesimal sensitivity. Alongside this, they incorporate a dispersion matrix denoted as B. Additionally, they define a function h(x, B) which serves as a leverage measure. The function W adjusts the influence of unusual observations and ensures that the function Ψ remains bounded. Consequently, the resulting M-estimator exhibits limited influence due to this adjustment. The function $c\left(\beta^T x, \frac{b}{h(x,\beta)}\right)$ serves as a bias-correction term, specifically chosen to satisfy the condition in Equation (29). Let us define the corrected residual in the following manner:

$$r(y, x, \beta, b, B) = y - g\left(\beta^T x\right) - c\left(\beta^T x, \frac{b}{h(x,\beta)}\right) \tag{31}$$

Thus, the weights are of the form

$$W(y, x, \beta, b, B) = W_b(r(y, x, \beta)h(x, B)) \tag{32}$$

where W_b is the Huber weight function given by

$$W_b(x) = min\left\{1, \frac{b}{|x|}\right\} \tag{33}$$

Additionally, like the Schweppe-type GM estimators, the weight matrix W is tailored to reduce the influence of observations with high values in the product of corrected residuals and leverage. This down-weighting scheme aims to mitigate the impact of influential observations during the estimation process. The dispersion matrix B fulfills certain requirements, as follows:

$$\mathrm{E}\left(\Psi(y, x, \beta, b, B)\Psi^T(y, x, \beta, b, B)\right) = B \tag{34}$$

For concise implementation instructions for the CE, please refer to Künsch et al. [31]. The next section presents the methodology employed in this research to address the challenges of multicollinearity and outliers in logistic regression models.

3. Methodology

The methodology employed in this study addressed the challenges of multicollinearity and outliers in logistic regression models.

3.1. Proposed Estimators

We obtained the proposed estimators for the logistic regression model by combining the ridge, Liu, and KL estimators with two robust estimators: the Bianco and Yohai (BY) estimator and the conditionally unbiased bounded influence estimator (CE). To create the LR-BY estimator, we integrate LR with the BY estimator, resulting in a hybrid approach denoted as the LR-BY estimator. The LR-BY estimator of β is defined as

$$\hat{\beta}_{BY}^{LR} = \left(X^T \hat{G}_n X + k_{m1} I_p\right)^{-1} X^T \hat{G}_n X \hat{\beta}_{BY} \tag{35}$$

where $k_{m1} = \frac{1}{\max\left(\hat{\alpha}_{j,BY}^2\right)}$.

Similarly, we propose the LR-CE estimator, which combines the ridge estimator with the conditionally unbiased bounded influence estimator. The LR-CE estimator of β is defined as:

$$\hat{\beta}_{CE}^{LR} = \left(X^T \hat{G}_n X + k_{m2} I_p\right)^{-1} X^T \hat{G}_n X \hat{\beta}_{CE} \tag{36}$$

where $k_{m2} = \frac{1}{\max\left(\hat{\alpha}_{j,CE}^2\right)}$. The robust Liu estimator is developed by integrating the Liu estimator with the BY estimator, resulting in the LL-BY estimator. The LL-BY estimator of β is defined as

$$\hat{\beta}_{BY}^{LL} = \left(X^T \hat{G}_n X + I_p\right)^{-1} \left(X^T \hat{G}_n X + d_{m1} I_p\right) \hat{\beta}_{BY} \tag{37}$$

where d_{m1} is the robust Liu parameter and is defined as follows:

$$d_{m1} = \min\left\{ \frac{\hat{\alpha}_{j,BY}^2}{\frac{1}{\lambda_j} + \hat{\alpha}_{j,BY}^2} \right\}.$$

Also, we propose another robust Liu estimator by integrating the Liu estimator with the CE, resulting in the LL-BY estimator. The LL-CE estimator of β is defined as

$$\hat{\beta}_{CE}^{LL} = \left(X^T \hat{G}_n X + I_p\right)^{-1} \left(X^T \hat{G}_n X + d_{m2} I_p\right) \hat{\beta}_{CE} \tag{38}$$

where d_{m2} is the robust Liu parameter and is defined as follows:

$$d_{m2} = \min\left\{ \frac{\hat{\alpha}_{j,CE}^2}{\frac{1}{\lambda_j} + \hat{\alpha}_{j,CE}^2} \right\}.$$

We combine the KL estimator with the BY estimator, leading to a hybrid approach denoted as the KL-BY estimator. The KL-BY estimator of β is defined as

$$\hat{\beta}_{BY}^{KL} = \left(X^T \hat{G}_n X + k_{m1} I_p\right)^{-1} \left(X^T \hat{G}_n X + k_{m1} I_p\right) \hat{\beta}_{BY} \tag{39}$$

where $k_{m1} = \frac{1}{\max\left(\hat{\alpha}_{j,BY}^2\right)}$.

Likewise, we suggest the KL-CE estimator, which combines the KL estimator with the conditionally unbiased bounded influence estimator. The LK-CE estimator of β is defined as

$$\hat{\beta}_{CE}^{KL} = \left(X^T \hat{G}_n X + k_{m2} I_p\right)^{-1} \left(X^T \hat{G}_n X + k_{m2} I_p\right) \hat{\beta}_{CE} \tag{40}$$

where $k_{m2} = \frac{1}{\max\left(\hat{\alpha}_{j,CE}^2\right)}$.

We conduct extensive simulations and real-life data analysis to evaluate the efficacy of the proposed estimators.

3.2. Scalar Mean Squared Errors of the Estimators

We retain the spectral decomposition of the estimated information matrix $(X^T\hat{G}_nX)$ to offer the precise form of the matrix mean squared errors (MMSEs) and the scalar mean squared errors (SMSEs) for the previously stated biased estimator. Assume that a matrix Q exists such that $\hat{\alpha}^{LMLE} = Q^T\hat{\beta}^{LMLE}$, where $Q^TX^T\hat{G}_nXQ = \Lambda = \text{diag}(\lambda_1, \ldots \lambda_{p+1})$. Here, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{p+1}$, and Λ represents the matrix of eigenvalues of $(X^T\hat{G}_nX)$, while Q is a matrix in which the columns are the eigenvectors of $(X^T\hat{G}_nX)$. It is important to note that there exists a relationship between the estimator α and β, namely $\hat{\alpha} = Q^T\hat{\beta}^{LMLE}$, resulting in $SMSE(\hat{\alpha}^{LMLE}) = SMSE(\hat{\beta}^{LMLE})$. Consequently, it is sufficient to focus on the canonical form exclusively. Equations (41)–(50) provide the scalar mean squared error of the following estimators: MLE, LR, LL, LL-BY, LL-CE, LR-BY, LR-CE, KL, KL-BY, and KL-CE. The SMSEs of the mentioned estimators are as follows:

$$SMSE\left(\hat{\alpha}^{LMLE}\right) = \sum_{j=1}^{p^*} \frac{1}{\lambda_j} \tag{41}$$

$$SMSE\left(\hat{\alpha}^{LL}\right) = \sum_{j=1}^{p^*} \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + \sum_{j=1}^{p^*} \frac{(d-1)^2\alpha_j^2}{(\lambda_j + 1)^2} \tag{42}$$

$$SMSE\left(\hat{\alpha}_{BY}^{LL}\right) = \sum_{j=1}^{p^*} \frac{(\lambda_j + d_{m1})^2}{\lambda_j(\lambda_j + 1)^2} + \sum_{j=1}^{p^*} \frac{(d_{m1}-1)^2\alpha_{j,BY}^2}{(\lambda_j + 1)^2} \tag{43}$$

$$SMSE\left(\hat{\alpha}_{CE}^{LL}\right) = \sum_{j=1}^{p^*} \frac{(\lambda_j + d_{m2})^2}{\lambda_j(\lambda_j + 1)^2} + \sum_{j=1}^{p^*} \frac{(d_{m2}-1)^2\alpha_{j,CE}^2}{(\lambda_j + 1)^2} \tag{44}$$

$$SMSE\left(\hat{\alpha}^{LR}\right) = \sum_{j=1}^{p^*} \frac{\lambda_j}{(\lambda_j + k)^2} + \sum_{j=1}^{p^*} \frac{\alpha_j^2}{(\lambda_j + k)^2} \tag{45}$$

$$SMSE\left(\hat{\alpha}_{BY}^{LR}\right) = \sum_{j=1}^{p^*} \frac{\lambda_j}{(\lambda_j + k_{m1})^2} + \sum_{j=1}^{p^*} \frac{\alpha_{j,BY}^2}{(\lambda_j + k_{m1})^2} \tag{46}$$

$$SMSE\left(\hat{\alpha}_{CE}^{LR}\right) = \sum_{j=1}^{p^*} \frac{\lambda_j}{(\lambda_j + k_{m2})^2} + \sum_{j=1}^{p^*} \frac{\alpha_{j,CE}^2}{(\lambda_j + k_{m2})^2} \tag{47}$$

$$SMSE\left(\hat{\alpha}^{KL}\right) = \sum_{j=1}^{p^*} \frac{(\lambda_j - k)^2}{\lambda_j(\lambda_j + k)^2} + 4k^2\sum_{j=1}^{p^*} \frac{\alpha_j^2}{(\lambda_j + k)^2} \tag{48}$$

$$SMSE\left(\hat{\alpha}_{BY}^{KL}\right) = \sum_{j=1}^{p^*} \frac{(\lambda_j - k_{m1})^2}{\lambda_j(\lambda_j + k_{m1})^2} + 4k_{m1}^2\sum_{j=1}^{p^*} \frac{\alpha_{j,BY}^2}{(\lambda_j + k_{m1})^2} \tag{49}$$

$$SMSE\left(\hat{\alpha}_{CE}^{KL}\right) = \sum_{j=1}^{p^*} \frac{(\lambda_j - k_{m2})^2}{\lambda_j(\lambda_j + k_{m2})^2} + 4k_{m2}^2\sum_{j=1}^{p^*} \frac{\alpha_{j,CE}^2}{(\lambda_j + k_{m2})^2} \tag{50}$$

3.3. Theoretical Comparisons Between the Estimators

3.3.1. The $\hat{\alpha}_{BY}^{LR}$ Estimator and the ML Estimator

The estimator $\hat{\alpha}_{BY}^{LR}$ is superior to the estimator $\hat{\alpha}^{MLE}$ in the sense of the SMSE criterion, i.e., $SMSE(\hat{\alpha}_{BY}^{LR}) - SMSE(\hat{\alpha}^{MLE}) < 0$ if $(\lambda_j + k_{m1})^2 > \lambda_j\left(\lambda_j + \alpha_{j,BY}^2\right)$ and $k_{m1} > 0$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{BY}^{LR}\right)$ and $SMSE\left(\hat{\alpha}^{MLE}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{BY}^{LR}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_{j,BY}^2\right)}{\left(\lambda_j + k_{m1}\right)^2} - \sum_{j=1}^{p^*} \frac{1}{\lambda_j} \tag{51}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\lambda_j\left(\lambda_j + \alpha_{j,BY}^2\right) - \left(\lambda_j + k_{m1}\right)^2}{\lambda_j\left(\lambda_j + k_{m1}\right)^2} \tag{52}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j + k_{m1}\right)^2 > \lambda_j\left(\lambda_j + \alpha_{j,BY}^2\right)$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{LR}$ is superior to the estimator $\hat{\alpha}^{MLE}$ when $k_{m1} > 0$ for all $j = 1, 2, \ldots, p^*$. □

3.3.2. The $\hat{\alpha}_{CE}^{LR}$ Estimator and the ML Estimator

The estimator $\hat{\alpha}_{CE}^{LR}$ is superior to the estimator $\hat{\alpha}^{MLE}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{CE}^{LR}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) < 0$ if $\left(\lambda_j + k_{m2}\right)^2 > \lambda_j\left(\lambda_j + \alpha_{j,CE}^2\right)$ and $k_{m2} > 0$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{CE}^{LR}\right)$ and $SMSE\left(\hat{\alpha}^{MLE}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{CE}^{LR}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_{j,\mathrm{CE}}^2\right)}{\left(\lambda_j + k_{m2}\right)^2} - \sum_{j=1}^{p^*} \frac{1}{\lambda_j} \tag{53}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\lambda_j\left(\lambda_j + \alpha_{j,\mathrm{CE}}^2\right) - \left(\lambda_j + k_{m2}\right)^2}{\lambda_j\left(\lambda_j + k_{m2}\right)^2} \tag{54}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j + k_{m2}\right)^2 > \lambda_j\left(\lambda_j + \alpha_{j,\mathrm{CE}}^2\right)$, and then it can be observed that the estimator $\hat{\alpha}_{CE}^{LR}$ is superior to the estimator $\hat{\alpha}^{MLE}$ when $k_{m2} > 0$ for all $j = 1, 2, \ldots, p^*$. □

