Research Article

Some New Robust Estimators for Circular Logistic Regression Model with Applications on Meteorological and Ecological Data

Shokrya S. Alshqaq ¹, Abdullah A. Ahmadini ¹ and Ali H. Abuzaid ²

¹Department of Mathematics, Jazan University, Jazan, Saudi Arabia

²Department of Mathematics, Al-Azhar University-Gaza, Gaza, State of Palestine

Correspondence should be addressed to Shokrya S. Alshqaq; salshekak@jazanu.edu.sa

Received 2 April 2021; Accepted 6 May 2021; Published 26 May 2021

Academic Editor: Ishfaq Ahmad

Copyright © 2021 Shokrya S. Alshqaq et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Maximum likelihood estimation (MLE) is often used to estimate the parameters of the circular logistic regression model due to its efficiency under a parametric model. However, evidence has shown that the classical MLE extremely affects the parameter estimation in the presence of outliers. This article discusses the effect of outliers on circular logistic regression and extends four robust estimators, namely, Mallows, Schweppe, Bianco and Yohai estimator (BY), and weighted BY estimators, to the circular logistic regression model. These estimators have been successfully used in linear logistic regression models for the same purpose. The four proposed robust estimators are compared with the classical MLE through simulation studies. They demonstrate satisfactory finite sample performance in the presence of misclassified errors and leverage points. Meteorological and ecological datasets are analyzed for illustration.

1. Introduction

Circular data arise whenever the values of a random variable can present the circumference of the unit circle. It is measured by angles with values between 0 and 2π or 0° and 360° , for example, wind directions, animal navigation, or the values of any periodic phenomena such as a 24-hour clock or days of the year which can be converted to circular data [1]. The modeling of the relationship between circular variables is so-called “circular regression,” and it is classified into three main classes, namely, circular-circular, circular-linear, and linear-circular regression models [2].

The applications on circular regression models are widely spread in many applied fields. Several regression models were proposed to predict a continuous circular variable from other circular or linear predictors [3–6].

Logistic regression analysis is a useful statistical tool that analyses the relationship between a binary response and a predictor. The theory of logistic regression is well developed [7]. Daffaie and Khan [8] proposed the circular logistic model to predict a binary variable from a circular variable,

such as modeling the rainfall (yes, no) and wind direction, fatal road accident (yes, no), and time of accident.

The existence of outliers is a common problem in regression analysis. For the linear logistic model, Feser and Pia [9] showed that maximum likelihood estimation can be influenced by outliers. Croux et al. [10] found that the most dangerous outliers termed “bad leverage points” are misclassified observations that are outlying in the design space of predicted variables.

Circular logistic regression is also subjected to the existence of outliers as shown in [11], where they have proposed an outlier detection procedure based on the penalized maximum likelihood and applied it on different real datasets.

Several robust estimators that are less affected by outliers are proposed in the literature to improve the estimation performance in linear logistic regression models [12]. The authors in [13] introduced weights depending on the response and covariates and proposed the Mallows-type estimator. This estimator was analyzed deeply by Carroll and Pederson [14] using Mahalanobis distance to reduce the

weights in terms of leverage, and Bianco and Yohai [15] proposed BY methods.

Since the published work just considered the detection of outliers in the circular logistic regression model, this article attempts to overcome the problem of outliers in the circular logistic regression model by extending some robust estimators from the classical linear logistic to the circular logistic case.

The rest of this paper is organized as follows. Section 2 reviews the formulation of the circular logistic regression model and its parameter estimation via MLE. Section 3 presents the types of outliers in circular logistic regression and derives the proposed robust estimators in a logistic circular regression model. Section 4 discusses the effect of outliers on the circular logistic estimators by computing the influence functions. Section 5 investigates the performance of the considered robust estimators. Section 6 applies the proposed estimators on meteorological and ecological datasets. Section 7 provides the conclusion.

2. Model Formulation

A circular logistic regression describes the relationship between a binary response and circular predictors. It shows potential with various applications in the field of environmental sciences. The authors in [8] assumed that n binomial observations with a probability of success, $p_i = (v_i/n_i)$, for $i = 1, \dots, k$, depend on circular random variable u_i , and the proposed model is given as follows:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta^* \cos(u_i - u_0), \quad (1)$$

where β_0 is the value of the logit (log odds) when $(u_i - u_0) = 90^\circ$ and u_0 is the angle where the logit reaches its highest value. Let $\beta^* = \sqrt{\beta_1^2 + \beta_2^2}$ and $u_0 = \arctan(\beta_2/\beta_1)$, and equation (1) can be written as

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 \cos u_i + \beta_2 \sin u_i. \quad (2)$$

Suppose that binomial data of the form v_i successes out of n_i trials are observations from a binomial distribution, the likelihood function is then given by

$$L(\beta) = \prod_{i=1}^k \binom{n_i}{v_i} p_i^{v_i} (1-p_i)^{n_i-v_i}. \quad (3)$$

Let $\eta_i = \beta_0 + \beta_1 \cos u_i + \beta_2 \sin u_i$, and by using the exponential function, we obtain

$$\frac{p_i}{1-p_i} = e^{\eta_i} \Rightarrow p_i = \frac{e^{\eta_i}}{(1+e^{\eta_i})}. \quad (4)$$

The maximum likelihood estimation (MLE) is classically used for parameter estimation and is defined by an objective function as

$$\hat{\beta}_{\text{MLE}} = \text{argmax}_{\beta} \sum_{i=1}^k \log L(v_i, u_i, \beta), \quad (5)$$

where

$$\log L(v_i, u_i, \beta) = \left\{ \log \binom{n_i}{v_i} + v_i \log p_i + (n_i - v_i) \log (1 - p_i) \right\}. \quad (6)$$

The maximum likelihood equations are given as follows:

$$\frac{\partial \log L(\beta)}{\partial \beta_0} = \sum_{i=1}^k v_i - \sum_{i=1}^k \frac{n_i}{1+e^{-\eta_i}}, \quad (7)$$

$$\frac{\partial \log L(\beta)}{\partial \beta_1} = \sum_{i=1}^k v_i \cos u_i - \sum_{i=1}^k \frac{n_i \cos u_i}{1+e^{-\eta_i}}, \quad (8)$$

$$\frac{\partial \log L(\beta)}{\partial \beta_2} = \sum_{i=1}^k v_i \sin u_i - \sum_{i=1}^k \frac{n_i \sin u_i}{1+e^{-\eta_i}}. \quad (9)$$

These equations are solved iteratively by using the Newton–Raphson method. Recently, Abuzaid and ElShekh Ahmed [11] used the penalized maximum likelihood estimator (PMLE) to identify outliers in the circular logistic regression model and investigated its performance via simulation. The following section discusses some possible robust estimators for the circular logistic regression model.

3. Robust Estimators for Circular Logistic Regression

3.1. Outliers in Circular Logistic Regression. This section distinguishes different cases of outlying observations in circular logistic regression, where outliers might occur in the dependent variable, independent variable, or both of them.

For binary data, all the v_i 's values are either 0 or 1. Hence, an error in the v_i direction can only occur as a transposition of 0 to 1 or vice versa. This type of outlier is known as residual outlier or misclassification-type error [16,17].

A leverage outlier or leverage point occurs when the circular observation at position d (e.g., u_d) is contaminated as follows: $u_d^* = u_d + \gamma \pi \text{mod}(2\pi)$, where u_d^* is the value after contamination and γ is the degree of contamination in the range $0 \leq \gamma \leq 1$. A leverage point can be considered as a good leverage point when $v_i = 1$ with a large value of $P(v_i = 1|u_i)$ while it is a bad leverage point when $v_i = 0$ with small value of $P(v_i = 1|u_i)$ and vice versa. Abuzaid and ElShekh Ahmed [11] considered the misclassification-type error outlier in their simulation study without any investigation of the leverage point detection.

Alternatively, robust estimators are the common methods used for handling the problem of outliers in logistic models. The following section presents four robust estimators for the parameters in the circular logistic model as follows.

3.2. Circular Mallows Class. The proposed estimator (MALLOWS_C) is extended from the Mallows class in [18] to the circular logistic model for weighting the maximum likelihood estimator.

