Robustness issues in circular-circular regression

Technical Report No. ASU/2017/7
Dated 24 April, 2017

Jayant Jha

and

Atanu Biswas

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Robustness issues in circular-circular regression

Jayant Jha[1], Atanu Biswas[1] and Tsung-Chi Cheng[2]

[1]Applied Statistics Unit, Indian Statistical Institute

203 B. T. Road, Kolkata - 700 108, India

[2]National Chengchi University, Taipei, Taiwan

E-mails: jayantjha@gmail.com, atanu@isical.ac.in, chengt@nccu.edu.tw

Abstract

In this paper, we attempt to address the robustness issues in circular-circular regression. We consider the Möbius transformation based circular-circular regression model of Kato *et al.* (2008). Then, we discuss the robustness issue of the estimators in this model. We propose maximum trimmed cosine estimator in this context and discuss the algorithm for its computation. We also discuss some properties including the breakdown point of the estimators. Simulation studies and a real data analysis are used to illustrate the proposed methodology.

1 Introduction

Circular data or directional data is a relatively unexplored area of statistics despite having numerous applications in meterology (e.g. Kato *et.al.*, 2008), astronomy (e.g. Protheroe, 1985), geophysics (e.g. Chang *et. al.*, 1990) and opthalomology (e.g. Biswas *et al.*, 2015). Some of the common examples of circular random variables are concerned with the directions, e.g. direction of migration of birds (e.g. Busse and Trociñska, 1999), wind direction (e.g. Kato *et al.*, 2008), etc. Other not so obvious examples are the periodic random variables such as the date of the year and the time of the day (e.g. Jha *et.al.*, 2016). Due to the difference in topology, the usual methods applicable for linear random variables can not be directly applied in the case of circular random variables. Mardia and Jupp (2000) gives a comprehensive review on the analysis of circular random variables.

Circular-circular regression is the modelling used to explain the relationship between a circular covariate and a circular response variable. There have been numerous studies on robustness of linear regression models, but in the case of circular-circular regression the existing literature primarily consists of regression modelling. So far our knowledge goes, the robustness issue has not been considered in the literature for the circular-circular regression models. The problem of outlier detection for such models was considered in Ibrahim *et al.* (2013) where they used the COVRATIO statistic to identify the outliers. But, they did not consider robust circular-circular regression. The trimming techniques used in robust linear regression modelling have not been considered in circular-circular regression modelling, as far our knowledge goes.

In this paper, we consider the circular-circular regression model of Kato *et al.* (2008). We try to address the issues of robustness in this model by using a *Maximum Trimmed Cosine* approach in the spirit of the *Least Trimmed Square* residual as mentioned in Rosseuw (1984). In Section 2, we illustrate the Möbius transformation based circular-circular regression model of Kato *et al.* (2008). Then, we shall show why the robustness problem is generally relevant for this model with special reference to von Mises distribution of the angular error. In fact, Kato *et al.* (2008) analysed a data set on wind direction in Milwaukee where the sample size was 21; and in their spokeplot (cf. Zubairi *et al.*, 2008) Kato *et al.* (2008) observed that the number of outliers in the dataset is 5, which is about 24% of the data size. However, they did not carry out any kind of robust analysis, possibly due to the unavailability of any methodology in the literature. Note that robust estimation for circular data has got some attention of researchers (see, e.g., Fisher, 1982; Ko, 1992; Agostinelli, 2007; Kato and Eguchi, 2014; Laha and Mahesh, 2012). Also Otieno and Anderson-Cook (2006) pointed out the existence of number of outliers in several real datasets. However, so far our knowledge goes, no work on robust circular-circular regression is available in the literature. The present paper aims at fulfilling that gap. The Milwaukee data set is certainly a good example where robust statistical methods are needed in circular-circular regression.

In Section 3, the Maximum trimmed cosine (MTC) estimator is defined and the breakdown point of the estimator is discussed. Some computational algorithm of the proposed estimator is discussed in Section 4. Simulations results are presented in Section 5. In Section 6, we revisit the Milwaukee data set to illustrate our proposed estimators. Section 7 concludes.