3.3.3. The $\hat{\alpha}_{BY}^{LL}$ Estimator and the ML Estimator

The estimator $\hat{\alpha}_{BY}^{LL}$ is superior to the estimator $\hat{\alpha}^{MLE}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{BY}^{LL}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) < 0$ if $\left(\lambda_j + 1\right)^2 > \left(\lambda_j + d_{m1}\right)^2 + (d_{m1} - 1)^2\alpha_{j,BY}^2$ and $0 \leq d_{m1} < 1$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{BY}^{LL}\right)$ and $SMSE\left(\hat{\alpha}^{MLE}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{BY}^{LL}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + d_{m1}\right)^2 + (d_{m1} - 1)^2\alpha_{j,BY}^2}{\lambda_j\left(\lambda_j + 1\right)^2} - \sum_{j=1}^{p^*} \frac{1}{\lambda_j} \tag{55}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + d_{m1}\right)^2 + (d_{m1} - 1)^2\alpha_{j,BY}^2 - \left(\lambda_j + 1\right)^2}{\lambda_j\left(\lambda_j + 1\right)^2} \tag{56}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j + 1\right)^2 > \left(\lambda_j + d_{m1}\right)^2 + (d_{m1} - 1)^2\alpha_{j,BY}^2$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{LL}$ is superior to the estimator $\hat{\alpha}^{MLE}$ when $0 \leq d_{m1} < 1$ for all $j = 1, 2, \ldots, p^*$. □

3.3.4. The $\hat{\alpha}_{CE}^{LL}$ Estimator and the ML Estimator

The estimator $\hat{\alpha}_{CE}^{LL}$ is superior to the estimator $\hat{\alpha}^{MLE}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{CE}^{LL}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) < 0$ if $\left(\lambda_j + 1\right)^2 > \left(\lambda_j + d_{m2}\right)^2 + (d_{m2} - 1)^2\alpha_{j,CE}^2$ and $0 \leq d_{m2} < 1$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{CE}^{LL}\right)$ and $SMSE\left(\hat{\alpha}^{MLE}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{CE}^{LL}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + d_{m2}\right)^2 + \left(d_{m2} - 1\right)^2 \alpha_{j,CE}^2}{\lambda_j \left(\lambda_j + 1\right)^2} - \sum_{j=1}^{p^*} \frac{1}{\lambda_j} \tag{57}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + d_{m2}\right)^2 + \left(d_{m2} - 1\right)^2 \alpha_{j,CE}^2 - \left(\lambda_j + 1\right)^2}{\lambda_j \left(\lambda_j + 1\right)^2} \tag{58}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j + 1\right)^2 > \left(\lambda_j + d_{m2}\right)^2 + \left(d_{m2} - 1\right)^2 \alpha_{j,CE}^2$, and then it can be observed that the estimator $\hat{\alpha}_{CE}^{LL}$ is superior to the estimator $\hat{\alpha}^{MLE}$ when $0 \leq d_{m2} < 1$ for all $j = 1, 2, \ldots, p^*$. □

3.3.5. The $\hat{\alpha}_{BY}^{KL}$ Estimator and the ML Estimator

The estimator $\hat{\alpha}_{BY}^{KL}$ is superior to the estimator $\hat{\alpha}^{MLE}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{BY}^{KL}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) < 0$ if $\left(\lambda_j + k_{m1}\right)^2 > \left(\lambda_j - k_{m1}\right)^2 + 4k_{m1}^2 \alpha_{j,BY}^2$ and $k_{m1} > 0$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{BY}^{KL}\right)$ and $SMSE\left(\hat{\alpha}^{MLE}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{BY}^{KL}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j - k_{m1}\right)^2 + 4k_{m1}^2 \alpha_{j,BY}^2}{\lambda_j \left(\lambda_j + k_{m1}\right)^2} - \sum_{j=1}^{p^*} \frac{1}{\lambda_j} \tag{59}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j - k_{m1}\right)^2 + 4k_{m1}^2 \alpha_{j,BY}^2 - \left(\lambda_j + k_{m1}\right)^2}{\lambda_j \left(\lambda_j + k_{m1}\right)^2} \tag{60}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j + k_{m1}\right)^2 > \left(\lambda_j - k_{m1}\right)^2 + 4k_{m1}^2 \alpha_{j,BY}^2$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{KL}$ is superior to the estimator $\hat{\alpha}^{MLE}$ when $k_{m1} > 0$ for all $j = 1, 2, \ldots, p^*$. □

3.3.6. The $\hat{\alpha}_{CE}^{KL}$ Estimator and the ML Estimator

The estimator $\hat{\alpha}_{CE}^{KL}$ is superior to the estimator $\hat{\alpha}^{MLE}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{CE}^{KL}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) < 0$ if $\left(\lambda_j + k_{m2}\right)^2 > \left(\lambda_j - k_{m2}\right)^2 + 4k_{m2}^2 \alpha_{j,\mathrm{CE}}^2$ and $k_{m2} > 0$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{CE}^{KL}\right)$ and $SMSE\left(\hat{\alpha}^{MLE}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{CE}^{KL}\right) - SMSE\left(\hat{\alpha}^{MLE}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j - k_{m2}\right)^2 + 4k_{m2}^2 \alpha_{j,\mathrm{CE}}^2}{\lambda_j \left(\lambda_j + k_{m2}\right)^2} - \sum_{j=1}^{p^*} \frac{1}{\lambda_j} \tag{61}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j - k_{m2}\right)^2 + 4k_{m2}^2 \alpha_{j,\mathrm{CE}}^2 - \left(\lambda_j + k_{m2}\right)^2}{\lambda_j \left(\lambda_j + k_{m2}\right)^2} \tag{62}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j + k_{m2}\right)^2 > \left(\lambda_j - k_{m2}\right)^2 + 4k_{m2}^2 \alpha_{j,\mathrm{CE}}^2$, and then it can be observed that the estimator $\hat{\alpha}_{CE}^{KL}$ is superior to the estimator $\hat{\alpha}^{MLE}$ when $k_{m2} > 0$ for all $j = 1, 2, \ldots, p^*$. □

3.3.7. The $\hat{\alpha}_{BY}^{LR}$ Estimator and the $\hat{\alpha}^{LR}$ Estimator

The estimator $\hat{\alpha}_{BY}^{LR}$ is superior to the estimator $\hat{\alpha}^{LR}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{BY}^{LR}\right) - SMSE\left(\hat{\alpha}^{LR}\right) < 0$ if $\left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + k_{m1}\right)^2 > \left(\lambda_j + \alpha_{j,BY}^2\right)\left(\lambda_j + k\right)^2$ and $0 < k < k_{m1}$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{BY}^{LR}\right)$ and $SMSE\left(\hat{\alpha}^{LR}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{BY}^{LR}\right) - SMSE\left(\hat{\alpha}^{LR}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_{j,BY}^2\right)}{\left(\lambda_j + k_{m1}\right)^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_j^2\right)}{\left(\lambda_j + k\right)^2} \tag{63}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_{j,BY}^2\right)\left(\lambda_j + k\right)^2 - \left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + k_{m1}\right)^2}{\left(\lambda_j + k\right)^2\left(\lambda_j + k_{m1}\right)^2} \tag{64}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + k_{m1}\right)^2 > \left(\lambda_j + \alpha_{j,BY}^2\right)\left(\lambda_j + k\right)^2$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{LR}$ is superior to the estimator $\hat{\alpha}^{LR}$ when $0 < k < k_{m1}$ for all $j = 1, 2, \ldots, p^*$. □

3.3.8. The $\hat{\alpha}_{CE}^{LR}$ Estimator and the $\hat{\alpha}^{LR}$ Estimator

The estimator $\hat{\alpha}_{CE}^{LR}$ is superior to the estimator $\hat{\alpha}^{LR}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{CE}^{LR}\right) - SMSE\left(\hat{\alpha}^{LR}\right) < 0$ if $\left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + k_{m2}\right)^2 > \left(\lambda_j + \alpha_{j,\mathrm{CE}}^2\right)\left(\lambda_j + k\right)^2$ and $0 < k < k_{m2}$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{CE}^{LR}\right)$ and $SMSE\left(\hat{\alpha}^{LR}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{CE}^{LR}\right) - SMSE\left(\hat{\alpha}^{LR}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_{j,\mathrm{CE}}^2\right)}{\left(\lambda_j + k_{m2}\right)^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_j^2\right)}{\left(\lambda_j + k\right)^2} \tag{65}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_{j,\mathrm{CE}}^2\right)\left(\lambda_j + k\right)^2 - \left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + k_{m2}\right)^2}{\left(\lambda_j + k\right)^2\left(\lambda_j + k_{m2}\right)^2} \tag{66}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + k_{m2}\right)^2 > \left(\lambda_j + \alpha_{j,\mathrm{CE}}^2\right)\left(\lambda_j + k\right)^2$, and then it can be observed that the estimator $\hat{\alpha}_{CE}^{LR}$ is superior to the estimator $\hat{\alpha}^{LR}$ when $0 < k < k_{m2}$ for all $j = 1, 2, \ldots, p^*$. □

3.3.9. The $\hat{\alpha}_{BY}^{LL}$ Estimator and the $\hat{\alpha}^{LR}$ Estimator

The estimator $\hat{\alpha}_{BY}^{LL}$ is superior to the estimator $\hat{\alpha}^{LR}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{BY}^{LL}\right) - SMSE\left(\hat{\alpha}^{LR}\right) < 0$ if $\lambda_j\left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + 1\right)^2 > \left(\lambda_j + d_{m1}\right)^2\left(\lambda_j + k\right)^2 + \left(\lambda_j + k\right)^2(d_{m1} - 1)^2\alpha_{j,BY}^2\lambda_j$ and $0 < k$, $0 \leq d_{m1} < 1$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{BY}^{LL}\right)$ and $SMSE\left(\hat{\alpha}^{LR}\right)$ is as follows:

$$SMSE\left(\hat{\alpha}_{BY}^{LL}\right) - SMSE\left(\hat{\alpha}^{LR}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + d_{m1}\right)^2 + (d_{m1} - 1)^2\alpha_{j,BY}^2\lambda_j}{\lambda_j\left(\lambda_j + 1\right)^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j + \alpha_j^2\right)}{\left(\lambda_j + k\right)^2} \tag{67}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j + d_{m1}\right)^2\left(\lambda_j + k\right)^2 + \left(\lambda_j + k\right)^2(d_{m1} - 1)^2\alpha_{j,BY}^2\lambda_j - \lambda_j\left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + 1\right)^2}{\lambda_j\left(\lambda_j + 1\right)^2\left(\lambda_j + k\right)^2} \tag{68}$$

The difference $\nabla < 0$ occurs if
$\lambda_j\left(\lambda_j + \alpha_j^2\right)\left(\lambda_j + 1\right)^2 > \left(\lambda_j + d_{m1}\right)^2\left(\lambda_j + k\right)^2 + \left(\lambda_j + k\right)^2(d_{m1} - 1)^2\alpha_{j,BY}^2\lambda_j$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{LL}$ is superior to the estimator $\hat{\alpha}^{LR}$ when $0 < k$, $0 \leq d_{m1} < 1$ for all $j = 1, 2, \ldots, p^*$. □

3.3.10. The $\hat{\alpha}_{CE}^{LL}$ Estimator and the $\hat{\alpha}^{LR}$ Estimator

The estimator $\hat{\alpha}_{CE}^{LL}$ is superior to the estimator $\hat{\alpha}^{LR}$ in the sense of the SMSE criterion, i.e., $SMSE(\hat{\alpha}_{CE}^{LL}) - SMSE(\hat{\alpha}^{LR}) < 0$ if $\lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+1)^2 > (\lambda_j+d_{m2})^2(\lambda_j+k)^2 + (\lambda_j+k)^2(d_{m2}-1)^2\alpha_{j,CE}^2\lambda_j$ and $0<k,\ 0\le d_{m2}<1$.