Assume F is a continuous and increasing distribution function and is given by

$$F(\eta_i) = \frac{e^{-\eta_i}}{(1 + e^{-\eta_i})}. \quad (10)$$

Then, the partial derivatives in (7)–(9) become

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta_0} &= \sum_{i=1}^k [v_i - F(\eta_i)], \\ \frac{\partial l(\beta)}{\partial \beta_1} &= \sum_{i=1}^k \cos u_i [v_i - F(\eta_i)], \\ \frac{\partial l(\beta)}{\partial \beta_2} &= \sum_{i=1}^k \sin u_i [v_i - F(\eta_i)], \end{aligned} \quad (11)$$

respectively. The robust estimates for the circular logistic regression model in equation (10) are given by the solution of β obtained by

$$\frac{\partial l(\beta)}{\partial \beta_0} = \sum_{i=1}^k w_i \{v_i - F(\eta_i) - c(u_i, \beta_0)\} = 0, \quad (12)$$

$$\frac{\partial l(\beta)}{\partial \beta_1} = \sum_{i=1}^k w_i \cos u_i \{v_i - F(\eta_i) - c(u_i, \beta_1)\} = 0, \quad (13)$$

$$\frac{\partial l(\beta)}{\partial \beta_2} = \sum_{i=1}^k w_i \sin u_i \{v_i - F(\eta_i) + c(u_i, \beta_2)\} = 0, \quad (14)$$

where w_i are the weights that may depend on u_i , v_i , or both and $c(\cdot)$ is a correction function needed to ensure consistency. If $w_i = 1$ and $c(\cdot) = 0$, then equations (12)–(14) give the usual circular logistic regression estimate. If $w_i = w(u_i, \eta_i)$ and $c(\cdot) = 0$, then the weights depend only on u_i .

3.3. Circular Schweppe Class. Stefanski [19] stated that w_i is robust Mahalanobis distance for the u_i vector that depends on the covariance matrix of the regression model, which is given by

$$\hat{\Sigma} = (u_i' W u_i)^{-1}, \quad (15)$$

where W is a diagonal matrix with $w_{ii} = p_i(1 - p_i)$ and $(1 - p_i)$ is the probability p_i of $v_i = 1$; then, the Mahalanobis distance is given by

$$D_M = \sqrt{(u_i - \bar{u}_i)' \hat{\Sigma}^{-1} (u_i - \bar{u}_i)}. \quad (16)$$

If $w_i = w(u_i, \eta_i, v_i)$, then the estimator is the same as the linear logistic Schweppe class proposed in [13]. Here, w_i depends on v_i and circular u_i . This estimator is called circular conditionally unbiased bounded influence function or CUBIF_C estimator.

3.4. Circular BY Estimators. Bianco and Yohai [15] proposed BY methods for the linear logistic model, and in this section, we extend it to the circular case and referred to as BY_C.

Let the MLE be obtained by minimizing the deviance,

$$\hat{\beta}_{MLE} = \operatorname{argmin}_{\beta} \sum_{i=1}^k D_i(\beta), \quad (17)$$

where $D_i(\beta) = \{-v_i \ln[F(\eta_i)] - (1 - v_i) \ln[1 - F(\eta_i)]\}$. By replacing the deviance function in (17) with function ρ , the robust BY_C estimator is defined by

$$\hat{\beta}_{BY_C} = \operatorname{argmin}_{\beta} \sum_{i=1}^k \rho[D(F(\eta_i), v_i) + G(F(\eta_i)) + G(1 - F(\eta_i))], \quad (18)$$

where ρ is a bounded, differentiable, and nondecreasing function defined in [15] and given by

$$\rho(u) = \begin{cases} u - \left(\frac{u^2}{2m}\right), & \text{if } u \leq m, \\ \frac{m}{2}, & \text{otherwise,} \end{cases} \quad (19)$$

where m is a positive number, $\rho'(u) = \psi(u)$, and $G(u) = \int_0^u \psi(-\ln g) dg$.

3.5. Circular WBY Estimators. The extension of BY estimator by including weights is reducing the influence of outliers in u_i space. This weighted BY (WBY) estimator is extended to the circular logistic regression model and defined as

$$\hat{\beta}_{WBY_C} = \operatorname{argmin}_{\beta} \sum_{i=1}^k w(u_i) \rho[D(F(\eta_i), v_i) + G(F(\eta_i)) + G(1 - F(\eta_i))], \quad (20)$$

where the weights $w(u_i)$ are distances which are computed using the minimum covariance determinant (MCD) estimator, to be a decreasing function of robust Mahalanobis distances (RD_i), and given by (see [20])

$$w(u_i) = \begin{cases} 1, & \text{if } RD_i^2 \leq \chi_{p,0.975}^2, \\ 0, & \text{elsewhere.} \end{cases} \quad (21)$$

4. Influence Function

Suppose two source populations with k -dimensional circular variables, which are both von Mises distribution with different means but the same concentration parameter, κ . Circular variable u_i can arise from one of these populations

$$u_i \sim \begin{cases} H_1 = v_i M(\mu_1, \kappa), & \text{with probability } p_1, \\ H_0 = v_i M(\mu_0, \kappa), & \text{with probability } p_0, \end{cases} \quad (22)$$

where $i = 1, \dots, k$ and $p_1 + p_0 = 1$. Let the binary variable v_i indicates the source population of the corresponding u_i , then

$$v_i \sim \begin{cases} 1, & \text{with probability } p_1, \\ 0, & \text{with probability } p_0. \end{cases} \quad (23)$$

Let the joint distribution of (u_i, v_i) be denoted by H and T be an estimator of the circular logistic parameters. Then, the influence function [21] is defined as

$$\begin{aligned} \text{IF}((u_i, v_i), T, H) &= \lim_{\varepsilon \rightarrow 0} \frac{T((1 - \varepsilon)H + \varepsilon \Delta_{(u_i, v_i)}) - T(H)}{\varepsilon} \\ &= \frac{\partial}{\partial \varepsilon} T((1 - \varepsilon)H + \varepsilon \Delta_{(u_i, v_i)})|_{\varepsilon=0}. \end{aligned} \quad (24)$$

If $T = \text{MLE}$, as shown in equations (7)–(9), then the influence function is an unbounded function for spaces u_i and v_i . Specifically, a small amount of contamination in the training data due to the presence of possible outlier in u_i or v_i intensively affects the MLE, as shown in the simulation section.

Suppose $T = \text{MALLOWS}_C$, where the weights depend on the robust distance of observation u_i , and robust distance RD_i is equal to the Mahalanobis distance of u_i to the center of the data cloud. This condition reduces the influence of outlying observations in the u_i space. Thus, $\text{IF}((u_i, v_i), \text{MALLOWS}_C, H)$ is bounded with respect to u_i but unbounded with respect to v_i . Similar conclusions as $T = \text{BY}_C$ can be derived.

If $T = \text{CUBIF}_C$ or $T = \text{WBY}_C$, which adds a weight to BY_C , then a fully bounded influence is obtained.

5. Simulation Study

5.1. Settings. This simulation aims to compare the robustness of the proposed robust estimators and the classical MLE. The independent circular variable is generated from von Mises distribution with mean 60° and concentration parameter of $\kappa = 1, 2, 6, 10$, and 15 , with sample size

$N = 200, 250$, and 300 ; a large sample size is chosen to avoid separation problems. The true parameter values are $\beta = (0, 2, 2)$.

The simulation study is reported in a variety of situations. Initially, the data without contamination are simulated. The robust properties of all estimators with contaminated data are examined in three different ways. First, proportions are taken from the responses, and v is chosen randomly and changed from either 0 to 1 or 1 to 0. This process constitutes the misclassification-type error. For each contaminated case, 5%, 10%, 20%, 30%, and 40% of the original data are contaminated. Second, the same proportions are taken to contaminate u with $\gamma = 0.5$ for good leverage points. Finally, the same proportions are considered, and the generated data are contaminated with two types of outliers simultaneously. This process constitutes bad leverage points.