2 Breakdown point of the model

2.1 The model

The circular-circular regression considered in Kato *et al.* (2008) is a Möbius transformation based model where the circular covariate and the circular response are considered to be complex numbers with unit Modulus. The regression function defined in the model is the following:

$$y = \beta_0 \frac{x + \beta_1}{1 + \overline{\beta_1} x} \epsilon, \tag{2.1}$$

where $y \in \Omega$ is the response, $x \in \Omega$ is the covariate, $\beta_0, \epsilon \in \Omega$ and $\beta_1 \in \mathbb{C}$. Here $\Omega = \{z : |z| = 1\}$. In this model, $\arg(\epsilon)$ is the angular error which follows a distribution on the unit circle with mean direction 0 degree. When $|\beta_1|| = 1$, the above model gives no direction at $x == \beta_1$. This problem can be resolved with geometrical continuity by defining $y = \beta_0 \beta_1 \epsilon$ for all $x \in \Omega$ whenever $|\beta_1| = 1$. This definition at $|\beta_1| = 1$ is used and should be understood like this only throughout the paper. The geometry of this model is also shown in Kato *et al.* (2008) where they have shown different geometries for the case of $|\beta_1| < 1$ and $|\beta_1| > 1$. The two cases can be combined together by the same geometry which is explained in Kato *et al.* (2008) for $||\beta_1|| < 1$.

Circular-circular regression can be defined and interpreted as follows. Let $\theta_x, \theta_y \in \Omega$ be the covariate and the response, respectively, and let x, y are the corresponding unit complex numbers. The circular-circular regression model (2.1) used the Wrapped Cauchy distribution due to some of its elegant properties. The regression function is a form of Möbius transformation; it is a mapping of unit circle $|z| = 1$ onto itself. In the model (2.1), β_0 is a rotation parameter while β_1 is a fixed point in the complex plane which converts the point x to a point x_β on the unit circle which is the intersection of the unit circle $|z| = 1$ with the line joining $-x$ and β_1. Considering $\mu(\beta_0, \beta_1, x) = \beta_0 \frac{x+\beta_1}{1+\overline{\beta_1} x}$, we have $x_\beta = \mu(1, \beta_1, x)$. See Figure 1. The case of $|\beta_1| > 1$ is also discussed in the paper of Kato *et al.* (2008), where first the fixed point is taken to be $\frac{1}{\overline{\beta_1}}$ and then it is joined to $\frac{\beta_1}{|\beta_1|} \frac{\beta_1}{|\beta_1|} \overline{x}$. The intersection of this line with the unit circle $|z| = 1$ is taken as x_β. Thus, if $|\beta_1|$ is closer to 1, the function results in the resultant values of x_β concentrating around $\frac{\beta_1}{|\beta_1|}$. The role of β_0 is then a rotation of

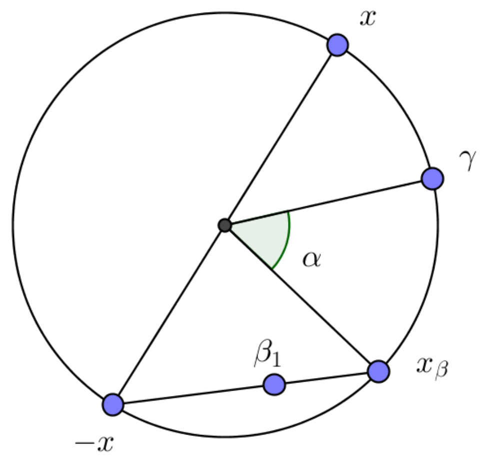

Figure 1: Circular-circular regression model.

x_β to an angle $\alpha = \arg(\beta_0)$ on the unit circle. Thus, $\mu(\beta_0, \beta_1, x)$ is the rotation of $\mu(1, \beta_1, x)$ by $\arg(\beta_0)$. The resultant point is γ, say. Then, y has the error distribution given by ϵ, but with the circular mean given by γ. For example, if ϵ follows a Wrapped Cauchy distribution with circular mean 0, then y follows a Wrapped Cauchy distribution with circular mean γ. See Figure 1 for the illustration of the model (2.1). Clearly, when $|\beta_1| = 0$, we have $\gamma = \beta_0 x$; which is just a rotation of x to an angle $\alpha = \arg(\beta_0)$ on the unit circle. This model is a Möbius transformation model which combines the rotation with the projection via a point β_1.