Proof. The difference between $SMSE(\hat{\alpha}_{CE}^{LL})$ and $SMSE(\hat{\alpha}^{LR})$ is as follows:

$$SMSE\left(\hat{\alpha}_{CE}^{LL}\right) - SMSE\left(\hat{\alpha}^{LR}\right) = \nabla = \sum_{j=1}^{p^*} \frac{(\lambda_j+d_{m2})^2 + (d_{m2}-1)^2\alpha_{j,CE}^2\lambda_j}{\lambda_j(\lambda_j+1)^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j+\alpha_j^2\right)}{(\lambda_j+k)^2} \tag{69}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{(\lambda_j+d_{m2})^2(\lambda_j+k)^2 + (\lambda_j+k)^2(d_{m2}-1)^2\alpha_{j,CE}^2\lambda_j - \lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+1)^2}{\lambda_j(\lambda_j+1)^2(\lambda_j+k)^2} \tag{70}$$

The difference $\nabla < 0$ occurs if
$\lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+1)^2 > (\lambda_j+d_{m2})^2(\lambda_j+k)^2 + (\lambda_j+k)^2(d_{m2}-1)^2\alpha_{j,CE}^2\lambda_j$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{LL}$ is superior to the estimator $\hat{\alpha}^{LR}$ when $0<k$, $0\le d_{m2}<1$ for all $j=1,2,\ldots,p^*$. □

3.3.11. The $\hat{\alpha}_{BY}^{KL}$ Estimator and the $\hat{\alpha}^{LR}$ Estimator

The estimator $\hat{\alpha}_{BY}^{KL}$ is superior to the estimator $\hat{\alpha}^{LR}$ in the sense of the SMSE criterion, i.e., $SMSE(\hat{\alpha}_{BY}^{KL}) - SMSE(\hat{\alpha}^{LR}) < 0$ if $\lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+k_{m1})^2 > (\lambda_j-k_{m1})^2(\lambda_j+k)^2 + 4k_{m1}^2\alpha_{j,BY}^2(\lambda_j+k)^2$ and $0<k<k_{m1}$.

Proof. The difference between $SMSE(\hat{\alpha}_{BY}^{KL})$ and $SMSE(\hat{\alpha}^{LR})$ is as follows:

$$SMSE\left(\hat{\alpha}_{BY}^{KL}\right) - SMSE\left(\hat{\alpha}^{LR}\right) = \nabla = \sum_{j=1}^{p^*} \frac{(\lambda_j-k_{m1})^2 + 4k_{m1}^2\alpha_{j,BY}^2}{\lambda_j(\lambda_j+k_{m1})^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j+\alpha_j^2\right)}{(\lambda_j+k)^2} \tag{71}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{(\lambda_j-k_{m1})^2(\lambda_j+k)^2 + 4k_{m1}^2\alpha_{j,BY}^2(\lambda_j+k)^2 - \lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+k_{m1})^2}{\lambda_j(\lambda_j+k_{m1})^2(\lambda_j+k)^2} \tag{72}$$

The difference $\nabla < 0$ occurs if
$\lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+k_{m1})^2 > (\lambda_j-k_{m1})^2(\lambda_j+k)^2 + 4k_{m1}^2\alpha_{j,BY}^2(\lambda_j+k)^2$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{KL}$ is superior to the estimator $\hat{\alpha}^{LR}$ when $0<k<k_{m1}$ for all $j=1,2,\ldots,p^*$. □

3.3.12. The $\hat{\alpha}_{CE}^{KL}$ Estimator and the $\hat{\alpha}^{LR}$ Estimator

The estimator $\hat{\alpha}_{CE}^{KL}$ is superior to the estimator $\hat{\alpha}^{LR}$ in the sense of the SMSE criterion, i.e., $SMSE(\hat{\alpha}_{CE}^{KL}) - SMSE(\hat{\alpha}^{LR}) < 0$ if $\lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+k_{m2})^2 > (\lambda_j-k_{m2})^2(\lambda_j+k)^2 + 4k_{m2}^2\alpha_{j,CE}^2(\lambda_j+k)^2$ and $0<k<k_{m2}$.

Proof. The difference between $SMSE(\hat{\alpha}_{CE}^{KL})$ and $SMSE(\hat{\alpha}^{LR})$ is as follows:

$$SMSE\left(\hat{\alpha}_{CE}^{KL}\right) - SMSE\left(\hat{\alpha}^{LR}\right) = \nabla = \sum_{j=1}^{p^*} \frac{(\lambda_j-k_{m2})^2 + 4k_{m2}^2\alpha_{j,\mathrm{CE}}^2}{\lambda_j(\lambda_j+k_{m2})^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j+\alpha_j^2\right)}{(\lambda_j+k)^2} \tag{73}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{(\lambda_j-k_{m2})^2(\lambda_j+k)^2 + 4k_{m2}^2\alpha_{j,CE}^2(\lambda_j+k)^2 - \lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+k_{m2})^2}{\lambda_j(\lambda_j+k_{m2})^2(\lambda_j+k)^2} \tag{74}$$

The difference $\nabla < 0$ occurs if
$\lambda_j\left(\lambda_j+\alpha_j^2\right)\left(\lambda_j+k_{m2}\right)^2 > \left(\lambda_j-k_{m2}\right)^2\left(\lambda_j+k\right)^2 + 4k_{m2}^2\alpha_{j,CE}^2\left(\lambda_j+k\right)^2$, and then it can be observed that the estimator $\hat{\alpha}_{CE}^{KL}$ is superior to the estimator $\hat{\alpha}^{LR}$ when $0 < k < k_{m2}$ for all $j = 1, 2, \ldots, p^*$. □

3.3.13. The $\hat{\alpha}_{BY}^{LL}$ Estimator and the $\hat{\alpha}^{KL}$ Estimator

The estimator $\hat{\alpha}_{BY}^{LL}$ is superior to the estimator $\hat{\alpha}^{KL}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{BY}^{LL}\right) - SMSE\left(\hat{\alpha}^{KL}\right) < 0$ if $\left(\lambda_j+1\right)^2\left(\lambda_j-k\right)^2 + \left(\lambda_j+1\right)^2 4k^2\alpha_j^2 > \left(\lambda_j+d_{m1}\right)^2\left(\lambda_j+k\right)^2 + \left(\lambda_j+k\right)^2(d_{m1}-1)^2\alpha_{j,BY}^2$ and $k > 0$ and $0 \le d_{m1} < 1$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{BY}^{LL}\right)$ and $SMSE\left(\hat{\alpha}^{KL}\right)$ is as follows:

$$MSE\left(\hat{\alpha}_{BY}^{LL}\right) - SMSE\left(\hat{\alpha}^{KL}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j+d_{m1}\right)^2 + (d_{m1}-1)^2\alpha_{j,BY}^2}{\lambda_j\left(\lambda_j+1\right)^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j-k\right)^2 + 4k^2\alpha_j^2}{\lambda_j\left(\lambda_j+k\right)^2} \tag{75}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j+d_{m1}\right)^2\left(\lambda_j+k\right)^2 + \left(\lambda_j+k\right)^2(d_{m1}-1)^2\alpha_{j,BY}^2 - \left(\lambda_j+1\right)^2\left(\lambda_j-k\right)^2 + \left(\lambda_j+1\right)^2 4k^2\alpha_j^2}{\lambda_j\left(\lambda_j+1\right)^2\left(\lambda_j+k\right)^2} \tag{76}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j+1\right)^2\left(\lambda_j-k\right)^2 + \left(\lambda_j+1\right)^2 4k^2\alpha_j^2 > \left(\lambda_j+d_{m1}\right)^2 \left(\lambda_j+k\right)^2 + \left(\lambda_j+k\right)^2(d_{m1}-1)^2\alpha_{j,BY}^2$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{LL}$ is superior to the estimator $\hat{\alpha}^{KL}$ when $k > 0$ and $0 \le d_{m1} < 1$ for all $j = 1, 2, \ldots, p^*$. □

3.3.14. The $\hat{\alpha}_{CE}^{LL}$ Estimator and the $\hat{\alpha}^{KL}$ Estimator

The estimator $\hat{\alpha}_{CE}^{LL}$ is superior to the estimator $\hat{\alpha}^{KL}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{CE}^{LL}\right) - SMSE\left(\hat{\alpha}^{KL}\right) < 0$ if $\left(\lambda_j+1\right)^2\left(\lambda_j-k\right)^2 + \left(\lambda_j+1\right)^2 4k^2\alpha_j^2 > \left(\lambda_j+d_{m2}\right)^2\left(\lambda_j+k\right)^2 + \left(\lambda_j+k\right)^2(d_{m2}-1)^2\alpha_{j,CE}^2$ and $k > 0$ and $0 \le d_{m2} < 1$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{CE}^{LL}\right)$ and $SMSE\left(\hat{\alpha}^{KL}\right)$ is as follows:

$$MSE\left(\hat{\alpha}_{CE}^{LL}\right) - SMSE\left(\hat{\alpha}^{KL}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j+d_{m2}\right)^2 + (d_{m2}-1)^2\alpha_{j,CE}^2}{\lambda_j\left(\lambda_j+1\right)^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j-k\right)^2 + 4k^2\alpha_j^2}{\lambda_j\left(\lambda_j+k\right)^2} \tag{77}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j+d_{m2}\right)^2\left(\lambda_j+k\right)^2 + \left(\lambda_j+k\right)^2(d_{m2}-1)^2\alpha_{j,CE}^2 - \left(\lambda_j+1\right)^2\left(\lambda_j-k\right)^2 + \left(\lambda_j+1\right)^2 4k^2\alpha_j^2}{\lambda_j\left(\lambda_j+1\right)^2\left(\lambda_j+k\right)^2} \tag{78}$$

The difference $\nabla < 0$ occurs if
$\left(\lambda_j+1\right)^2\left(\lambda_j-k\right)^2 + \left(\lambda_j+1\right)^2 4k^2\alpha_j^2 > \left(\lambda_j+d_{m2}\right)^2\left(\lambda_j+k\right)^2 + \left(\lambda_j+k\right)^2(d_{m2}-1)^2$ $\alpha_{j,CE}^2$, and then it can be observed that the estimator $\hat{\alpha}_{CE}^{LL}$ is superior to the estimator $\hat{\alpha}^{KL}$ when $k > 0$ and $0 \le d_{m2} < 1$ for all $j = 1, 2, \ldots, p^*$. □

3.3.15. The $\hat{\alpha}_{BY}^{LR}$ Estimator and the $\hat{\alpha}^{KL}$ Estimator

The estimator $\hat{\alpha}_{BY}^{LR}$ is superior to the estimator $\hat{\alpha}^{KL}$ in the sense of the SMSE criterion, i.e., $SMSE\left(\hat{\alpha}_{BY}^{LR}\right) - SMSE\left(\hat{\alpha}^{KL}\right) < 0$ if $\left(\lambda_j-k\right)^2\left(\lambda_j+k_{m1}\right)^2 + \left(\lambda_j+k_{m1}\right)^2 4k^2\alpha_j^2 > \lambda_j\left(\lambda_j+k\right)^2\left(\lambda_j+\alpha_{j,BY}^2\right)$ and $0 < k < k_{m1}$.

Proof. The difference between $SMSE\left(\hat{\alpha}_{BY}^{LR}\right)$ and $SMSE\left(\hat{\alpha}^{KL}\right)$ is as follows:

$$MSE\left(\hat{\alpha}_{BY}^{LR}\right) - SMSE\left(\hat{\alpha}^{KL}\right) = \nabla = \sum_{j=1}^{p^*}\sum_{j=1}^{p^*} \frac{\left(\lambda_j+\alpha_{j,BY}^2\right)}{\left(\lambda_j+k_{m1}\right)^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j-k\right)^2 + 4k^2\alpha_j^2}{\lambda_j\left(\lambda_j+k\right)^2} \tag{79}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\lambda_j(\lambda_j+k)^2\left(\lambda_j+\alpha_{j,BY}^2\right) - (\lambda_j-k)^2(\lambda_j+k_{m1})^2 + (\lambda_j+k_{m1})^2 4k^2\alpha_j^2}{\lambda_j(\lambda_j+k_{m1})^2(\lambda_j+k)^2} \tag{80}$$

The difference $\nabla < 0$ occurs if $(\lambda_j-k)^2(\lambda_j+k_{m1})^2 + (\lambda_j+k_{m1})^2 4k^2\alpha_j^2 > \lambda_j(\lambda_j+k)^2\left(\lambda_j+\alpha_{j,BY}^2\right)$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{LR}$ is superior to the estimator $\hat{\alpha}^{KL}$ when $0 < k < k_{m1}$ for all $j = 1, 2, \ldots, p^*$. □

3.3.16. The $\hat{\alpha}_{CE}^{LR}$ Estimator and the $\hat{\alpha}^{KL}$ Estimator

The estimator $\hat{\alpha}_{CE}^{LR}$ is superior to the estimator $\hat{\alpha}^{KL}$ in the sense of the SMSE criterion, i.e., $SMSE(\hat{\alpha}_{CE}^{LR}) - SMSE(\hat{\alpha}^{KL}) < 0$ if $(\lambda_j-k)^2(\lambda_j+k_{m2})^2 + (\lambda_j+k_{m2})^2 4k^2\alpha_j^2 > \lambda_j(\lambda_j+k)^2\left(\lambda_j+\alpha_{j,\mathrm{CE}}^2\right)$ and $0 < k < k_{m2}$.