Each simulation includes 1000 replications. The performance of the estimators is evaluated based on the bias and the median squared error (MSE) for each parameter, which are defined as follows:

$$\begin{aligned} \text{bias} &= \frac{1}{1000} \sum_{i=1}^{1000} \hat{\beta}_i - \beta_i, \\ \text{MSE} &= \text{median} \left(\sqrt{\sum_{i=1}^{1000} (\hat{\beta}_i - \bar{\beta}_i)^2} \right). \end{aligned} \quad (25)$$

A good estimator has bias and MSE that are relatively small or close to zero.

The simulation used the standard available “Robust” package in R to obtain the estimators and the “CircStats” package for generating the circular variable u .

5.2. Results. The bias and MSE of the five estimators are shown in Tables 1 to 7. Table 1 shows the results for uncontaminated data (i.e., clean data), where the biases and MSEs of all five estimators are fairly close to each other. However, the MALLOWS_C 's MSE is larger than the others for large concentration parameter ($\kappa = 6, 10$, and 15). Hence, MALLOWS_C performs worse compared with the other estimators in this situation.

Tables 2 and 3 show the results of data with misclassified errors. The bias and MSE of the MLE estimates are immediately affected by 5% misclassified type error. The results suggest that the MLE becomes biased with 5% contamination, CUBIF_C with 20% contamination, and BY_C with 30% contamination. MALLOWS_C and WBY_C are good robust estimators and considered as the best methods.

As shown in Tables 4 and 5, small differences are found between the classical MLE and the robust methods when contaminated data with good leverage points are used, where the biases and MSEs for all methods are relatively small. Hence, good leverage points do not affect the data.

The results in the case of bad leverage points change the performance of estimators intensively. As shown in Tables 6 and 7, the CUBIF_C estimator can only withstand up to 5%

TABLE 1: Bias and MSE of all methods for uncontaminated data.

κ	Methods	$N = 200$		$N = 250$		$N = 300$	
		Bias	MSE	Bias	MSE	Bias	MSE
1	MLE	0.0049	0.0166	0.0023	0.0127	0.0025	0.0114
	MALLOWS _C	0.0029	0.0194	0.0017	0.0140	0.0028	0.0123
	CBIF _C	0.0038	0.0166	0.0021	0.0128	0.0022	0.0114
	BY _C	0.0051	0.0179	0.0028	0.0135	0.0028	0.0117
	WBY _C	0.0050	0.0179	0.0030	0.0135	0.0030	0.0118
2	MLE	0.0064	0.0213	0.0012	0.0166	1.2e-03	0.0131
	MALLOWS _C	0.0113	0.1496	0.0055	0.0601	2.3e-03	0.0406
	CBIF _C	0.0062	0.0233	0.0010	0.0177	1.6e-03	0.0136
	BY _C	0.0083	0.0220	0.0024	0.0171	1.0e-03	0.0136
	WBY _C	0.0057	0.1396	0.0010	0.0182	4.2e-05	0.0143
6	MLE	0.0170	0.2221	0.0110	0.0849	0.0066	0.0642
	MALLOWS _C	0.1099	0.8734	0.0767	0.3499	0.0417	0.2006
	CBIF _C	0.0197	0.2346	0.0176	0.0963	0.0105	0.0749
	BY _C	0.0060	0.2248	0.0043	0.0908	0.0033	0.0641
	WBY _C	0.0307	0.2402	0.0007	0.0316	0.0471	0.0632
10	MLE	0.0216	0.1392	0.0162	0.1155	0.0168	0.0938
	MALLOWS _C	0.0634	0.3600	0.0624	0.2769	0.0609	0.2530
	CBIF _C	0.0341	0.1529	0.0319	0.1311	0.0160	0.1030
	BY _C	0.0326	0.1419	0.0173	0.1182	0.0177	0.0948
	WBY _C	0.0764	0.1264	0.0719	0.1090	0.0752	0.0871
15	MLE	0.0895	0.2501	0.0706	0.1992	0.0330	0.1789
	MALLOWS _C	0.1518	1.0260	0.1158	0.4844	0.1414	0.4132
	CBIF _C	0.0784	0.2749	0.0716	0.2137	0.0529	0.1945
	BY _C	0.0859	0.2560	0.0626	0.2035	0.0324	0.1805
	WBY _C	0.0934	0.9754	0.0835	0.1949	0.0816	0.1661

TABLE 2: Bias and MSE of all methods for data with various percentages of misclassified errors (5%, 10%, and 20%).

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	5	MLE	1.4570	2.1918	1.5570	2.2018	1.6581	2.3018
		MALLOWS _C	0.1400	0.3073	0.2510	0.4162	0.1367	0.3612
		CBIF _C	0.5148	0.4482	0.4648	0.3472	0.4636	0.3371
		BY _C	0.1344	0.3765	0.1233	0.2756	0.1144	0.2565
		WBY _C	0.1745	0.3842	0.0865	0.3945	0.0841	0.2951
2		MLE	1.5570	2.2918	1.6570	2.3018	1.7581	2.4018
		MALLOWS _C	0.2400	0.4073	0.3510	0.5162	0.2367	0.4612
		CBIF _C	0.6148	0.5482	0.5648	0.4472	0.5636	0.4371
		BY _C	0.2344	0.4765	0.2233	0.3756	0.2144	0.3565
		WBY _C	0.2745	0.4842	0.1865	0.4945	0.1841	0.3951
6		MLE	1.0570	2.1818	1.0570	2.2018	1.0581	2.3018
		MALLOWS _C	0.1400	0.3173	0.2510	0.6262	0.1267	0.5512
		CBIF _C	0.5048	0.4282	0.4748	0.3372	0.4536	0.2271
		BY _C	0.0244	0.3265	0.0133	0.2656	0.0144	0.2465
		WBY _C	0.0645	0.3442	0.0765	0.3845	0.0741	0.2851
10		MLE	1.5580	2.2018	1.6680	2.3128	1.7691	2.4128
		MALLOWS _C	0.0310	0.2183	0.1400	0.3052	0.0257	0.2502
		CBIF _C	0.4038	0.3372	0.3538	0.2362	0.3526	0.2261
		BY _C	0.0234	0.2655	0.0123	0.1646	0.0034	0.1455
		WBY _C	0.0645	0.2732	0.0755	0.2845	0.0731	0.1841
15		MLE	1.5681	2.2029	1.6681	2.3129	1.7692	2.4129
		MALLOWS _C	0.1311	0.3162	0.2401	0.4051	0.1256	0.3501
		CBIF _C	0.5037	0.4371	0.4537	0.3361	0.4525	0.3260
		BY _C	0.1233	0.3654	0.1122	0.2645	0.1033	0.2454
		WBY _C	0.1634	0.3731	0.0754	0.3834	0.0730	0.2840

TABLE 2: Continued.