2.2 Breakdown point

Before proceeding to address the robustness issue in this model, first it is necessary to check if such a problem is at all relevant for the circular-circular regression case. The Breakdown point was first introduced in Donoho and Huber (1983). Finite sample *breakdown point* (BDP), defined in Theorem 1 of Neykov and Neytchev (1990), is the measure of robustness of the estimators in a regression model. The breakdown point is defined in terms of the smallest fraction of observations which can be contaminated such that the estimator goes arbitrarily far from the estimator based on all the observations. Formally,

$$BDP = \inf\left\{\frac{m}{n} : \sup ||T(X) - T'(X)|| = \infty\right\}.$$

Here $T(X)$ is the estimator based on all the n observations, while $T'(X)$ is the estimator when m out of n observations are contaminated arbitrarily. The BDP of OLS in the simple linear regression setup is $\frac{1}{n}$. Continuing in similar fashion, the robustness of the MLE of β_1 can be judged from the BDP of the estimator. For obtaining the BDP, we shall first show that the parameter β_1 can take any arbitrary value based on an exact fit from three points.

Suppose (x_1, y_1) and (x_2, y_2) be two points. Then, different values of β_0 give different values of β_1 based on the intersection of the line joining $-x_2$ and y_2 and the line joining $-x_1$ and y_1. These all values of β_1 give an exact fit for these two points for different values of β_0. The locus of all such β_1 can be given by the equation

$$\frac{y_1(1+\overline{\beta_1}x_1)}{x_1+\beta_1} = \frac{y_2(1+\overline{\beta_1}x_2)}{x_2+\beta_1}'. \tag{2.2}$$

Next, we shall show that these values of β_1 can be unbounded.

Theorem 1: *The solution of β_1 in the equation (2.2) can be unbounded.*

Proof: Without loss of generality, let $x_1 = y_1 = 1$. Then, (2.2) reduces to

$$\frac{1+\overline{\beta_1}}{1+\beta_1} = \frac{y_2(1+\overline{\beta_1}x_2)}{x_2+\beta_1} = \frac{y_2(1+\overline{\beta_1}x_2)}{x_2(1+\beta_1\overline{x_2})} = \frac{B(1+\overline{\beta_1}x_2)}{1+\beta_1\overline{x_2}},$$

where $\frac{y_2}{x_2} = B$. Clearly, $|B| = |x_2| = 1$. Thus,

$$(1+\overline{\beta_1})(1+\beta_1\overline{x_2}) = B(1+\overline{\beta_1}x_2)(1+\beta_1).$$

Defining $Z = (1+\overline{\beta_1})(1+\beta_1\overline{x_2})$, the above equation becomes

$$Z = B\overline{Z}.$$

Let $B = 1$. Then, we get $Z = \overline{Z}$, implying that Z is real and hence its imaginary part is 0. Let $\beta_1 = re^{i\theta_1}$ and $x_2 = e^{i\theta_{x,2}}$. Then, the imaginary part of Z is $Im(Z) = -r^2\sin\theta_{x,2} - r\sin\theta_1 + r\sin(\theta_1 - \theta_{x,2})$. Now, $Im(Z) = 0$ implies either $r = 0$ or $r = \frac{\sin(\theta_1-\theta_{x,2})-\sin(\theta_1)}{\sin(\theta_{x,2})}$, which can be unbounded as the denominator can go to 0. □

If the angular error in the regression model follows a von Mises distribution, then the MLE of the parameters of the model can be found in the following way. Let us assume that L is the log likelihood based on a sample of size n. Then,

$$L = -n\log 2\pi - n\log I_0(\kappa) - \kappa\sum_{j=1}^{n}\cos(\theta_{y_{p,j}} - \theta_{y_j}), \tag{2.3}$$

where $\theta y_{p,j}$ is the predicted jth angle based on x_j, θ_{y_j} is the observed jth angle, κ is the concentration parameter, and $I_u(\kappa)$ is the modified Bessel function of the first kind of order $u \geq 0$. Maximizing L we get $\widehat{\kappa} = A^{-1}\left(\frac{\sum_{j=1}^{n}\cos(\theta_{y_{p,j}}-\theta_{y_j})}{n}\right)$, where $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$.