Proof. The difference between $SMSE(\hat{\alpha}_{CE}^{LR})$ and $SMSE(\hat{\alpha}^{KL})$ is as follows:

$$MSE\left(\hat{\alpha}_{CE}^{LR}\right) - SMSE\left(\hat{\alpha}^{KL}\right) = \nabla = \sum_{j=1}^{p^*}\sum_{j=1}^{p^*} \frac{\left(\lambda_j+\alpha_{j,\mathrm{CE}}^2\right)}{(\lambda_j+k_{m2})^2} - \sum_{j=1}^{p^*} \frac{(\lambda_j-k)^2 + 4k^2\alpha_j^2}{\lambda_j(\lambda_j+k)^2} \tag{81}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\lambda_j(\lambda_j+k)^2\left(\lambda_j+\alpha_{j,\mathrm{CE}}^2\right) - (\lambda_j-k)^2(\lambda_j+k_{m2})^2 + (\lambda_j+k_{m2})^2 4k^2\alpha_j^2}{\lambda_j(\lambda_j+k_{m2})^2(\lambda_j+k)^2} \tag{82}$$

The difference $\nabla < 0$ occurs if
$(\lambda_j-k)^2(\lambda_j+k_{m2})^2 + (\lambda_j+k_{m2})^2 4k^2\alpha_j^2 > \lambda_j(\lambda_j+k)^2\left(\lambda_j+\alpha_{j,\mathrm{CE}}^2\right)$, and then it can be observed that the estimator $\hat{\alpha}_{CE}^{LR}$ is superior to the estimator $\hat{\alpha}^{KL}$ when $0 < k < k_{m2}$ for all $j = 1, 2, \ldots, p^*$. □

3.3.17. The $\hat{\alpha}_{BY}^{KL}$ Estimator and the $\hat{\alpha}^{KL}$ Estimator

The estimator $\hat{\alpha}_{BY}^{KL}$ is superior to the estimator $\hat{\alpha}^{KL}$ in the sense of the SMSE criterion, i.e., $SMSE(\hat{\alpha}_{BY}^{KL}) - SMSE(\hat{\alpha}^{KL}) < 0$ if $\lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+k_{m1})^2 > (\lambda_j-k_{m1})^2(\lambda_j+k)^2 + 4k_{m1}^2\alpha_{j,BY}^2(\lambda_j+k)^2$ and $0 < k < k_{m1}$.

Proof. The difference between $SMSE(\hat{\alpha}_{BY}^{KL})$ and $SMSE(\hat{\alpha}^{KL})$ is as follows:

$$SMSE\left(\hat{\alpha}_{BY}^{KL}\right) - SMSE\left(\hat{\alpha}^{KL}\right) = \nabla = \sum_{j=1}^{p^*} \frac{(\lambda_j-k_{m1})^2 + 4k_{m1}^2\alpha_{j,BY}^2}{\lambda_j(\lambda_j+k_{m1})^2} - \sum_{j=1}^{p^*} \frac{(\lambda_j-k)^2 + 4k^2\alpha_j^2}{\lambda_j(\lambda_j+k)^2} \tag{83}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{(\lambda_j-k_{m1})^2(\lambda_j+k)^2 + 4k_{m1}^2\alpha_{j,BY}^2(\lambda_j+k)^2 - (\lambda_j-k)^2(\lambda_j+k_{m1})^2 + 4k^2\alpha_j^2(\lambda_j+k_{m1})^2}{\lambda_j(\lambda_j+k_{m1})^2(\lambda_j+k)^2} \tag{84}$$

The difference $\nabla < 0$ ocurs if
$(\lambda_j-k)^2(\lambda_j+k_{m1})^2 + 4k^2\alpha_j^2(\lambda_j+k_{m1})^2 > (\lambda_j-k_{m1})^2(\lambda_j+k)^2 + 4k_{m1}^2\alpha_{j,BY}^2(\lambda_j+k)^2$, and then it can be observed that the estimator $\hat{\alpha}_{BY}^{KL}$ is superior to the estimator $\hat{\alpha}^{KL}$ when $0 < k < k_{m1}$ for all $j = 1, 2, \ldots, p^*$. □

3.3.18. The $\hat{\alpha}_{CE}^{KL}$ Estimator and the $\hat{\alpha}^{KL}$ Estimator

The estimator $\hat{\alpha}_{CE}^{KL}$ is superior to the estimator $\hat{\alpha}^{KL}$ in the sense of the SMSE criterion, i.e., $SMSE(\hat{\alpha}_{CE}^{KL}) - SMSE(\hat{\alpha}^{KL}) < 0$ if $\lambda_j\left(\lambda_j+\alpha_j^2\right)(\lambda_j+k_{m2})^2 > (\lambda_j-k_{m2})^2(\lambda_j+k)^2 + 4k_{m1}^2\alpha_{j,\mathrm{CE}}^2(\lambda_j+k)^2$ and $0 < k < k_{m2}$.

Proof. The difference between $SMSE(\hat{\alpha}_{CE}^{KL})$ and $SMSE(\hat{\alpha}^{KL})$ is as follows:

$$SMSE\left(\hat{\alpha}_{CE}^{KL}\right) - SMSE\left(\hat{\alpha}^{KL}\right) = \nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j - k_{m2}\right)^2 + 4k_{m2}^2\alpha_{j,\mathrm{CE}}^2}{\lambda_j\left(\lambda_j + k_{m2}\right)^2} - \sum_{j=1}^{p^*} \frac{\left(\lambda_j - k\right)^2 + 4k^2\alpha_j^2}{\lambda_j\left(\lambda_j + k\right)^2} \tag{85}$$

$$\nabla = \sum_{j=1}^{p^*} \frac{\left(\lambda_j - k_{m2}\right)^2\left(\lambda_j + k\right)^2 + 4k_{m2}^2\alpha_{j,\mathrm{CE}}^2\left(\lambda_j + k\right)^2 - \left(\lambda_j - k\right)^2\left(\lambda_j + k_{m2}\right)^2 + 4k^2\alpha_j^2\left(\lambda_j + k_{m2}\right)^2}{\lambda_j\left(\lambda_j + k_{m2}\right)^2\left(\lambda_j + k\right)^2} \tag{86}$$

The difference $\nabla < 0$ occurs if $\left(\lambda_j - k\right)^2\left(\lambda_j + k_{m2}\right)^2 + 4k^2\alpha_j^2\left(\lambda_j + k_{m2}\right)^2 > \left(\lambda_j - k_{m2}\right)^2\left(\lambda_j + k\right)^2 + 4k_{m2}^2\alpha_{j,\mathrm{CE}}^2\left(\lambda_j + k\right)^2$, and then it can be observed that the estimator $\hat{\alpha}_{CE}^{KL}$ is superior to the estimator $\hat{\alpha}^{KL}$ when $0 < k < k_{m2}$ for all $j = 1, 2, \ldots, p^*$. □

4. Monte Carlo Simulation

In this section, we conduct a comparative analysis of logistic regression estimators through a simulation study. Numerous researchers have undertaken simulation studies to evaluate the performance of estimators for linear and logistic regression models [32,33].

The MSE is a function of β and p and is minimized subject to constraint $\beta^T\beta = 1$ [34,35]. Schaeffer et al. [3] demonstrated that the logistic regression model can be constructed using a similar methodology to the linear regression model. The procedure outlined by Kibria et al. [36] can be employed to generate the correlated explanatory variables for this purpose:

$$x_{ij} = \left(1 - \rho^2\right)^{1/2}\omega_{ij} + \rho\omega_{i(j+1)},\ i = 1, 2, \ldots, n\ ,\ j = 1, 2, \ldots, p \tag{87}$$

In the given context, the variables ω_{ij} are assumed to be independent standard normal random variables, and ρ^2 represents the correlation between any two independent variables. We considered four levels of multicollinearity, ρ = 0.8, 0.9, 0.95, and 0.99. The response variable follows the Bernoulli distribution, i.e., $y_i \sim Be(\pi_i)$, where $\pi_i = \frac{exp(X_i^T\beta)}{1+exp(X_i^T\beta)}$. Sample size n is varied, i.e., 30, 50, 100, or 200. In this study, two cases of contamination were examined. In case 1, the model was contaminated with 10% outliers in the y-direction, while in case 2, the contamination level was increased to 20%. The simulation study was carried out using R-Studio (2024.12.0+467). We determined the number of outliers by multiplying the length of the response variable by 0.1 or 0.2, representing 10 percent or 20 percent of the total observations. In R, we utilize the round () function to obtain the nearest integer value for the number of outliers. To introduce outliers, we use the sample () function to randomly select indices from the response variable Y, which correspond to the observations where we will add outliers. Finally, we flip the values at the selected indices by subtracting them from 1. This operation effectively converts 0 s to 1 s and 1 s to 0 s [22]. The estimated MSE is calculated as

$$MSE\left(\hat{\beta}\right) = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{\beta}_i - \beta\right)^T\left(\hat{\beta}_i - \beta\right) \tag{88}$$

In the experiment, the vector $\hat{\beta}_i$ represents the estimated regression coefficients in the *ith* replication, while β denotes the vector of true parameter values. The true parameter values β are chosen such that $\beta^T\beta = 1$. The experiment was replicated 1000 times. The estimated mean squared errors (MSEs) are presented in Tables 1 and 2 for p = 3, with 10 and 20 percent outliers, respectively. Similarly, for p = 5, the results are provided in Tables 3 and 4, and for p = 7, the results are presented in Tables 5 and 6. Tables 1–6 show the estimated MSE for different estimators with different correlation coefficients (ρ) in the presence of outliers. Based on the provided tables (Tables 1–6), we can discuss the results as follows.

Table 1. Estimated MSE for different estimators when $p = 3$ with 10% outliers.

n	ρ	MLE	LL	LL-BY	LL-CE	LR	LR-BY	LR-CE	KL	KL-BY	KL-CE
30	0.8	2.3667	1.3344	2.0663	1.7486	1.1316	0.9932	1.0617	0.8906	0.7762	0.8679
	0.9	4.0135	1.9736	2.1172	1.9905	1.0952	1.0386	1.0363	1.0587	1.0223	1.0245
	0.95	7.7846	2.8569	4.4857	4.1533	2.0742	2.9920	2.0298	2.5645	2.1007	2.3126
	0.99	35.7174	9.6487	12.0985	12.6381	3.8876	3.8369	3.8497	3.2078	2.8987	2.6757
50	0.8	1.0746	0.9135	0.9128	0.9153	0.8449	0.8436	0.8451	0.7232	0.7219	0.7222
	0.9	1.4745	1.1337	1.1086	1.1303	0.9148	0.9120	0.9131	0.9103	0.8772	0.8789
	0.95	2.5356	1.6298	1.5467	1.6228	1.0073	1.0034	1.0048	0.9963	0.8892	0.8905
	0.99	9.4855	5.0952	4.7838	5.0980	2.9185	2.9139	2.9183	1.4127	1.4022	1.4115
100	0.8	0.7866	0.7495	0.7449	0.7489	0.7080	0.7126	0.7120	0.6569	0.6530	0.6565
	0.9	1.0340	0.9086	0.8991	0.9077	0.8187	0.8133	0.8176	0.8027	0.8013	0.8018
	0.95	1.7614	1.2665	1.2529	1.2575	1.0255	1.0201	1.0206	0.8678	0.8556	0.8567
	0.99	6.3481	2.6469	2.5970	2.6244	1.1340	1.1255	1.1278	0.9719	0.9654	0.9679
200	0.8	0.5321	0.5294	0.5316	0.5294	0.5290	0.5269	0.5269	0.5259	0.5242	0.5243
	0.9	0.6230	0.6062	0.6089	0.6063	0.5880	0.5859	0.5859	0.5625	0.5599	0.5601
	0.95	0.8556	0.7787	0.7815	0.7786	0.7168	0.7132	0.7134	0.5840	0.5811	0.5813
	0.99	2.3353	1.4464	1.4590	1.4382	0.8981	0.8861	0.8871	0.6113	0.6093	0.6101