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	10	MLE	2.4648	6.1013	2.3537	6.0002	2.2526	6.1113
		MALLOWS _C	0.3484	0.2098	0.2393	0.1087	0.1292	0.0076
		CBIF _C	0.8933	1.4598	0.9922	0.9497	0.9811	0.9386
		BY _C	0.2542	0.1716	0.1431	0.0605	0.0320	0.0504
		WBY _C	0.0565	0.1120	0.0454	0.0110	0.0343	0.0121
2		MLE	2.3648	6.0013	2.2537	6.0102	2.1526	6.0113
		(MALLOWS _C)	0.2484	0.1098	0.1393	0.0087	0.0292	0.0066
		(CBIF _C)	0.6933	0.7598	0.8922	0.8497	0.8811	0.8386
		BY _C	0.1542	0.0716	0.0431	0.0505	0.0220	0.0404
		WBY _C	0.0465	0.0120	0.0354	0.0010	0.0243	0.0021
6		MLE	2.3548	6.0013	2.2437	6.0112	2.1426	6.0013
		(MALLOWS _C)	0.2384	0.1198	0.1293	0.0187	0.0192	0.0176
		(CBIF _C)	0.9833	.6498	0.8822	0.8397	0.8711	0.8286
		BY _C	0.1442	0.0616	0.0331	0.1505	0.1220	0.1404
		WBY _C	0.1465	0.0020	0.1354	0.1010	0.1243	0.1021
10		MLE	2.5758	6.2123	2.4647	6.1112	2.1416	6.2223
		MALLOWS _C	0.2374	0.1188	0.1283	0.0077	0.0182	0.0166
		CBIF _C	0.8823	0.5488	0.8812	0.8387	0.8701	0.8276
		BY _C	0.1432	0.0606	0.0321	0.1515	0.1210	0.1414
		WBY _C	0.0455	0.0030	0.0344	0.0000	0.0233	0.0011
15		MLE	2.5759	6.2124	2.4648	6.1113	2.3637	6.2224
		MALLOWS _C	0.2373	0.1087	0.1282	0.0076	0.0181	0.0065
		CBIF _C	0.6822	0.7487	0.8811	0.8386	0.8700	0.8275
		BY _C	0.1431	0.0605	0.0321	0.0504	0.0210	0.0403
		WBY _C	0.0454	0.0010	0.0343	0.0000	0.0232	0.0010
1	20	MLE	2.7288	7.4773	2.6177	7.3662	2.5066	7.2551
		MALLOWS _C	0.4309	0.6467	0.3208	0.5356	0.2107	0.4245
		CBIF _C	1.6603	1.8031	1.5502	2.7020	1.4401	2.6010
		BY _C	0.7688	1.3257	0.6577	1.2146	0.5466	1.1035
		WBY _C	0.0703	0.3217	0.0602	0.2106	0.0501	0.1005
2		MLE	3.7288	8.4773	3.6177	8.3662	3.5066	8.2551
		MALLOWS _C	0.5309	0.7467	0.4208	0.6356	0.3107	0.5245
		CBIF _C	2.6603	3.8031	2.5502	3.7020	2.4401	3.6010
		BY _C	0.8688	2.3257	0.7577	2.2146	0.6466	2.1035
		WBY _C	0.1703	0.4217	0.1602	0.3106	0.1501	0.2005
6		MLE	2.8388	7.5873	2.7277	7.4762	2.6166	7.3651
		MALLOWS _C	0.5209	0.7567	0.4308	0.6456	0.3207	0.5345
		CBIF _C	1.7703	2.9131	1.6602	2.8120	1.5501	2.7110
		BY _C	0.8788	1.4357	0.7677	1.3246	0.6566	1.2135
		WBY _C	0.1803	0.4317	0.1702	0.2106	0.1601	0.2105
10		MLE	3.9388	8.6873	3.8277	8.5762	3.5166	8.4651
		MALLOWS _C	0.4109	0.6467	0.3208	0.5356	0.2107	0.4245
		CBIF _C	0.6703	1.8131	0.5602	1.7120	0.4501	1.6110
		BY _C	0.7688	0.5357	0.6577	0.2246	0.5466	0.1135
		WBY _C	0.0703	0.3217	0.0602	0.1006	0.0501	0.1005
15		MLE	4.0488	9.7973	4.9377	9.6862	4.6266	9.5751
		MALLOWS _C	0.4008	0.6356	0.3107	0.5245	0.2006	0.4134
		CBIF _C	0.6602	1.8020	0.5501	1.7010	0.4400	1.6000
		BY _C	0.7577	0.5246	0.6466	0.2135	0.5355	0.1024
		WBY _C	0.0602	0.3106	0.0501	0.1005	0.0400	0.1004

TABLE 3: Bias and MSE of all methods for data with various percentages of misclassified errors (30% and 40%).

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	30	MLE	2.8247	8.0053	3.7247	7.0043	2.6147	6.0032
		MALLOWS _C	0.4354	0.5614	0.3254	0.6514	0.2154	0.5414
		CBIF _C	2.0318	4.1658	1.0208	3.1548	1.0108	3.0448
		BY _C	2.8258	8.0112	2.7158	8.0012	2.6058	8.0001
		WBY _C	0.3752	0.5560	0.2642	0.4450	0.1532	0.3340
MLE		3.8247	9.0053	4.7247	8.0043	3.6147	7.0032	
MALLOWS _C		0.5354	0.6614	0.4254	0.7514	0.3154	0.6414	
CBIF _C		3.0318	5.1658	2.0208	4.1548	2.0108	4.0448	
BY _C		3.8258	9.0112	3.7158	9.0012	3.6058	9.0001	
WBY _C		0.4752	0.6560	0.3642	0.5450	0.2532	0.4340	
MLE		3.9347	9.1153	4.8347	8.1143	3.7247	7.1132	
MALLOWS _C		0.5464	0.6724	0.4364	0.7624	0.3264	0.6524	
CBIF _C		3.0428	5.1768	2.0318	4.1658	2.0218	4.0558	
BY _C		3.8368	9.0222	3.7268	9.0122	3.6168	9.0000	
WBY _C		0.4862	0.6670	0.3752	0.5560	0.2642	0.4450	
MLE		3.9358	9.1164	4.8358	8.1154	3.7258	7.1143	
MALLOWS _C		0.3243	0.4503	0.2143	0.5403	0.1043	0.4303	
CBIF _C		2.0207	4.0547	1.0107	3.0437	1.0007	3.0337	
BY _C		2.7147	8.0001	2.6047	8.0001	2.5047	8.0000	
WBY _C		0.2641	0.6650	0.1531	0.3340	0.0421	0.2230	
MLE		3.8258	9.0064	4.7258	8.0054	3.6158	7.0043	
MALLOWS _C		0.2143	0.3403	0.1043	0.4303	0.0043	0.3203	
CBIF _C		2.1307	4.1647	1.1207	3.1537	1.1107	3.1437	
BY _C		2.8247	8.1101	2.7147	8.1101	2.6147	8.1100	
WBY _C		0.3741	0.7750	0.2631	0.4440	0.1521	0.3330	
MLE	2.8838	8.5320	2.7727	8.4210	2.6616	8.3100		
MALLOWS _C	0.7716	0.8849	0.7605	0.8738	0.7504	0.8627		
CBIF _C	2.4287	5.9337	2.3177	5.0227	2.2066	5.0116		
BY _C	2.8771	8.3148	2.7660	8.2037	2.6550	8.1026		
WBY _C	0.0515	0.3961	0.0404	0.2870	0.0314	0.2980		
MLE	2.9938	8.6420	2.8827	8.5310	2.7716	8.4200		
MALLOWS _C	0.8816	0.9949	0.8705	0.9838	0.8604	0.9727		
CBIF _C	2.5387	5.0437	2.4277	5.1327	2.3166	5.1216		
BY _C	2.9871	8.4248	2.8760	8.3137	2.7650	8.2126		
WBY _C	0.1615	0.4071	0.0515	0.2981	0.0425	0.2091		
MLE	3.0049	9.5421	3.9828	9.6311	3.8717	9.5201		
MALLOWS _C	0.9926	1.0030	0.9817	1.0949	0.9715	1.0838		
CBIF _C	2.8753	5.3704	2.7742	5.2713	2.6631	5.1612		
BY _C	2.7189	8.4842	2.6087	8.3731	2.5076	8.2621		
WBY _C	0.5161	0.7140	0.1505	0.8129	0.2504	0.2013		
MLE	3.1150	9.6532	3.0939	9.7422	3.9828	9.6312		
MALLOWS _C	1.0937	1.1141	1.9817	1.1050	1.0626	1.9747		
CBIF _C	2.9863	5.4817	2.8853	5.3824	2.7741	5.2723		
BY _C	2.8290	8.5953	2.7198	8.4842	2.6187	8.3732		
WBY _C	0.6272	0.8251	0.2616	0.9230	0.3615	0.3124		

contamination. The BY_C estimator can tolerate up to 30% contamination. However, the performance of WBY_C and MALLOWS_C is equally good: they can withstand up to 40% contamination.

6. Application to Real Data

Two real datasets are considered to illustrate the performance of the proposed robust estimators. The results of the estimated coefficients with their standard errors (SEs) and p values are presented. A good estimator has small SE and the

lower p values. The analyzed datasets are available from the corresponding author upon request.

6.1. Meteorological Data. A total of 365 daily observations of rainfall (yes, no) and wind directions are measured in degrees at 0.00 a.m. from January 1 to December 31, 2019. The data are obtained from the Palestinian Meteorological Department for the Hebron meteorology station in the West Bank, Palestine, with geographical coordinates (latitude = 31°32 N and longitude = 35°06 W) and 1005 m above sea

TABLE 4: Bias and MSE of all methods for data with various percentages of good leverage points (5%, 10%, and 20%).