Now, $\theta_{y_{p,j}} = \theta_0 + \arg\left(\frac{x_j+\beta_1}{1+\overline{\beta_1}x_j}\right)$, where $\theta_0 = \arg(\beta_0)$. Now, the MLE of θ_0 can be found from the MLE of β_1 by taking the circular mean of $\arg(\frac{y_j(1+\overline{\widehat{\beta}_1}x_j)}{x_j+\widehat{\beta}_1})$, and thus we have to maximize $\left|\sum_{j=1}^{n}\frac{y_j(1+\overline{\beta}_1x_j)}{x_j+\beta_1}\right|^2$. Now,

$$\left|\sum_{j=1}^{n}\frac{y_j(1+\overline{\widehat{\beta}_1}x_j)}{x_j+\widehat{\beta}_1}\right|^2 = n + 2\sum_{j=1}^{n}\sum_{l>j}^{n}\cos\left(\theta_{y_j}-\theta_{y_l}-\arg\left(\frac{x_j+\beta_1}{1+\overline{\beta_1}x_j}\right)+\arg\left(\frac{x_l+\beta_1}{1+\overline{\beta_1}x_l}\right)\right).$$

Thus, we have to maximize

$$S(\beta_1) = \sum_{j=1}^{n}\sum_{l>j}^{n}\cos\left(\theta_{y_j}-\theta_{y_l}-\arg\left(\frac{x_j+\beta_1}{1+\overline{\beta_1}x_j}\right)+\arg\left(\frac{x_l+\beta_1}{1+\overline{\beta_1}x_l}\right)\right).$$

The BDP can thus be defined in terms of the BDP of β_1. Proceeding as per the definition of BDP, the BDP of MLE of β_1 is

$$\inf\left\{\frac{m}{n} : \sup||\widehat{\beta}_1 - \widehat{\beta}_1'|| = \infty\right\},$$

where $\widehat{\beta}_1$ is the estimator based on n observations and $\widehat{\beta}_1'$ is the estimator when m out of n observations are contaminated.

Theorem 1 implies that a third pair of observation can always be chosen based on any value of β_1 which gives an exact fit. Hence, the contamination in (x_3, y_3) can make the MLE of β_1 unbounded.

Theorem 2: *The BDP of MLE of β_1 in the circular-circular regression model when the angular error follows von Mises distribution is $\frac{1}{n}$.*

Proof: Let the BDP of $\widehat{\beta}_1$ be greater than $\frac{1}{n}$. Then, for any change in the value of (x_n, y_n), the MLE of β_1 will remain in a compact set.

Let

$$C(\beta_1) = \sum_{j=1}^{n-1}\sum_{l>j}\cos\left(\theta_{y_j}-\theta_{y_l}-\arg\left(\frac{x_j+\beta_1}{1+\overline{\beta_1}x_j}\right)+\arg\left(\frac{x_l+\beta_1}{1+\overline{\beta_1}x_l}\right)\right),$$

$$D(\beta_1) = \sum_{l=1}^{n-1}\cos\left(\theta_{y_n}-\theta_{y_l}-\arg\left(\frac{x_n+\beta_1}{1+\overline{\beta_1}x_n}\right)+\arg\left(\frac{x_l+\beta_1}{1+\overline{\beta_1}x_l}\right)\right).$$