Table 2. Estimated MSE for different estimators when $p = 3$ with 20% outliers.

n	ρ	MLE	LL	LL-BY	LL-CE	LR	LR-BY	LR-CE	KL	KL-BY	KL-CE
30	0.8	2.6073	1.3384	1.3781	1.3375	1.1018	1.0747	1.0762	0.8898	0.8700	0.8733
	0.9	4.6440	1.6009	1.6574	1.5864	1.1509	1.1348	1.1297	1.1465	1.1174	1.1206
	0.95	7.7352	2.0290	1.9976	1.9885	2.1242	2.1193	2.1172	2.1213	2.1107	2.1134
	0.99	36.1123	7.0375	6.2995	6.6513	3.9684	3.9544	3.9570	3.4126	3.2316	3.2343
50	0.8	1.1544	1.0301	1.0305	1.0299	0.9444	0.9431	0.9441	0.8174	0.8156	0.8162
	0.9	1.7432	1.3693	1.3717	1.3702	1.1207	1.1201	1.1205	0.9804	0.9730	0.9735
	0.95	2.6725	1.7791	1.8135	1.7916	1.1699	1.1598	1.1625	1.1263	1.1212	1.1215
	0.99	10.6493	5.5129	5.4835	5.5508	1.2289	1.2223	1.2225	1.2127	1.2082	1.2116
100	0.8	0.8673	0.8414	0.8396	0.8415	0.8175	0.8160	0.8176	0.7801	0.7791	0.7800
	0.9	1.0665	0.9790	0.9771	0.9789	0.9017	0.9003	0.9014	0.7919	0.7912	0.7918
	0.95	1.5526	1.2585	1.2555	1.2579	1.0396	1.0369	1.0384	1.0244	1.0125	1.0129
	0.99	5.2257	2.5697	2.5920	2.5578	1.1937	1.1919	1.1934	1.1855	1.1753	1.1760
200	0.8	0.6581	0.6540	0.6519	0.6540	0.6694	0.6513	0.6513	0.6464	0.6451	0.6461
	0.9	0.7321	0.7120	0.7097	0.7122	0.6957	0.6933	0.6953	0.6680	0.6668	0.6675
	0.95	0.9755	0.8935	0.8947	0.8941	0.8294	0.8280	0.8285	0.7211	0.7191	0.7207
	0.99	2.4798	1.5666	1.5336	1.5677	1.0288	1.0231	1.0301	1.0012	0.9878	0.9882

Table 3. Estimated MSE for different estimators when $p = 5$ with 10% outliers.

n	ρ	MLE	LL	LL-BY	LL-CE	LR	LR-BY	LR-CE	KL	KL-BY	KL-CE
30	0.8	4.7509	1.7719	1.8694	2.0446	1.3837	1.2094	1.4701	1.2945	1.1547	1.1841
	0.9	9.3889	2.7966	3.1155	2.7313	1.3917	1.2229	1.3199	1.3183	1.2213	1.2945
	0.95	20.1485	5.4596	8.2924	6.4925	3.5911	3.3180	3.3518	3.5437	3.0907	3.0924
	0.99	117.9439	48.1616	54.7636	24.7987	4.4837	4.9376	4.9512	4.3078	4.1847	4.1930
50	0.8	2.2180	1.4932	1.5367	1.4802	1.3211	1.2848	1.2748	0.8161	0.8069	0.8081
	0.9	4.0343	1.9980	2.0500	1.9610	1.5669	1.5198	1.5117	1.3023	1.2140	1.2231
	0.95	7.2837	2.6262	2.8973	2.6182	1.5750	1.5416	1.5325	1.3113	1.3043	1.3050
	0.99	36.4630	8.6211	7.8963	8.5427	2.2457	2.2124	2.2079	2.2132	2.2003	2.2045
100	0.8	1.0057	0.8932	0.9075	0.8981	0.8632	0.8523	0.8548	0.7544	0.7447	0.7453
	0.9	1.9873	1.4432	1.4599	1.4322	1.2313	1.2084	1.2013	1.2234	1.1379	1.1388
	0.95	3.4729	2.0386	2.0546	2.0149	1.4784	1.4307	1.4271	1.4687	1.4211	1.4234
	0.99	16.2475	6.3017	5.3991	6.0793	1.6421	1.5000	1.5969	1.5719	1.4543	1.4579
200	0.8	0.6969	0.6765	0.6778	0.6765	0.6702	0.6690	0.6690	0.6699	0.6439	0.6440
	0.9	0.9258	0.8331	0.8390	0.8340	0.8072	0.8025	0.8033	0.8244	0.7188	0.7192
	0.95	1.4154	1.1343	1.1511	1.1378	1.0429	1.0260	1.0277	1.0384	0.7518	0.7525
	0.99	5.1578	2.0883	2.1673	2.0945	1.3964	1.3695	1.3711	1.3745	1.2682	1.2706

Table 4. Estimated MSE for different estimators when $p = 5$ with 20% outliers.

n	ρ	MLE	LL	LL-BY	LL-CE	LR	LR-BY	LR-CE	KL	KL-BY	KL-CE
30	0.8	2.6521	1.2165	1.4282	1.2282	0.9910	0.9358	0.9677	0.9772	0.9222	0.9231
	0.9	5.0177	1.4555	2.1246	1.4569	1.3338	1.2300	1.2331	1.2565	1.2267	1.2286
	0.95	17.0628	3.8666	11.5752	5.0429	2.9537	2.4343	2.5087	2.9234	2.2355	2.2357
	0.99	111.8208	31.9524	40.0553	29.4738	4.1616	4.0202	4.0476	4.1026	3.3863	3.3897
50	0.8	1.9790	1.4599	1.5175	1.4510	1.3402	1.2733	1.2702	1.3395	1.2663	1.2673
	0.9	3.3806	1.8582	1.9332	1.8460	1.5521	1.4791	1.4750	1.5478	1.4708	1.4732
	0.95	5.9931	2.2945	2.4638	2.2794	1.6484	1.5489	1.5501	1.6278	1.5312	1.5321
	0.99	27.3854	5.1222	5.3875	5.0187	2.3050	2.2347	2.2354	2.2279	2.2097	2.2107
100	0.8	1.2265	1.0973	1.1089	1.0995	1.0462	1.0370	1.0378	1.0379	1.0287	1.0296
	0.9	1.8024	1.3999	1.4261	1.4005	1.2605	1.2474	1.2471	1.2519	1.2411	1.2418
	0.95	2.8681	1.8185	1.8560	1.8147	1.4418	1.4202	1.4170	1.4123	1.4118	1.4122
	0.99	11.3100	3.6191	3.4978	3.6768	1.8016	1.6766	1.6758	1.7864	1.6742	1.6756
200	0.8	0.8165	0.7935	0.7953	0.7935	0.7845	0.7729	0.7829	0.7776	0.7516	0.7549
	0.9	1.1018	1.0175	1.0195	1.0174	0.9781	0.9662	0.7961	0.9712	0.9568	0.9572
	0.95	1.6116	1.3422	1.3471	1.3419	1.2118	1.2083	1.2080	1.2093	1.2060	1.2078
	0.99	5.4312	2.4311	2.4556	2.4310	1.6087	1.5966	1.5967	1.5987	1.4576	1.4587

Table 5. Estimated MSE for different estimators when $p = 7$ with 10% outliers.

n	ρ	MLE	LL	LL-BY	LL-CE	LR	LR-BY	LR-CE	KL	KL-BY	KL-CE
30	0.8	3.5601	1.6900	1.9975	1.6859	1.6334	1.4737	1.4819	1.6232	1.4521	1.4552
	0.9	7.0527	2.2594	2.4219	2.2155	1.7429	1.5984	1.5994	1.7290	1.5317	1.5378
	0.95	13.4857	4.0066	4.6556	4.8187	4.7041	4.5423	4.5421	4.6683	4.4570	4.4598
	0.99	69.6288	11.8602	14.3567	12.0148	5.1370	5.0616	5.0646	5.3075	5.2836	5.2860
50	0.8	3.1278	1.6719	1.6435	1.6370	1.6184	1.4276	1.4274	1.5009	1.4226	1.4239
	0.9	7.006	2.1247	2.3751	2.1309	1.6678	1.5519	1.5521	1.6455	1.5214	1.5248
	0.95	13.1860	2.7181	3.8245	2.7436	1.9197	1.6478	1.6426	1.8762	1.6407	1.6417
	0.99	65.6741	10.2304	13.5697	9.7620	3.2418	3.1151	3.1131	3.2099	3.1120	3.1125
100	0.8	1.4038	1.1460	1.1438	1.1435	1.0812	1.0714	1.0790	1.0676	1.0543	1.0578
	0.9	2.3730	1.5736	1.5674	1.5710	1.3954	1.3800	1.3769	1.3723	1.3709	1.3711
	0.95	4.3467	2.1428	2.1554	2.1331	1.6501	1.6339	1.6286	1.5391	1.5222	1.5234
	0.99	4.7679	2.3280	2.2543	2.2571	1.7519	1.7435	1.7482	1.7479	1.7413	1.7456
200	0.8	0.8818	0.8488	0.8348	0.8283	0.8198	0.8139	0.8137	0.8122	0.8093	0.8100
	0.9	1.3480	1.1533	1.1413	1.1533	1.1255	1.1086	1.1087	1.1146	1.1119	1.1127
	0.95	2.2587	1.6193	1.6124	1.6198	1.5336	1.4932	1.4932	1.5327	1.4128	1.4187
	0.99	4.2735	1.9602	1.8145	1.8689	1.7301	1.7293	1.7279	1.7295	1.7268	1.7270

Table 6. Estimated MSE for different estimators when $p = 7$ with 20% outliers.

n	ρ	MLE	LL	LL-BY	LL-CE	LR	LR-BY	LR-CE	KL	KL-BY	KL-CE
30	0.8	4.1740	2.7526	2.8720	2.7515	2.6059	2.5138	2.5136	2.6034	2.5052	2.5121
	0.9	8.3509	3.0744	2.9618	2.8790	2.7169	2.6184	2.6191	2.7123	2.5826	2.5902
	0.95	19.3763	5.8239	5.7925	5.6500	5.6890	5.5807	5.5827	5.6436	5.5483	5.5498
	0.99	85.0216	17.9551	16.0241	15.3306	6.2436	6.1813	6.1838	6.2355	6.1726	6.1742
50	0.8	3.1585	1.9454	1.9100	1.7810	1.8651	1.6766	1.6774	1.7735	1.5673	1.5699
	0.9	8.2203	2.6601	2.5870	2.2841	1.9091	1.7605	1.7625	1.8595	1.7376	1.7424
	0.95	14.8272	3.9815	3.9791	3.8465	2.8527	2.7509	2.7538	2.5662	2.3470	2.5400
	0.99	78.4817	15.1810	15.4688	15.2021	5.3170	5.2484	5.2515	4.8867	4.3109	4.4150
100	0.8	1.7304	1.4146	1.4280	1.4131	1.3586	1.3408	1.3399	1.3468	1.3302	1.3389
	0.9	3.7697	1.8752	1.8666	1.8466	1.6917	1.6673	1.6653	1.6728	1.6609	1.6618
	0.95	4.9520	2.7641	2.6117	2.5436	2.0317	1.9787	1.9770	2.0253	1.8345	1.8423
	0.99	20.7055	6.7451	4.7749	4.7526	1.7976	1.7646	1.7630	1.7650	1.7612	1.7677
200	0.8	0.9386	0.8998	0.8981	0.8979	0.8860	0.8752	0.8854	0.8843	0.8736	0.8766
	0.9	2.2683	1.1610	1.1362	1.1310	1.1333	1.1303	1.1309	1.1326	1.1249	1.1286
	0.95	3.0432	1.6767	1.5841	1.5762	1.5661	1.5557	1.5560	1.5627	1.5212	1.5218
	0.99	7.6175	3.8367	3.0594	2.8387	2.0025	1.9661	1.9661	1.9959	1.8276	1.8317

As the sample size increases, the mean squared error (MSE) values generally decrease across all estimators. This trend indicates that the estimators provide more accurate and precise estimates with larger sample sizes. This behavior is illustrated in Figure 1. For a fixed sample size, higher correlation coefficients typically result in increased MSE values

across most estimators. This suggests that as the variables in the dataset become more correlated, the estimators face greater difficulty in capturing the underlying relationships, leading to larger estimation errors. This behavior is demonstrated in Figure 2. The MSE values increase as the number of predictors grows, indicating a rise in estimation errors with more predictors. This trend is depicted in Figure 3. To rank the estimators based on their performance, we consider the MSE values. Lower MSE values indicate better performance in terms of estimation accuracy. KL-BY appears to have the lowest MSE values, indicating the best performance. KL-CE, LR-BY, and LR-CE also demonstrate competitive performance, while LL and MLE show relatively higher MSE values, suggesting poor performance when compared to others.