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	5	MLE	0.8072	0.8122	0.8162	0.8232	0.8272	0.8342
		MALLOWS _C	0.8115	0.8188	0.8225	0.8298	0.8105	0.8188
		CBIF _C	0.8122	0.8219	0.8232	0.8329	0.8342	0.8439
		BY _C	0.8138	0.8356	0.8248	0.8466	0.8358	0.8576
		WBY _C	0.8209	0.8623	0.8319	0.8733	0.8429	0.8843
MLE		0.8720	0.8221	0.8621	0.8322	0.8722	0.8423	
MALLOWS _C		0.8151	0.8881	0.8252	0.8982	0.8051	0.8881	
CBIF _C		0.8221	0.8192	0.8322	0.8293	0.8423	0.8394	
BY _C		0.8381	0.8563	0.8482	0.8664	0.8583	0.8765	
WBY _C		0.8092	0.8236	0.8193	0.8337	0.8294	0.8438	
MLE		0.8172	0.9222	0.9262	0.9332	0.9372	0.9442	
MALLOWS _C		0.9215	0.9288	0.9325	0.9398	0.9205	0.9288	
CBIF _C		0.9222	0.9319	0.9332	0.9429	0.9442	0.9539	
BY _C		0.9238	0.9456	0.9348	0.9566	0.9458	0.9676	
WBY _C		0.9309	0.9723	0.9419	0.9833	0.9529	0.9943	
MLE		0.8712	0.8222	0.8622	0.8332	0.8732	0.8442	
MALLOWS _C		0.8152	0.8828	0.8253	0.8938	0.8025	0.8828	
CBIF _C		0.8222	0.8139	0.8332	0.8249	0.8442	0.8359	
BY _C		0.8328	0.8546	0.8438	0.8656	0.8548	0.8766	
WBY _C		0.8039	0.8273	0.8149	0.8383	0.8259	0.8493	
MLE		0.8217	0.8223	0.8226	0.8233	0.8237	0.8244	
MALLOWS _C		0.8251	0.8288	0.8352	0.8839	0.8520	0.8228	
CBIF _C		0.8223	0.8931	0.8233	0.8942	0.8244	0.8953	
BY _C		0.8823	0.8645	0.8834	0.8566	0.8845	0.8667	
WBY _C		0.8930	0.8372	0.8941	0.8833	0.8952	0.8394	
MLE	0.8035	0.8096	0.7135	0.7196	0.7025	0.7086		
MALLOWS _C	0.8083	0.8183	0.7183	0.7282	0.7291	0.7393		
CBIF _C	0.8060	0.8143	0.7160	0.7243	0.7270	0.7353		
BY _C	0.8121	0.8416	0.7221	0.7316	0.7331	0.7426		
WBY _C	0.8118	0.8506	0.7218	0.7606	0.7128	0.7066		
MLE	0.8355	0.8966	0.7355	0.7966	0.7255	0.7866		
MALLOWS _C	0.8833	0.8823	0.7833	0.7822	0.7911	0.7933		
CBIF _C	0.8600	0.8433	0.7600	0.7433	0.7700	0.7533		
BY _C	0.8211	0.8166	0.7211	0.7166	0.7311	0.7266		
WBY _C	0.8188	0.8066	0.7188	0.7066	0.7288	0.7666		
MLE	0.9153	0.9196	0.8153	0.8169	0.8152	0.8168		
MALLOWS _C	0.9138	0.9238	0.8238	0.8228	0.8219	0.8339		
CBIF _C	0.9106	0.9134	0.8106	0.8234	0.8207	0.8335		
BY _C	0.9112	0.9461	0.8212	0.8361	0.8313	0.8462		
WBY _C	0.9181	0.9560	0.8281	0.8660	0.8182	0.8065		
MLE	0.8145	0.8107	0.7245	0.7207	0.7135	0.7196		
MALLOWS _C	0.8193	0.8293	0.7293	0.7392	0.7302	0.7404		
CBIF _C	0.8170	0.8253	0.7270	0.7353	0.7380	0.7463		
BY _C	0.8231	0.8526	0.7331	0.7426	0.7441	0.7536		
WBY _C	0.8228	0.8616	0.7328	0.7716	0.7238	0.7176		
MLE	0.8254	0.8207	0.7354	0.7370	0.7253	0.7269		
MALLOWS _C	0.8239	0.8339	0.7339	0.7429	0.7420	0.7540		
CBIF _C	0.8207	0.8335	0.7307	0.7435	0.7408	0.7536		
BY _C	0.8313	0.8662	0.7413	0.7562	0.7514	0.7663		
WBY _C	0.8382	0.8761	0.7482	0.7861	0.7383	0.7267		

TABLE 4: Continued.

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	20	MLE	0.7910	0.7903	0.6091	0.6094	0.6102	0.6105
		MALLOWS _C	0.7954	0.7979	0.6064	0.6089	0.6175	0.6190
		CBIF _C	0.7911	0.7922	0.6811	0.6822	0.6722	0.6733
		BY _C	0.8019	0.7867	0.7119	0.6967	0.7220	0.6078
		WBY _C	0.7867	0.8101	0.6767	0.7001	0.6656	0.7223
2		MLE	0.7911	0.7922	0.6811	0.6822	0.6722	0.6733
		MALLOWS _C	0.8019	0.7867	0.7119	0.6967	0.7220	0.6078
		CBIF _C	0.7910	0.7903	0.6091	0.6094	0.6102	0.6105
		BY _C	0.7954	0.7979	0.6064	0.6089	0.6175	0.6190
		WBY _C	0.7786	0.8011	0.6677	0.7010	0.6566	0.7232
6		MLE	0.7910	0.7903	0.6091	0.6094	0.6102	0.6105
		MALLOWS _C	0.7954	0.7979	0.6064	0.6089	0.6175	0.6190
		CBIF _C	0.7911	0.7922	0.6811	0.6822	0.6722	0.6733
		BY _C	0.8019	0.7867	0.7119	0.6967	0.7220	0.6078
		WBY _C	0.7867	0.8101	0.6767	0.7001	0.6656	0.7223
10		MLE	0.8020	0.8013	0.7101	0.7104	0.7212	0.7215
		MALLOWS _C	0.8064	0.8089	0.7174	0.7199	0.7285	0.7200
		CBIF _C	0.8021	0.8032	0.7921	0.7932	0.7832	0.7843
		BY _C	0.9129	0.8977	0.8229	0.7077	0.8330	0.7188
		WBY _C	0.8977	0.9211	0.7877	0.8111	0.7766	0.8333
15		MLE	0.8131	0.8124	0.7212	0.7215	0.7323	0.7326
		MALLOWS _C	0.8175	0.8190	0.7285	0.7200	0.7396	0.7311
		CBIF _C	0.8132	0.8143	0.7032	0.7043	0.7943	0.7954
		BY _C	0.9231	0.8088	0.8330	0.7188	0.8221	0.7299
		WBY _C	0.8088	0.9322	0.7988	0.8211	0.7877	0.8433

TABLE 5: Bias and MSE of all methods for data with various percentages of good leverage points (30% and 40%).