Then,

$$S(\beta_1) = C(\beta_1) + D(\beta_1).$$

Now, let (x_n, y_n) be contaminated to (x_n*, y_n*). Then,

$$S^*(\beta_1) = C(\beta_1) + D^*(\beta_1),$$

where

$$\begin{aligned} D^*(\beta_1) &= \sum_{j=1}^{n-1} \cos\left(\theta_{y_n*} - \theta_{y_j} - \arg\left(\frac{x_n * + \beta_1}{1 + \overline{\beta_1} x_n *}\right) + \arg\left(\frac{x_j + \beta_1}{1 + \overline{\beta_1} x_j}\right)\right) \\ &= \sum_{j=1}^{n-1} \cos\left(\theta_{y_n} - \theta_{y_j} - \arg\left(\frac{x_n * + \beta_1}{1 + \overline{\beta_1} x_n *}\right) + \arg\left(\frac{x_j + \beta_1}{1 + \overline{\beta_1} x_j}\right) + \alpha\right), \end{aligned}$$

where $\alpha \in [0, 2\pi)$. If $BDP > \frac{1}{n}$, then MLE based on $S^*(\beta_1)$ is at finite distance from MLE based on $S(\beta_1)$ for all values of α.

When $\alpha = \pi$, we have $D^*(\beta_1) = -D(\beta_1)$, which implies that the MLE based on $C(\beta_1) + D(\beta_1)$ is at a finite distance from the MLE based on $C(\beta_1) - D(\beta_1)$.

By the continuity of $D(\beta_1)$, the MLE based on $C(\beta_1)$ is at a finite distance from the MLE based on $S(\beta_1)$. Thus, the MLE based on a sample of size n is at a finite distance from the MLE based on a sample of size $n-1$. Proceeding in the same fashion, it can be argued that the MLE based on a sample of size of three is at a finite distance from the MLE based on a sample of size n. But, the MLE based on a sample of size 3 can take any arbitrary value if we change any of the 3 points arbitrarily and the resulting MLE based on the sample of size n will be at a finite distance from the MLE based on a sample of size 3, and hence, it can also take any arbitrary value. Thus, $BDP = \frac{1}{n}$, and thus there is a definite need to address the robustness issue in the case of this circular-circular regression model. □

Corollary. Even if the angular error follows some other distribution, then also by noting that if $BDP > \frac{1}{n}$, x_n can take arbitrary value implies that the MLE based on $(n-1)$ observations is at a finite distance from the MLE based on n observations and continuing as in Theorem 2, it can be argued that there is a contradiction and hence, $BDP = \frac{1}{n}$.

3 Maximum Trimmed Cosine estimator

In the case of linear regression, the least square estimator has $BDP = \frac{1}{n}$. Thus, there was need to improve the robustness of the estimator because a single bad outlier has the potential to change the value of the least square estimate arbitrarily. In terms of the BDP, the problem remained with the introduction of new estimators such as least absolute values estimator proposed by Edgeworth (1987) and M-estimator (see Huber, 1973, p. 800). Thus, a more robust estimator with a higher BDP was needed.

In pursuit of such a robust estimator, Rousseuw (1985) introduced *Least Trimmed Squares* (LTS) estimator which has $BDP = ([n/2]+1)/n$ for simple linear regression, and hence $h = ([n/2]+1)$. Extending the idea in the circular-circular regression setup, in this present paper, we propose the *Maximum Trimmed Cosine Estimator* (MTCE). In the linear regression case, the prediction accuracy of an estimator can be judged by the discrepancy between the observed values and the predicted values. Smaller the values of the squares of the distance between the observed and predicted values, the more is the accuracy. Taking a cue from the linear regression setup, this discrepancy in the circular-circular regression setup can be measured by taking the cosine distance between the observed and predicted angles of the responses. The cosine residual is defined as $\cos(\theta_{y_{j,p}} - \theta_{y_j})$, where $\theta_{y,j}$ is the observed jth angular response and $\theta_{y_{j,p}}$ is the predicted jth angle based on x_j. The higher the value of this cosine residual, the better is the fit.

Thus, the Maximum Trimmed Cosine Residual is defined as the value of the parameters β_0, β_1 which maximizes

$$\sum_{j=1}^{h} \cos(\theta_{y_{j,p}} - \theta_{y_j}),$$

where $\left[\frac{n}{2}\right] + 1 \leq h \leq n$. This estimator is based on h observations out of n. Typically, the value of h is taken to be greater than $\left[\frac{n}{2}\right]$ because a lower value of h means that more than half of the observations are contaminated which does not make much sense.