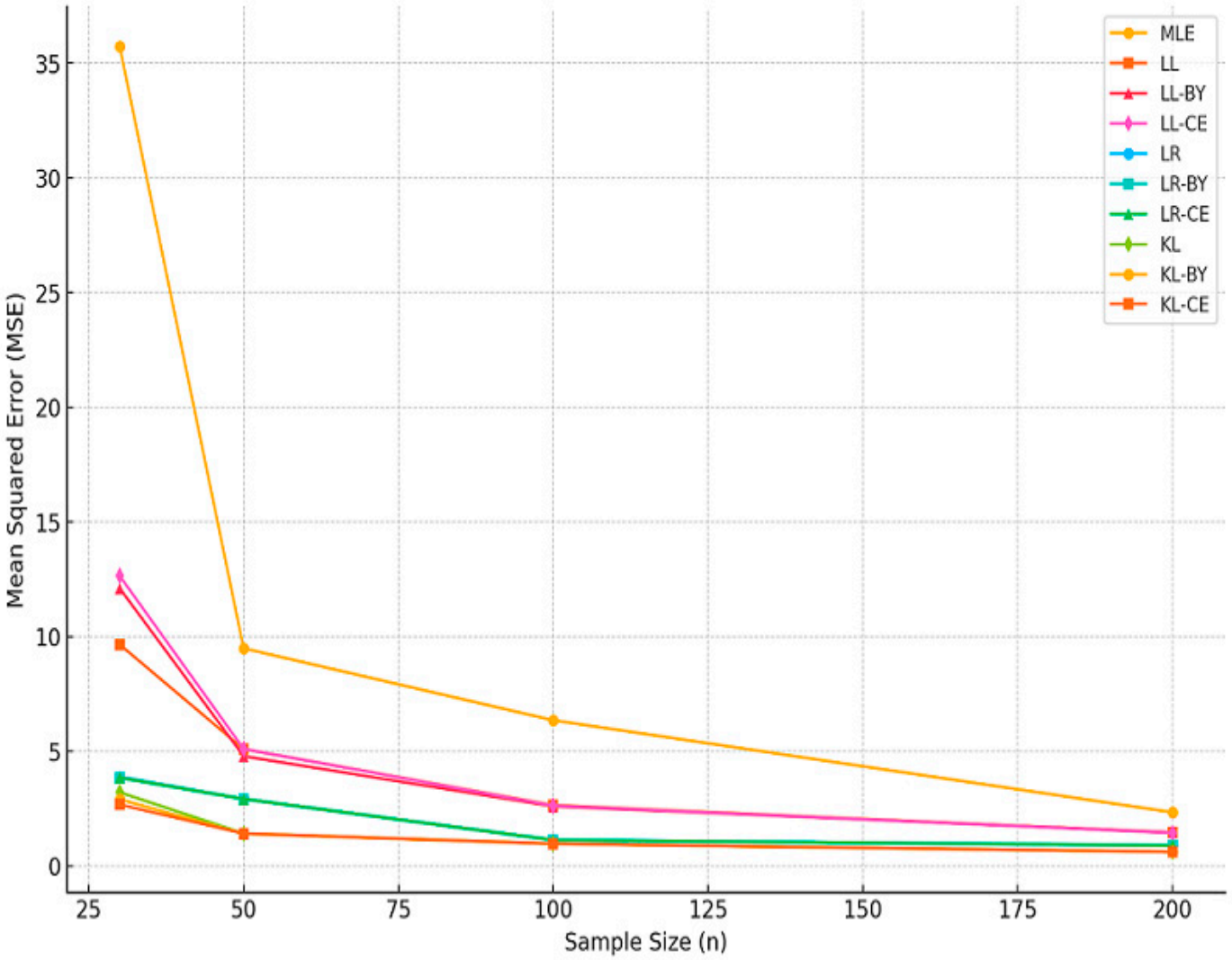

Figure 1. Graph of MSE against sample size n for $\rho = 0.99$ when $p = 3$ with 10% outliers.

Therefore, based on this ranking, KL-BY is the preferable estimator among the provided options, as it consistently exhibits the lowest MSE values and offers better estimation accuracy. However, it is important to note that the choice of estimator should also consider other factors, such as computational efficiency, robustness to outliers, and the specific requirements of the analysis or research question at hand.

It is worth noting that MLE, LL, KL, and LR are non-robust methods, whereas KL-BY, KL-CE, LL-BY, LL-CE, LR-BY, and LR-CE are robust methods. Therefore, it is expected that the robust estimators would generally outperform the non-robust estimators in scenarios with outliers or when the assumptions of the non-robust estimators are violated. Observing

the provided tables (Tables 1–6), it is apparent that the MSE values of the robust estimators (KL-BY, KL-CE, LL-BY, LL-CE, LR-BY, and LR-CE) are generally lower than those of the non-robust estimators (MLE, LL, KL, and LR) across different sample sizes (n) and correlation coefficients (ρ). This indicates that the robust estimators exhibit better performance in terms of estimation accuracy, particularly in the presence of outliers.

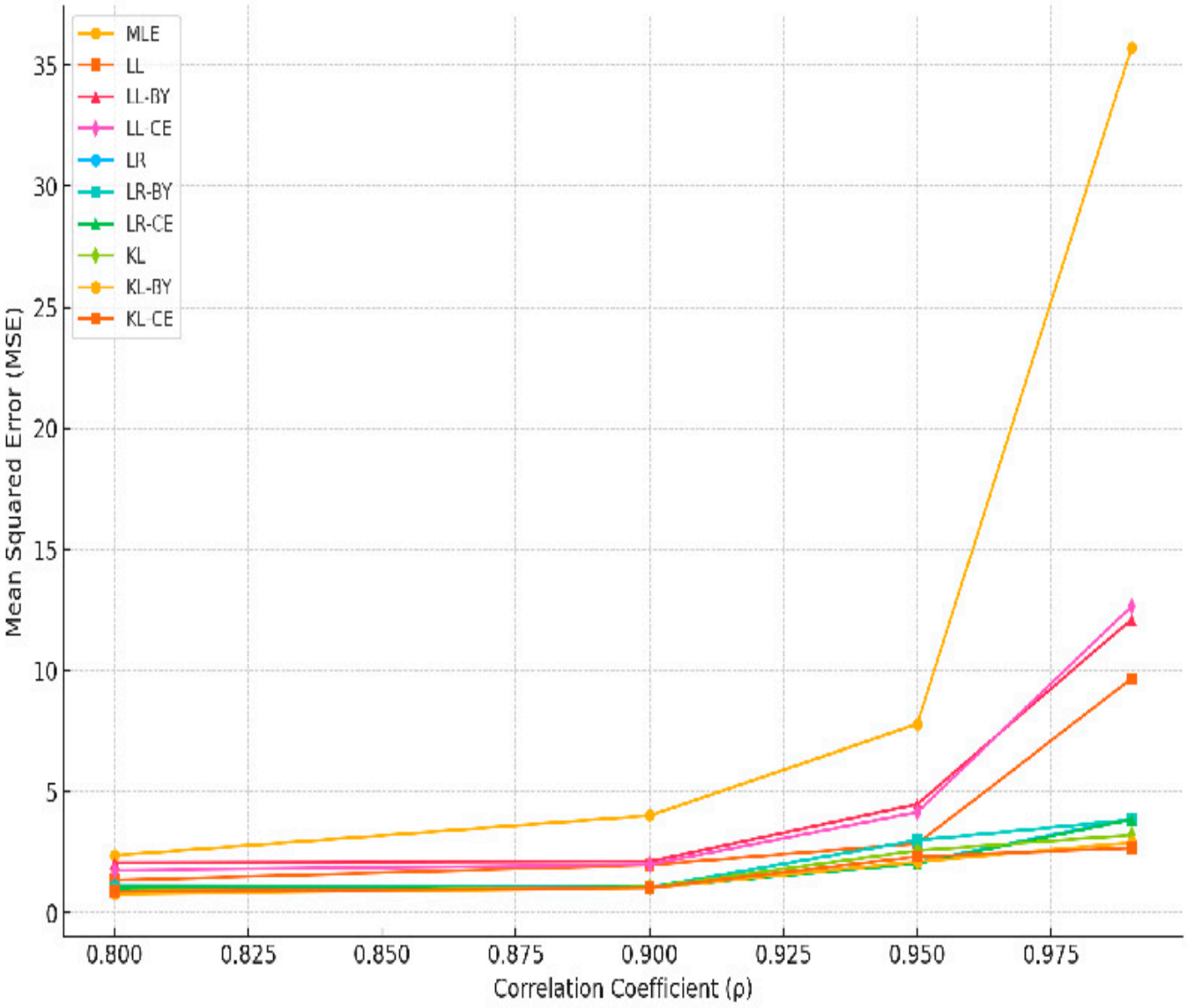

Figure 2. Graph of MSE against the level of multicollinearity for n = 30 when p = 3 with 10% outliers.

The superior performance of the robust estimators can be attributed to their ability to down-weight or ignore the influence of outliers, thereby producing more accurate estimates. On the other hand, the non-robust estimators are more sensitive to outliers, leading to higher MSE values when outliers are present. Based on this information, the revised ranking of the estimators would be as follows: KL-BY, KL-CE, LR-BY, LR-CE, LR, LL-BY, LL, LL-CE, and MLE (maximum likelihood estimator). Considering the robust nature of the estimators and their better performance in the presence of outliers, KL-BY remains the preferable estimator.

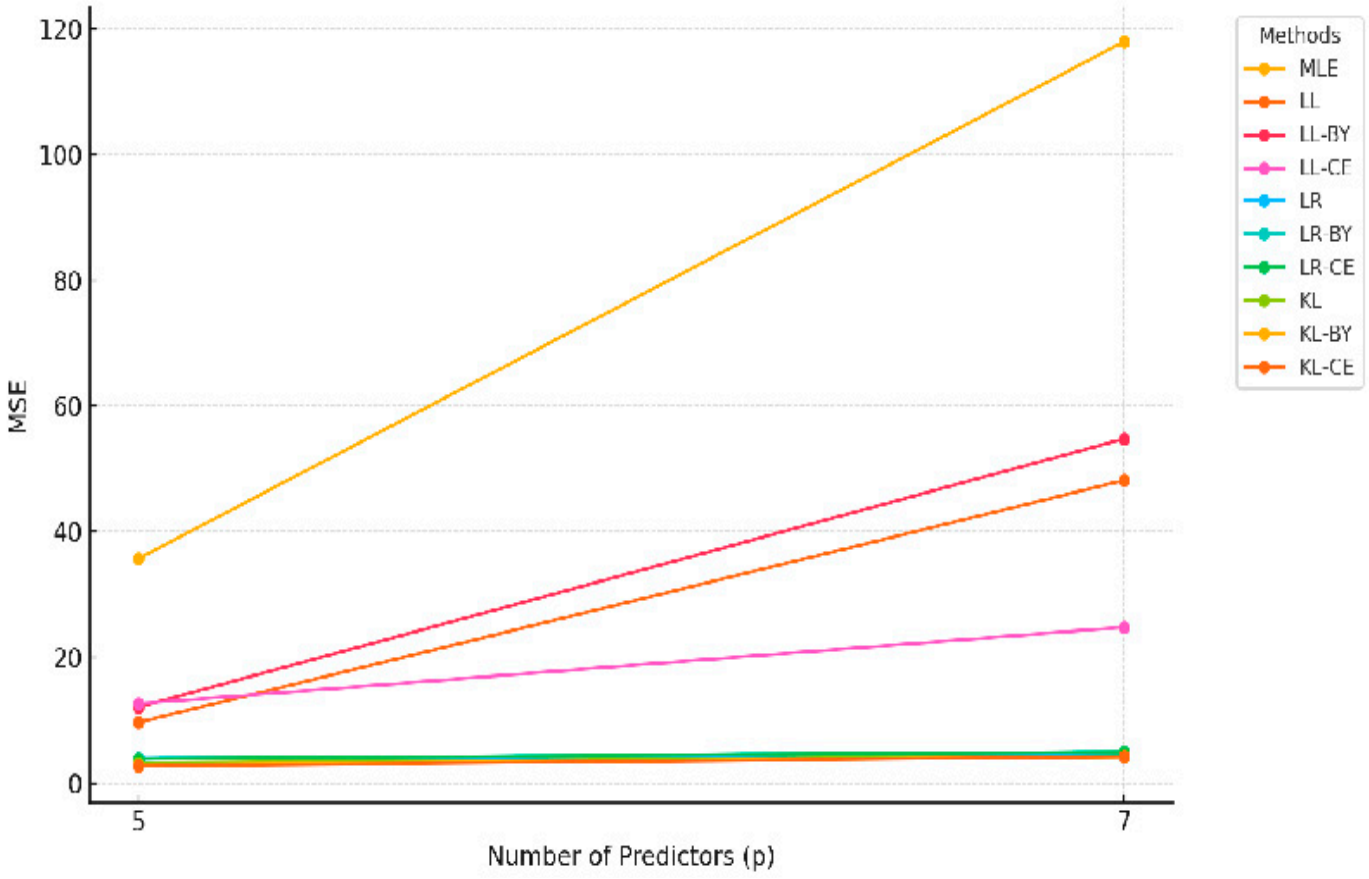

Figure 3. Graph of MSE against the number of predictors for n = 30 when $\rho = 0.99$ with 10% outliers.