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	30	MLE	0.8089	0.8392	0.7189	0.7492	0.7989	0.7922
		MALLOWS _C	0.8124	0.8452	0.7224	0.7552	0.7244	0.7522
		CBIF _C	0.8111	0.8442	0.7211	0.7542	0.7112	0.7425
		BY _C	0.8150	0.8601	0.7250	0.7701	0.7052	0.7107
		WBY _C	0.8109	0.8632	0.7209	0.7732	0.7092	0.7327
MLE		0.8190	0.8403	0.7290	0.7503	0.7090	0.7033	
2		MALLOWS _C	0.8235	0.8562	0.7335	0.7663	0.7355	0.7633
		CBIF _C	0.8221	0.8553	0.7322	0.7653	0.7223	0.7536
		BY _C	0.8261	0.8712	0.7361	0.7812	0.7163	0.7218
		WBY _C	0.8210	0.8743	0.7310	0.7843	0.7103	0.7438
		MLE	0.8890	0.8923	0.7891	0.7924	0.7899	0.7229
6		MALLOWS _C	0.8241	0.8524	0.7242	0.7525	0.7442	0.7215
		CBIF _C	0.8231	0.8424	0.7112	0.7425	0.7121	0.7254
		BY _C	0.8501	0.8016	0.7502	0.7017	0.7520	0.7071
		WBY _C	0.8091	0.8326	0.7092	0.7327	0.7920	0.7273
		MLE	0.8981	0.9012	0.7980	0.8013	0.7988	0.7337
10		MALLOWS _C	0.8352	0.8635	0.7353	0.7636	0.7553	0.7326
		CBIF _C	0.8342	0.8535	0.7223	0.7536	0.7232	0.7365
		BY _C	0.8612	0.8127	0.7613	0.7128	0.7631	0.7182
		WBY _C	0.8102	0.8437	0.7103	0.7438	0.8031	0.7384
		MLE	0.9189	0.9123	0.8089	0.9031	0.8889	0.8373
10		MALLOWS _C	0.9235	0.9356	0.8235	0.8356	0.8355	0.8236
		CBIF _C	0.9234	0.9355	0.8322	0.8365	0.8223	0.8256
		BY _C	0.9162	0.9271	0.8136	0.8281	0.8136	0.8281
		WBY _C	0.9021	0.9374	0.8031	0.8384	0.9103	0.8438
		MLE	0.9982	0.9232	0.8981	0.9311	0.8899	0.8734
15		MALLOWS _C	0.9353	0.9564	0.8353	0.8564	0.8554	0.8363
		CBIF _C	0.9343	0.9554	0.8224	0.8654	0.8233	0.8563
		BY _C	0.9622	0.9713	0.8362	0.8813	0.8362	0.8813
		WBY _C	0.9211	0.9744	0.8311	0.8844	0.9032	0.8384

TABLE 5: Continued.

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	40	MLE	0.8162	0.8421	0.7272	0.7311	0.7382	0.7421
		MALLOWS _C	0.8216	0.8519	0.7326	0.7629	0.7436	0.7739
		CBIF _C	0.8186	0.8458	0.7286	0.7558	0.7396	0.7658
		BY _C	0.8454	0.9045	0.7564	0.8035	0.7674	0.8145
		WBY _C	0.8463	0.9096	0.7573	0.9107	0.7683	0.9217
2		MLE	0.8273	0.8532	0.7383	0.7422	0.7493	0.7532
		MALLOWS _C	0.8327	0.8620	0.7437	0.7730	0.7546	0.7840
		CBIF _C	0.8297	0.8569	0.7397	0.7669	0.7407	0.7769
		BY _C	0.8565	0.9156	0.7674	0.8146	0.7785	0.8256
		WBY _C	0.8574	0.9107	0.7684	0.9218	0.7794	0.9328
6		MLE	0.8373	0.8326	0.7834	0.7225	0.7935	0.7326
		MALLOWS _C	0.8274	0.8207	0.7375	0.7308	0.7466	0.7409
		CBIF _C	0.8973	0.8696	0.7974	0.7697	0.7075	0.7698
		BY _C	0.8656	0.9562	0.7747	0.8462	0.7858	0.853
		WBY _C	0.8746	0.9072	0.7847	0.9183	0.7948	0.9284
10		MLE	0.9162	0.9421	0.8474	0.8311	0.8382	0.8421
		MALLOWS _C	0.9216	0.9511	0.8326	0.8621	0.8435	0.8731
		CBIF _C	0.9186	0.9458	0.8287	0.8558	0.8316	0.8650
		BY _C	0.9454	0.9267	0.8563	0.9035	0.8674	0.9145
		WBY _C	0.9463	0.9218	0.8574	0.9107	0.8684	0.9217
15		MLE	0.9273	0.9532	0.8585	0.8422	0.8493	0.8532
		MALLOWS _C	0.9327	0.9622	0.8437	0.8732	0.8546	0.8842
		CBIF _C	0.9297	0.9569	0.8398	0.8669	0.8427	0.8761
		BY _C	0.9566	0.9368	0.8674	0.9146	0.8785	0.9256
		WBY _C	0.9352	0.9107	0.8463	0.9006	0.8573	0.9106

TABLE 6: Bias and MSE of all methods for data with various percentages of bad leverage points (5%, 10%, and 20%).

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	5	MLE	1.4570	2.1918	1.3461	2.2028	1.2350	2.1017
		MALLOWS _C	0.1400	0.3073	0.2511	0.4184	0.2410	0.4293
		CBIF _C	0.5148	0.4482	0.6259	0.5593	0.5159	0.4472
		BY _C	0.1344	0.3765	0.1455	0.3654	0.1344	0.3543
		WBY _C	0.1745	0.3842	0.1634	0.3732	0.1523	0.3621
2		MLE	1.6681	2.3829	1.5572	2.4139	1.4461	2.3128
		MALLOWS _C	0.3511	0.5184	0.4622	0.6295	0.4521	0.6384
		CBIF _C	0.7259	0.6593	0.8368	0.7684	0.7268	0.6583
		BY _C	0.3455	0.5874	0.3666	0.5765	0.3455	0.5654
		WBY _C	0.3856	0.5953	0.3745	0.5843	0.3634	0.4732
6		MLE	1.5571	2.2719	1.4462	2.3029	1.3351	2.2018
		MALLOWS _C	0.2401	0.4074	0.3512	0.5185	0.3411	0.5274
		CBIF _C	0.6149	0.5483	0.7258	0.6574	0.6158	0.5473
		BY _C	0.2345	0.4764	0.2556	0.4655	0.2345	0.4544
		WBY _C	0.2746	0.4843	0.2635	0.4733	0.2524	0.3622
10		MLE	1.5560	2.2708	1.4451	2.3018	1.3340	2.2007
		MALLOWS _C	0.2400	0.4063	0.3501	0.5174	0.3400	0.5263
		CBIF _C	0.6138	0.5472	0.7247	0.6563	0.6147	0.5462
		BY _C	0.2334	0.4753	0.2545	0.4644	0.2334	0.4533
		WBY _C	0.2735	0.4832	0.2624	0.4722	0.2513	0.3611
15		MLE	1.4450	2.1607	1.3341	2.2007	1.2230	2.1106
		MALLOWS _C	0.1310	0.3052	0.2400	0.4063	0.2310	0.4152
		CBIF _C	0.5027	0.5361	0.6136	0.5452	0.5036	0.4351
		BY _C	0.1233	0.3643	0.1434	0.3533	0.1223	0.3422
		WBY _C	0.1624	0.3721	0.1513	0.3611	0.1402	0.2500

TABLE 6: Continued.