Theorem 3. In the circular-circular regression setup,

$$BDP \geq \frac{1}{n}\left[\frac{n-1}{2}\right].$$

Proof: The lower limit of BDP of this estimator can be calculated by Theorem 1 of Vandev and Neykov (1998). In the Theorem, they have taken $f_j \geq 0$ for all j, the f_j's are arranged in increasing order and first h out of n f_j's are taken. These conditions can be satisfied by taking the ordering in decreasing sense of the $\left[1 + \cos(\theta_{y_{j,p}} - \theta_{y_j})\right]$ values. This does not change the ordering and make the assumption of Theorem 1 true. Then, the first h functions of out of n functions are taken. To find d such that the function $\sum_{j=1}^{d} f_j$ is d-full, we have assumed that for any set of 3 out of the n functions based on the n observations, we can find a unique maximum. We define $z_j = \frac{y_j(1+\overline{\beta}_1 x_j)}{x_j+\beta_1}$ for all $j = \{1, 2, \ldots, n\}$. Now, $\frac{1}{3}\sum_{j=1}^{3} f_j$ is maximized when $\frac{1}{3}\left|\sum_{j=1}^{3} z_j\right|^2$ is maximized. Now, the set of values of β_1 such that $\frac{1}{3}\left|\sum_{j=1}^{3} z_j\right|^2 \geq C$ for all $C \in (0, 1]$ is compact because of the condition of unique maximum. This expression in terms of z_j also removes β_0 from the expressions which has no bearing on the BDP, as the BDP is calculated in terms of β_1. This assumption is valid unless one of $(x_1, y_1), \ldots, (x_n, y_n)$ coincide. Under the assumption of continuous distributions for $(\theta_{x_j}, \theta_{y_j})$ and independence of each sample point, this assumption follows. Hence, the value of d is equal to 3. Thus, as explained in Theorem 2 of Müller and Neykov (2003), if $\left[\frac{n+3}{2}\right] \leq h \leq \left[\frac{n+4}{2}\right]$, the BDP is not less than $\frac{1}{n}\left[\frac{n-1}{2}\right]$. □

3.1 Maximum trimmed likelihood estimator

Clearly, MTCE is the value of the parameter which fits h observations out of the total n in the best way. This is obtained in the same spirit as the Maximum Trimmed Likelihood Estimator (MTLE) was obtained as the robust answer from the classical maximum likelihood estimator (MLE). See, e.g., Neykov and Neytchev (1990), Bednarski and Clarke (1993), Vandev and Neykov (1993), Cuesta-Albertos *et al.* (2008) for the descriptions, rationale and applicability of the MTLE.

For circular data having von Mises distribution, the likelihood given by (2.3) contains $\sum \cos(\theta_{y_{j,p}} - \theta_{y_j})$ in the exponent, and hence the MTLE will be same as the MTCE. However, if the underlying distribution is something other than von Mises, say Wrapped Cauchy or Wrapped normal, the MTLE is different from the MTCE. The more the underlying distribution deviates from the von Mises, the two estimators are likely to deviate. However, the cosine residual is the most natural way of representing the goodness of fit in case of circular data; and hence in this present paper we concentrate on the MTCE instead of the MTLE.

4 Computing MTCE

Looking at the definition of MTCE, the simplest algorithm for computing MTCE is by taking all $\binom{n}{h}$ combinations and then computing MTCE in every case. But, the complexity of computation is then $O(n^h)$. This complexity can be considerably reduced by using an algorithm mentioned in Klouda (2015). For the calculation of Least Trimmed Square Estimator (LTSE), only h observations are assigned with weight 1 and the rest have weight 0. In Klouda (2015), the whole parameter set is divided into different sets. The weights are assigned based on the ordering of residuals for the parameters at the boundary of two sets. Then, only these weights are included and on the basis of these weighted set of observations, the Trimmed Square Estimator is found for each case. Then, the LTSE is the parameter having the minimum value of least square. The complexity is thus reduced significantly by using the algorithm in the case of linear regression setup. The case of circular-circular regression is different as we do not have such linear equations and hence, we do not get unique solutions by solving appropriate number of linear equations. Thus, in our case, we need to modify their algorithm.