5. Application

In this section, we analyze two real-life examples to examine the performance of the estimators under consideration.

5.1. Numerical Example 1

The dataset known as the skin data was initially introduced by Finney [37], later by Pregibon [24], and recently by Croux and Haesbroecks [28] to highlight the significance of influential observations in logistic regression. It focuses on binary outcomes, specifically the presence or absence of vasoconstriction of the skin of the digits after air inspiration. We investigated the relationship between the binary outcomes and two explanatory variables (the volume of air inspired (x_1) and the inspiration rate (x_2), both measured in logarithms).

The residuals plotted against the fitted values in Figure 4 identified cases 4, 18, and 24 as outliers. The outliers in this plot indicate that these cases exhibit significant residuals compared to the other observations, suggesting the presence of potential unusual patterns or influential points. Also, the Q–Q plot identified cases 4, 18, and 24 as outliers, which indicate deviations from the expected normal distribution of the residuals. These cases may exhibit extreme values or patterns that deviate significantly from the assumed model. The square root Pearson residuals against the predicted values plot also identify cases 4, 18, and 24 as outliers. Outliers in this plot suggest potential issues with model fit or influential points that affect the predicted values. Cases with significant square root Pearson residuals indicate that the model may not accurately capture their characteristics. The standardized Pearson residual and the leverage plot identified cases 4 and 18 as outliers and case 13 as leverage. Cook's distance shows that cases 4, 13, and 18 are influential points. Influential

points are observations that exert undue influence on the regression model's coefficients and may significantly impact the overall fit and conclusions drawn from the model.

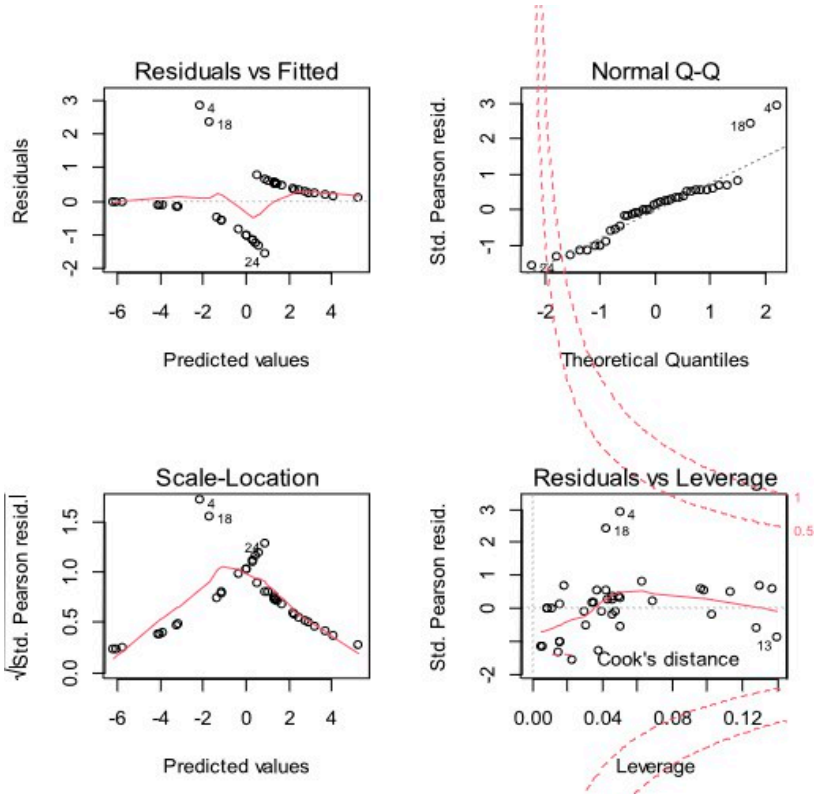

Figure 4. Diagnostic plots for the skin data.

In summary, cases 4, 18, and 24 consistently emerge as outliers across multiple diagnostic plots, indicating their display of unusual patterns or characteristics. Moreover, cases 4, 13, and 18 were identified as influential points, signifying their considerable impact on the regression model's coefficients and overall fit. Our findings align with Croux and Haesbroecks' [28] recognition of observations 4 and 18 as influential points, further reinforcing the importance of observations 4, 13, and 18 in shaping the results.

The correlation coefficient indicates a weak negative relationship between the two explanatory variables ($r = -0.38$). The variance inflation factor ($VIF = 2.838$) revealed the absence of multicollinearity in the model. Consequently, the model appears to be affected primarily by outliers. See Table 7 for the estimation results obtained using the selected methods.

Table 7. Estimated regression coefficients for the skin data.

Estimators	x_1	x_2	MSE
MLE	2.623	2.467	1.549
LL	2.414	2.268	1.416
LL-BY	2.240	2.006	1.394
LL-CE	2.212	2.069	1.398
LR	1.999	1.873	1.321
LR-BY	1.904	1.678	1.282
LR-CE	1.770	1.648	1.285
KL	1.375	1.278	1.315
KL-BY	1.262	1.071	1.217
KL-CE	1.120	1.029	1.287

Table 7 presents the coefficient estimates and mean squared error (MSE) for different estimators, including MLE, LL, LL-BY, LL-CE, LR, LR-BY, LR-CE, KL, KL-BY, and KL-CE. These results provide valuable insights into the performance of these estimators in terms of coefficient estimation and prediction accuracy. Comparing the coefficient estimates, we observe slight variations in the values of x_1 and x_2 and across the different estimators. While there are some differences, the general pattern is that the estimators provide relatively similar estimates for x_1 and x_2. Thus, this suggests a degree of consistency in the coefficient estimates among the different estimators.

Lower MSE values indicate better prediction accuracy. Among the considered estimators, KL-BY and KL-CE consistently demonstrate relatively lower MSE values, indicating better prediction performance than other estimators. However, it is worth noting that the differences in MSE values among the estimators are relatively small.

Generally, these results suggest that the estimators considered in the analysis provide reasonably accurate coefficient estimates and demonstrate comparable prediction accuracy. While slight variations exist in the coefficient estimates, the estimators generally perform well in connection with prediction accuracy, with KL-BY and KL-CE showing relatively better performance. These findings highlight the effectiveness of these estimators in capturing the relationship between the predictors and the dependent variable.

5.2. Numerical Example 2

Researchers have extensively studied the food stamp dataset in various research works. Stefanski et al. [38] provided an analysis of the dataset, acknowledging Rizek [39] as the first source of the study. Subsequently, Künsch et al. [31] and Carroll and Pederson [40] utilized the dataset to investigate the robustness of estimators for the logistic regression model. More recently, Kordzakhia et al. [41] employed the dataset. This dataset contains information from a survey conducted on over 2000 elderly citizens of the United States (U.S.), with a random sample of 150 individuals. The response variable "participation" indicates whether an individual participates in the U.S. Food Stamp Program, where "yes" is encoded as one and "no" as zero. In addition to the response variable, there are three predictor variables. "Tenancy" (x_1) denotes home ownership, with "yes" coded as one and "no" as zero. The variable "supplemental income" (x_2) represents whether an individual receives supplemental security income; a value of one indicates "yes", while zero means "no". The variable "income" (x_3) captures the individual monthly income in U.S. dollars.

Figure 5 illustrates the plot of residuals against the fitted values, revealing cases 66, 137, and 147 as outliers. These observations deviate significantly from the expected pattern, with their residuals differing considerably from the majority of the data points. This finding aligns with previous research by Kuinsch et al. [31], which also identified case 66 as an outlier. The Q–Q plot further confirms the presence of outliers, as cases 66, 137, and 147 exhibit departures from the normal distribution of residuals. These cases display unusual characteristics compared to the rest of the dataset. Additionally, the residual vs. leverage plot highlights cases 66, 137, and 147 as outliers and suggests potential issues with the model's fit or the presence of influential points. It is worth noting that case 66 stands out not only as an outlier but also as an influential point, as previously mentioned in the literature. Furthermore, based on Cook's distance, cases 66, 22, and 5 emerge as influential points. These observations exert undue influence on the regression model's coefficients, potentially impacting the overall fit, and the conclusions drawn from the analysis [40] also identified case 5 as a leverage point.

The correlation analysis reveals the presence of weak relationships among the variables under consideration. Specifically, the correlation coefficients indicate a low negative relationship between tenancy and supplemental income ($r = -0.19$) and income and supple-

mental income ($r = -0.18$). Conversely, they indicate a weak positive relationship between income and tenancy ($r = 0.28$). Furthermore, the variance inflation factor suggests the absence of multicollinearity within the model, indicating no multicollinearity. However, outliers influence the model. For detailed estimation outcomes obtained through the selected methods, please refer to Table 8.

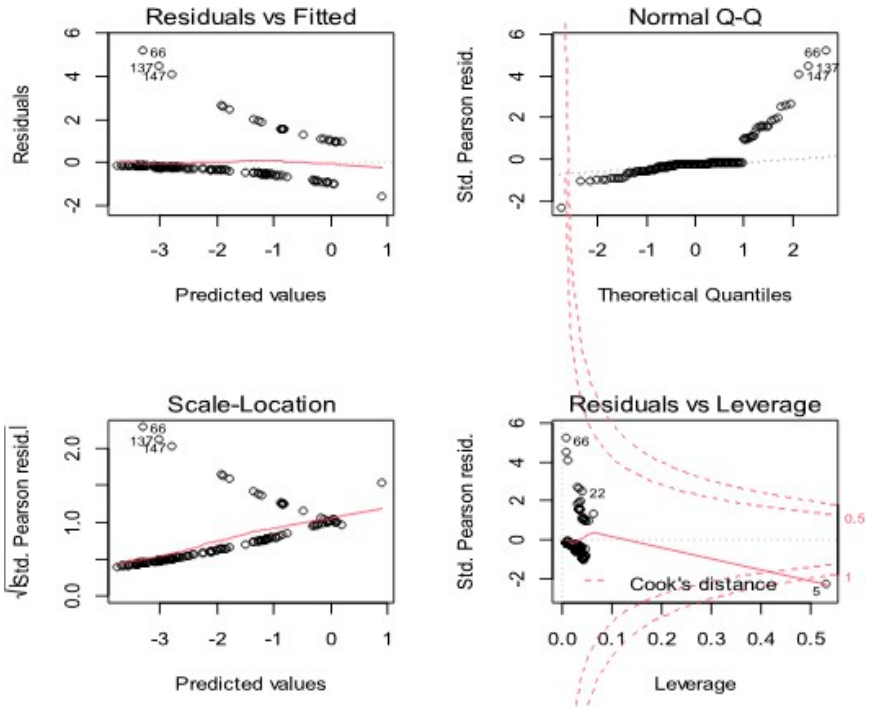

Figure 5. Diagnostic plots for the food stamp data.

Table 8. Estimated regression coefficients for the food stamp data.