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	10	MLE	2.4648	6.1013	2.3538	6.0903	2.2428	6.1813
		MALLOWS _C	0.3484	0.2098	0.2374	0.1988	0.2374	0.1988
		CBIF _C	1.1933	1.4598	1.0822	1.3487	1.0711	1.2376
		BY _C	0.2542	0.1716	0.1431	0.0605	0.0320	0.1514
		WBY _C	0.0565	0.1120	0.1454	0.0010	0.0343	0.0120
MLE		2.4846	6.1310	2.3835	6.0309	2.2824	6.1318	
MALLOWS _C		0.4584	0.2098	0.3474	0.2888	0.3474	0.2888	
CBIF _C		1.2833	1.5698	1.1722	1.4587	1.1811	1.3476	
BY _C		0.3642	0.2816	0.2531	0.1705	0.1420	0.2614	
WBY _C		0.0675	0.1230	0.1564	0.0120	0.0453	0.0220	
MLE		2.4956	6.1420	2.3945	6.0419	2.2934	6.1428	
MALLOWS _C		0.4694	0.2188	0.3584	0.2998	0.3584	0.2998	
CBIF _C		1.2943	1.5788	1.1832	1.4697	1.1921	1.3586	
BY _C		0.3752	0.2926	0.2641	0.1815	0.1530	0.2724	
WBY _C		0.0785	0.1340	0.1674	0.0230	0.0563	0.0320	
MLE		1.5866	5.0310	1.4035	5.2529	2.1824	6.2538	
MALLOWS _C		0.3505	0.1099	0.4695	0.1809	0.4673	0.3009	
CBIF _C		1.3053	1.6898	1.2942	1.5707	1.2031	1.3586	
BY _C		0.4862	0.3036	0.3751	0.2925	0.2640	0.3834	
WBY _C		0.0896	0.1451	0.1785	0.0341	0.0674	0.0431	
MLE		1.6977	5.1421	1.5146	5.3630	2.2935	6.3549	
MALLOWS _C		0.4616	0.2188	0.5706	0.2918	0.5784	0.4108	
CBIF _C		1.4164	1.7999	1.3953	1.6818	1.3142	1.4697	
BY _C		0.5973	0.4147	0.4861	0.3836	0.3751	0.4945	
WBY _C		0.1985	0.1662	0.1996	0.0552	0.0874	0.0541	
MLE	2.7288	7.4773	2.9308	7.6983	2.0918	7.7093		
MALLOWS _C	0.4309	0.6467	0.5410	0.7578	0.5521	0.7689		
CBIF _C	1.6603	2.8031	1.5512	2.7120	0.7623	1.9231		
BY _C	0.7688	1.3257	0.8876	1.5732	0.6987	0.4621		
WBY _C	0.0703	0.3217	0.0814	0.3328	0.0925	0.3439		
MLE	2.6178	7.3663	2.8218	7.5873	2.1808	7.6183		
MALLOWS _C	0.3219	0.5357	0.4300	0.6468	0.4411	0.6579		
CBIF _C	1.5413	2.7121	1.4322	2.6010	0.6513	1.8121		
BY _C	0.6578	1.2147	0.7766	1.4622	0.5877	0.3511		
WBY _C	0.0612	0.3106	0.0703	0.3217	0.0814	0.3328		
MLE	2.7288	7.4773	2.9328	7.6983	2.2918	7.7293		
MALLOWS _C	0.4329	0.6467	0.5410	0.7578	0.5521	0.7689		
CBIF _C	1.6523	2.8231	1.5432	2.7120	0.7623	1.9231		
BY _C	0.7688	1.3257	0.8876	1.5732	0.6987	0.4621		
WBY _C	0.0723	0.3217	0.0812	0.3106	0.0703	0.3217		
MLE	2.7288	7.4773	2.9328	7.6983	2.2918	7.7293		
MALLOWS _C	0.4329	0.6467	0.5410	0.7578	0.5521	0.7689		
CBIF _C	1.6523	2.8231	1.5432	2.7120	0.7623	1.9231		
BY _C	0.7688	1.3257	0.8876	1.5732	0.6987	0.4621		
WBY _C	0.0723	0.3217	0.0812	0.3106	0.0703	0.3217		
MLE	3.6178	6.3663	1.8418	6.5873	1.1808	6.6183		
MALLOWS _C	0.5439	0.7567	0.6520	0.8678	0.6631	0.8789		
CBIF _C	2.5412	1.7120	0.7543	1.8231	1.8534	2.8120		
BY _C	1.6577	1.6368	1.7985	2.4843	1.5876	0.6510		
WBY _C	0.1612	0.4106	0.1701	0.4105	0.1612	0.4106		

level. Forty-eight rainy days are reported in 2019 in Hebron governorate. Figure 1 shows the circular plot of wind direction data in Hebron.

Figure 2 suggests no outlier is found in the data (clean dataset). Abuzaid and ElShekh Ahmed [11] have not identified any outliers by their PMLE method and refereed

that to the low concentration parameter of the wind direction ($\hat{\kappa} = 0.883$).

Table 8 shows the estimated parameters and estimated SE for the different estimators. As observed in Table 8, all the five estimators are fairly close to each other. Thus, the model of the relation between the rainfall as a binary response and

TABLE 7: Bias and MSE of all methods for data with various percentages of bad leverage points (30% and 40%).

κ	% of misc error	Methods	$N = 200$		$N = 250$		$N = 300$	
			Bias	MSE	Bias	MSE	Bias	MSE
1	30	MLE	2.8247	8.0053	2.7137	8.0042	1.8248	9.1153
		MALLOWS _C	0.4354	0.5614	0.3243	0.4503	0.2132	0.3402
		CBIF _C	2.0318	4.1658	2.0207	4.0548	1.1318	5.1659
		BY _C	2.8258	8.0112	2.7147	8.0001	1.6036	7.9880
		WBY _C	0.3752	0.5560	0.3641	0.4451	0.2530	0.3340
2		MLE	2.7136	8.1142	2.7226	8.1131	1.9359	9.2242
		MALLOWS _C	0.5443	0.6703	0.4332	0.5602	0.3221	0.4501
		CBIF _C	2.0207	4.0547	2.1306	4.1637	1.2407	5.2748
		BY _C	2.9347	8.1201	2.8235	8.1100	1.7125	8.0970
		WBY _C	0.4841	0.6651	0.4730	0.5540	0.3620	0.4430
6		MLE	2.6025	8.0031	2.6115	8.0020	1.8248	9.1131
		MALLOWS _C	0.4332	0.5612	0.3221	0.4511	0.2110	0.3410
		CBIF _C	2.1116	4.1436	2.0215	4.0526	1.1316	5.1637
		BY _C	2.8236	8.0110	2.7124	8.0011	1.6014	8.0861
		WBY _C	0.3730	0.5540	0.3621	0.4430	0.2511	0.3321
10		MLE	1.8247	9.2253	3.8337	9.2242	2.0560	9.9352
		MALLOWS _C	0.6554	0.7834	0.5443	0.6733	0.4332	0.5632
		CBIF _C	2.3338	4.3758	2.2437	4.2748	1.3538	5.2859
		BY _C	2.0458	8.0332	2.7346	8.3322	1.8236	8.2983
		WBY _C	0.5952	0.7762	0.5843	0.6652	0.4733	0.5543
15		MLE	1.7137	9.1143	3.7227	9.1132	2.1450	9.8242
		MALLOWS _C	0.5444	0.6724	0.4333	0.5623	0.3222	0.4522
		CBIF _C	2.2228	4.2648	2.1327	4.1638	1.2428	5.1749
		BY _C	2.0347	8.0221	2.6236	8.2212	1.7126	8.1873
		WBY _C	0.4842	0.6652	0.4733	0.5542	0.3623	0.4433
1	40	MLE	2.8838	8.5320	2.7728	8.4210	1.6617	7.3101
		MALLOWS _C	0.7716	0.8849	0.6605	0.7738	0.6514	0.7627
		CBIF _C	2.4287	5.9337	2.3176	5.8226	1.5287	6.9337
		BY _C	2.8771	8.3148	2.7880	8.4037	2.6870	8.3126
		WBY _C	0.0515	0.3961	0.0404	0.2870	0.1313	0.1761
2		MLE	3.9948	9.6530	3.8838	9.5320	2.7727	8.4211
		MALLOWS _C	0.6606	0.7739	0.5515	1.6628	0.5404	0.6517
		CBIF _C	3.3177	6.8227	3.4286	6.9336	2.4177	7.8227
		BY _C	2.8771	8.3148	2.7880	8.4037	2.7970	8.4216
		WBY _C	0.0626	0.3072	0.0515	0.2981	0.1424	0.1872
6		MLE	3.0039	9.7641	3.9949	9.6431	2.8838	8.5322
		MALLOWS _C	0.7707	0.8849	0.6626	0.7739	0.4515	0.7628
		CBIF _C	3.4288	6.9338	3.5397	6.9447	2.5287	7.9337
		BY _C	2.9881	8.4268	2.8991	8.5148	3.6861	7.3105
		WBY _C	0.1737	0.4183	0.1626	0.3092	0.2535	0.2983
10		MLE	2.1128	8.9531	3.1038	9.7520	2.9927	8.6411
		MALLOWS _C	0.8806	0.9938	0.5515	0.8828	0.5604	0.8717
		CBIF _C	3.5378	6.0927	3.6408	6.8536	2.6376	7.0426
		BY _C	2.0870	8.3179	2.7802	8.6259	3.7950	7.4204
		WBY _C	0.2826	0.5272	0.2715	0.4181	0.3624	0.3172
15		MLE	2.2239	8.9642	3.2149	9.7631	2.0017	8.6522
		MALLOWS _C	0.9916	0.9049	0.6625	0.8939	0.6714	0.8828
		CBIF _C	3.6488	6.0038	3.7518	6.8647	2.7486	7.0315
		BY _C	2.1760	8.3288	2.8912	8.6369	3.8060	7.4315
		WBY _C	0.3936	0.5383	0.3815	0.4292	0.4734	0.3283

the wind directions as a circular predictor of the considered dataset is given by

$$\log it(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = -0.7894 + 0.2508 \cos u_i + 0.1789 \sin u_i. \quad (26)$$

6.2. Ecological Data. Leaf inclination angles are extremely important to model the light transmission through the vegetation canopies and to indirectly quantify the canopy attributes such as leaf area index and G-function. Two hundred measurements of leaf inclination angles of *Betula pendula* trees were measured on May 9, 2013, and their canopies either (top or bottom) were recorded [22].