Estimators	x_1	x_2	x_3	MSE
MLE	−1.879	0.953	0.182	0.530
LL	−1.517	0.806	−0.192	0.432
LL-BY	−1.467	0.773	−0.188	0.425
LL-CE	−1.528	0.792	−0.191	0.429
LR	−1.510	0.805	−0.193	0.375
LR-BY	−1.442	0.763	−0.189	0.366
LR-CE	−1.523	0.791	−0.191	0.374
KL	−1.142	0.656	−0.203	0.410
KL-BY	−1.065	0.613	−0.200	0.305
KL-CE	−1.153	0.647	−0.202	0.312

Table 8 displays the coefficient estimates and mean squared error (MSE) values for different estimators across variables, (x_1), (x_2), and (x_3). Analyzing the coefficient estimates, we observe slight variations in the magnitudes of the coefficients among the estimators. For instance, in terms of (x_1), the estimators MLE, LL, LL-BY, LL-CE, and LR-CE exhibit relatively similar coefficient estimates, while KL, KL-BY, and KL-CE have slightly different estimates. A similar pattern exists for variables (x_2) and (x_3), where the estimators generally produce similar coefficient estimates with some minor differences.

Focusing on the MSE values, we compared the performance of the estimators in terms of prediction accuracy. Among the estimators under study, KL-BY consistently demonstrates the lowest MSE values across all variables, indicating superior predictive performance. Next in performance is KL-CE, which exhibits relatively low MSE values.

MLE and LL have higher MSE values than other estimators, suggesting less accurate predictions. Generally, these results indicate that KL-BY is the most effective estimator regarding coefficient estimation and prediction accuracy. It consistently produces lower MSE values and exhibits stable coefficient estimates across the variables.

6. Conclusions

Logistic regression models encounter challenges when confronted with correlated predictors and influential outliers. The previous conclusion emphasized the potential of alternative methods, including the Liu and ridge estimators, in enhancing estimation accuracy for correlated predictors. However, these methods are susceptible to the presence of outliers, thereby compromising prediction stability. This study proposed the integration of robust estimators, specifically the Bianco–Yohai estimator (BY) and the conditionally unbiased bounded influence estimator (CE), with the shrinkage estimators. We evaluated the resulting estimators (LL-BY, LL-CE, LR-BY, LL-CE, KL-BY, and KL-CE) through simulations and real-life examples. The findings strongly favored the KL-BY, KL-CE, LR-BY, and LR-CE estimators, with KL-BY emerging as the preferred choice. Of the evaluated estimators, KL-BY consistently demonstrated superior performance in most scenarios, exhibiting a reduction in estimated mean squared error (MSE) values and displaying greater robustness against multicollinearity and outliers relative to other estimators. Our findings have practical implications for practitioners working with logistic regression models, as they frequently encounter challenges associated with multicollinearity and influential outliers. By adopting the KL-BY estimator proposed in this research, practitioners can achieve more stable and accurate predictions in their logistic regression analyses.

While this study addressed the impact of outliers in the response variable, further research is needed to explore the effects of outliers in the predictors. Future work could include extending the current framework to address outliers in predictor variables, which can significantly distort parameter estimates and model predictions. This could involve developing weighted or adaptive robust estimators that mitigate the influence of such outliers. Additionally, we examined binary logistic regression models with 3, 5, and 7 predictors, which reflect a variety of practical scenarios. However, we recognize the importance of higher-dimensional cases involving more predictors, as often encountered in some applied settings. Future research will focus on extending the methodology to address these higher-dimensional scenarios.

Although the current work is limited to binary logistic regression, the robust methods developed to address multicollinearity and regression outliers have the potential to be extended to more complex models, such as ordinal logistic regression and multinomial logistic regression. Exploring these extensions in future research could further enhance the utility of the proposed methods in handling a wider variety of regression frameworks, where outliers and multicollinearity remain significant challenges.

Author Contributions: Conceptualization: A.F.L. and S.M.; data analysis and interpretation: A.F.L., O.O. and R.A.F.; methodology: A.F.L., S.M., O.O. and R.A.F.; writing and review: A.F.L., S.M., O.O. and R.A.F. All authors have read and agreed to the published version of the manuscript.

Funding: The authors did not receive specific funding for this manuscript.

Data Availability Statement: Data is available on request from the authors. The code for reproducing the real-life application results presented in this manuscript is available on GitHub at https://github.com/olaluwoye9/Handling-multicollinearity-and-outliers-in-logistic-Regression-model (accessed on 18 December 2024).

Conflicts of Interest: There are no conflicts of interest to declare in this study.

References

1. Simpson, J.R.; Montgomery, D.C. A robust regression technique using compound estimation. *Nav. Res. Logist.* **1998**, *45*, 125–139. [CrossRef]
2. Lukman, A.F.; Ayinde, K.; Binuomote, S.; Clement, O.A. Modified ridge-type estimator to combat multicollinearity: Application to chemical data. *J. Chemom.* **2019**, *33*, e3125. [CrossRef]
3. Schaefer, R.L.; Roi, L.D.; Wolfe, R.A. A ridge logistic estimator. *Commun. Stat. Theory Methods* **1984**, *13*, 99–113. [CrossRef]
4. Aguilera, A.M.; Escabias, M.; Valderrama, M.J. Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Comput. Stat. Data Anal.* **2006**, *50*, 1905–1924. [CrossRef]
5. Mansson, G.; Golam Kibria, B.M.; Shukur, G. *On Liu Estimators for the Logit Regression Model*; The Royal Institute of Technology, Centre of Excellence for Science and Innovation Studies (CESIS): Stockholm, Sweden, 2012; p. 259.
6. Asar, Y.; Genc, A. New shrinkage parameters for the liu-type logistic estimators. *Commun. Stat.—Simul. Comput.* **2016**, *45*, 1094–1103. [CrossRef]
7. Jadhav, N.H. On linearized ridge logistic estimator in the presence of multicollinearity. *Comput. Stat.* **2020**, *35*, 667–687. [CrossRef]
8. Lukman, A.F.; Emmanuel, A.; Clement, O.A.; Ayinde, K. A modified ridge-type logistic estimator. *Iran. J. Sci. Technol. Trans. A Sci.* **2020**, *44*, 437–443. [CrossRef]
9. Lukman, A.F.; Kibria, B.M.G.; Nziku, C.K.; Amin, M.; Adewuyi, E.T.; Farghali, R. K-L estimator: Dealing with multicollinearity in the logistic regression model. *Mathematics* **2023**, *11*, 340. [CrossRef]
10. Barnett, V.; Lewis, T. *Outliers in Statistical Data*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 1994.
11. Rousseeuw, P.J.; Yohai, V. Robust regression by means of S estimators. In *Robust and Nonlinear Time Series Analysis*; Franke, J., Härdle, W., Martin, R.D., Eds.; Lecture Notes in Statistics; Springer: New York, NY, USA, 1984; Volume 26, pp. 256–274.
12. Yohai, V.J. High breakdown point and high efficiency robust estimates for regression. *Ann. Stat.* **1987**, *15*, 642–656. [CrossRef]
13. Rousseeuw, P.J.; van Driessen, K. Computing LTS regression for large data sets. *Data Min. Knowl. Discov.* **2006**, *12*, 29–45. [CrossRef]
14. Lukman, A.F.; Farghali, R.A.; Kibria, B.G.; Oluyemi, O.A. Robust-Stein estimator for overcoming outliers and multicollinearity. *Sci. Rep.* **2023**, *13*, 9066. [CrossRef] [PubMed]
15. Suhail, M.; Chand, S.; Kibria, B.G. Quantile-based robust ridge M-estimator for linear regression model in the presence of multicollinearity and outliers. *Commun. Stat.—Simul. Comput.* **2021**, *50*, 3194–3206. [CrossRef]
16. Lukman, A.F.; Ayinde, K.; Kibria, B.M.G.; Jegede, S.L. Two-parameter modified ridge-type M-estimator for linear regression model. *Sci. World J.* **2020**, *2020*, 3192852. [CrossRef] [PubMed]
17. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]
18. Hoerl, A.E.; Kennard, R.W.; Baldwin, K.F. Ridge regression: Some simulation. *Commun. Stat.—Theory Methods* **1975**, *4*, 105–123. [CrossRef]
19. Liu, K. A new class of biased estimate in linear regression. *Commun. Stat.* **1993**, *22*, 393–402.
20. Ozkale, M.R.; Kaciranlar, S. The restricted and unrestricted two-parameter estimators. *Commun. Stat.—Theory Methods* **2007**, *36*, 2707–2725. [CrossRef]
21. Kibria, B.M.; Lukman, A.F. A New Ridge-Type Estimator for the Linear Regression Model: Simulations and Applications. *Scientifica* **2020**, 9758378. [CrossRef]
22. Copas, J.B. Binary regression models for contaminated data. *J. R. Stat. Soc. Ser. B Methodol.* **1988**, *50*, 225–265. [CrossRef]
23. Victoria-Feser, M.P. Robust inference with binary data. *Psychometrika* **2002**, *67*, 21–32. [CrossRef]
24. Pregibon, D. Resistant fits for some commonly used logistic models with medical applications. *Biometrics* **1982**, *38*, 485–498. [CrossRef] [PubMed]
25. Croux, C.; Haesbroeck, G.; Rousseeuw, P.J. Location adjustment for the minimum volume ellipsoid estimator. *Stat. Comput.* **2002**, *12*, 191–200. [CrossRef]
26. Chen, C. Robust Regression and Outlier Detection with the ROBUSTREG Procedure. In Proceedings of the 27th Annual SAS Users Group International Conference, Williamsburg, VA, USA, 20–22 October 2002; SAS Institute Inc.: Cary, NC, USA, 2002.
27. Huber, P.J. Finite Sample Breakdown of M and P Estimators. *Ann. Stat.* **1984**, *12*, 119–126. [CrossRef]
28. Croux, C.; Haesbroeck, G. Implementing the Bianco and Yohai estimator for logistic regression. *Comput. Stat. Data Anal.* **2003**, *44*, 273–295. [CrossRef]
29. Habshah, M.; Syaiba, B.A. The Performance of Classical and Robust Logistic Regression Estimators in the Presence of Outliers. *Pertanika J. Sci. Technol.* **2012**, *20*, 313–325.
30. Bianco, A.M.; Yohai, V.J. Robust Estimation in the Logistic Regression Model. In *Robust Statistics, Data Analysis, and Computer Intensive Methods*; Rieder, H., Ed.; Springer: New York, NY, USA, 1996; Volume 109, pp. 17–34. [CrossRef]
31. Künsch, H.R.; Stefanski, L.A.; Carroll, R.J. Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *J. Am. Stat. Assoc.* **1989**, *84*, 460–466.

32. Varathan, N.; Wijekoon, P. Optimal generalized logistic estimator. *Commun. Stat.—Theory Methods* **2018**, *47*, 463–474. [CrossRef]
33. Farghali, R.A.; Lukman, A.F.; Ogunleye, A. Enhancing model predictions through the fusion of Stein estimator and principal component regression. *J. Stat. Comput. Simul.* **2024**, *94*, 1760–1775. [CrossRef]
34. Saleh, A.K.; Md, E.; Arashi, M.; Kibria, B.M.G. *Theory of Ridge Regression Estimation with Applications*; John Wiley: Hoboken, NJ, USA, 2019.
35. Newhouse, J.P.; Oman, S.D. *An Evaluation of Ridge Estimators*; Technical Report P-716-PR; Rand Corporation: Santa Monica, CA, USA, 1971.
36. Kibria, B.M.; Mansson, K.; Shukur, G. Performance of some logistic ridge regression estimators. *Comput. Econ.* **2012**, *40*, 401–414. [CrossRef]
37. Finney, D.J. *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve*; Cambridge University Press: Cambridge, MA, USA, 1947.
38. Stefanski, L.A.; Carroll, R.J.; Ruppert, D. Optimally bounded score functions for generalized linear models with applications to logistic regression. *Biometrika* **1986**, *73*, 413–425. [CrossRef]
39. Rizek, R. The 1977–78 Nationwide Food Consumption Survey. *Fam. Econ. Rev.* **1978**, *4*, 3–7.
40. Carroll, R.J.; Pederson, S. On robust estimation in the logistic regression model. *J. R. Stat. Soc. Ser. B Methodol.* **1993**, *55*, 693–706. [CrossRef]
41. Kordzakhia, N.; Mishra, G.D.; Reiersolmoen, L. Robust estimation in the logistic regression model. *J. Stat. Plan. Inference* **2001**, *98*, 211–223. [CrossRef]