FIGURE 1: Circular plot of daily wind directions in Hebron, 2019.

FIGURE 2: Scatter plot of meteorological data.

TABLE 8: Meteorological data: estimated parameters, SE, and p values.

Estimation	Intercept			cos(wind)			sin(wind)		
Method	Estimation	SE	p value	Estimation	SE	p value	Estimation	SE	p value
MLE	-0.7894	0.1149	≤ 0.001	0.2508	0.1627	0.0123	0.1789	0.1620	0.0269
MALLOWS _C	-0.7886	0.1149	≤ 0.001	0.2523	0.1627	0.0121	0.1802	0.1619	0.0266
CBIF _C	-0.7894	0.1149	≤ 0.001	0.2508	0.1627	0.0123	0.1789	0.1620	0.0269
BY _C	-0.7887	0.1148	≤ 0.001	0.2507	0.1685	0.0137	0.1684	0.1571	0.0284
WBY _C	-0.7875	0.1149	≤ 0.001	0.2513	0.1630	0.0123	0.1666	0.1623	0.0305

Figure 3 shows the circular plot of leaf inclination angles, and Figure 4 shows the scatter plot of the data. Six observations far from the mass of directions were identified as outliers by Abuzaid and ElShekh Ahmed [11]. Table 9 presents the parameter estimates, SE, and p values for the various robust estimators where the CUBIF_C and MALLOWS_C estimates are reasonably close to the MLE. However, BY_C and WBY_C have the smallest SE values. Therefore, BY_C and WBY_C estimators give the best results

for this dataset. Thus, the model of the relation between the vegetation canopy (v) as a binary response and the leaf inclination angles (u) as a circular predictor of the considered dataset is given by

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = -6.8431 + 6.5681 \cos u_i + 4.5951 \sin u_i. \quad (27)$$

FIGURE 3: Circular plot for leaf inclination angles ($N = 200$).

FIGURE 4: Scatter plot of ecological data.

TABLE 9: Ecological data: estimated parameters, standard errors, and p values.

Estimation Method	Intercept			cos(angle)			sin(angle)		
	Estimation	SE	p value	Estimation	SE	p value	Estimation	SE	p value
MLE	-0.0100	0.2523	0.0944	-0.0106	0.1999	0.0958	-0.1472	0.2026	0.0468
MALLOWS _C	-0.0063	0.2524	0.0964	-0.0157	0.2000	0.0937	-0.1428	0.2028	0.0481
CBIF _C	-0.0100	0.2523	0.0943	-0.0106	0.1999	0.0957	-0.1472	0.2026	0.0467
BY _C	-6.8431	0.1524	≤ 0.001	6.5681	0.1487	≤ 0.001	4.5951	0.1326	0.0200
WBY _C	-6.8141	0.1423	≤ 0.001	6.5381	0.1484	≤ 0.001	4.5651	0.1026	0.0100

7. Conclusion

This paper aimed to propose robust estimators for the circular logistic regression model. We have compared the performance of the MLE and the proposed robust estimators for the circular logistic model under clean and contaminated datasets. The findings indicate that the MLE tends to bias in the presence of misclassified error and bad leverage points. A good performance is obtained for the proposed robust estimator in estimating the parameters in such models. Some robust estimators, such as BY_C and WBY_C, show superiority over others depending on the type of contamination.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

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

References

- [1] K. V. Mardia and E. Peter, *Directional Statistics*, John Wiley & Sons, Hoboken, NJ, USA, 2009.
- [2] N. I. Fisher, *Statistical Analysis of Circular Data*, Cambridge University Press, London, UK, 1993.
- [3] A. H. Abuzaid and N. R. Allahham, "Simple circular regression model assuming wrapped cauchy error," *Pakistan Journal of Statistics*, vol. 31, no. 4, pp. 385–398, 2015.
- [4] A. G. Hussin, N. R. J. Fieller, and E. C. Stillman, "Linear regression model for circular variables with application to directional data," *Journal of Applied Science and Technology*, vol. 9, no. 1, pp. 1–6, 2004.

- [5] S. Kato, K. Shimizu, and G. S. Shieh, "A circular-circular regression model," *Statistica Sinica*, vol. 32, pp. 633–645, 2008.
- [6] Y. Sarma and S. Jammalamadaka, "Circular regression," *Statistical Science and Data Analysis*, vol. 34, pp. 109–128, 1993.
- [7] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Wiley, Hoboken, NJ, USA, 2nd edition, 2000.
- [8] K. Al-Daffaie and S. Khan, "Logistic regression for circular data," in *Proceedings of the AIP Conference Proceedings*, Bikaner, India, May 2017.
- [9] V. Feser and M. -Pia, "Robust inference with binary data," *Psychometrika*, vol. 67, no. 1, pp. 21–32, 2002.
- [10] C. Croux, C. Flandre, and G. Haesbroeck, "The breakdown behavior of the maximum likelihood estimator in the logistic regression model," *Statistics & Probability Letters*, vol. 60, no. 4, pp. 377–386, 2002.
- [11] A. H. Abuzaaid and H. I. ElShekh Ahmed, "On outliers detection in circular logistic regression," *Journal of Applied Probability and Statistics*, vol. 16, no. 1, pp. 95–110, 2021.
- [12] S. Ahmad, N. M. Ramli, and H. Midi, "Robust estimators in logistic regression: a comparative simulation study," *Journal of Modern Applied Statistical Methods*, vol. 9, no. 2, p. 18, 2010.
- [13] H. R. Kunsch and L. A. Stefanski, "Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models," *Journal of the American Statistical Association*, vol. 84, no. 6, pp. 460–466, 1989.
- [14] R. J. Carroll and S. Pederson, "On robustness in the logistic regression model," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 55, no. 3, pp. 693–706, 1993.
- [15] A. M. Bianco and V. J. Yohai, "Robust estimation in the logistic regression model," in *Proceedings of the Robust Statistics, Data Analysis, and Computer Intensive Methods*, pp. 17–34, New York, NY, USA, March 1996.
- [16] J. B. Copas, "Binary regression models for contaminated data," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 50, no. 2, pp. 225–253, 1988.
- [17] D. Pregibon, "Resistant fits for some commonly used logistic models with medical applications," *Biometrics*, vol. 38, no. 2, pp. 485–498, 1982.
- [18] S. Mallows and L. Colin, "On some topics in robustness," in *Proceedings of the Unpublished memorandum, Bell Telephone Laboratories*, Murray Hill, NJ, USA, March 1975.
- [19] L. A. Stefanski, "The effects of measurement error on parameter estimation," *Biometrika*, vol. 72, no. 3, pp. 583–592, 1985.
- [20] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, Hoboken, NJ, USA, 2005.
- [21] F. R. Hampel, E. M. Ronchetti, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, Hoboken, NJ, USA, 2011.
- [22] F. Chianucci, P. Jan, K. Raabe, L. Marchino, C. Ferrara, and P. Corona, "A dataset of leaf inclination angles for temperate and boreal broadleaf woody species," *Annals of Forest Science*, vol. 75, no. 2, pp. 1–7, 2018.