Let the parameters be represented as b_1, b_2, θ_0, where $\beta_1 = (b_1, b_2) \in \mathbb{R}^2$ and $\theta_0 \in [0, 2\pi)$. Now, we shall call this parameter set $\mathcal{B}$. So, $\mathcal{B} = \mathbb{R}^2 \times [0, 2\pi)$. Now, proceeding similar to Klouda (2015), we define two sets

$$
\begin{aligned}
U &= \left\{A \in \mathcal{B} : r_{(h)}(A) > r_{(h+1)}(A)\right\}, \\
H &= \left\{A \in \mathcal{B} : r_{(h)}(A) = r_{(h+1)}(A)\right\},
\end{aligned}
$$

where $r_k(A)$ is the ordered kth residual. The ordering is done such that $r_{(1)}(A) \geq r_{(2)}(A) \geq \cdots \geq r_{(n)}(A)$. Now, $\sum_{j=1}^{h} r_{(j)} = \sum_{j=1}^{n} w_j r_j$, where $w_j \in \{0, 1\}$ and $\sum_{j=1}^{n} w_j = h$. Let $Z(A) = w_A$, where w_A denote the set $\{w_1, \ldots, w_n\}$ with respect to A. Let $A_1, A_2 \in U$ and $tA_1 + (1-t)A_2 \in U$ for all $t \in (0, 1)$. Then, $Z(A_1) = Z(A_2)$. Thus, H divides U in connected sets U_js such that $\cup_j U_j = U$ and $A_1, A_2 \in U_j \Rightarrow Z(A_1) = Z(A_2)$. The set H contains all the weights. Thus, we shall only find the weights on set H and then find MTCE corresponding to each weighted set of observations and then choose the parameter with respect to which the sum of ordered h cosine residuals is maximum. But, when only two observations are used, then the number of points in the set H can be infinite. Thus, we have to find a set

$H_c \subset H$ which contains finite points and we shall consider only the weights corresponding to this set.

Now, we consider

$$\cos(\theta_{y_{l,p}} - \theta_{y_l}) = \cos(\theta_{y_{j,p}} - \theta_{y_j}) = \cos(\theta_{y_{k,p}} - \theta_{y_k}).$$

Note that $\cos(\theta_{y_{l,p}} - \theta_{y_l}) = \cos(\theta_{y_{j,p}} - \theta_{y_j})$ implies

$$Re\{z_l(\beta_0, \beta_1)\} = Re\{z_j(\beta_0, \beta_1)\}, \tag{4.1}$$

where $z_s(\beta_0, \beta_1) = y_s \frac{1+\overline{\beta_1} x_s}{\beta_0(x_s+\beta_1)}$ for $s = l, j$. Thus, either $z_l(\beta_0, \beta_1) = z_j(\beta_0, \beta_1)$ or $z_l(\beta_0, \beta_1) = \overline{z_j(\beta_0, \beta_1)}$. If $z_l(\beta_0^*, \beta_1^*) = z_j(\beta_0^*, \beta_1^*)$, then $z_l(\beta_0, \beta_1^*) = z_j(\beta_0, \beta_1^*)$ for all $\theta_0 \in [0, 2\pi)$. Now, $z_l = z_j$ implies that

$$\beta_1\overline{\beta_1} + \beta_1 \frac{y_l - y_j}{x_l y_l - x_j y_j} + \overline{\beta_1} \frac{x_l x_j (y_l - y_j)}{x_l y_l - x_j y_j} + \frac{x_j y_l - x_l y_j}{x_l y_l - x_j y_j} = 0. \tag{4.2}$$

Similarly, from $z_j = z_k$, we get

$$\beta_1\overline{\beta_1} + \beta_1 \frac{y_j - y_k}{x_j y_j - x_k y_k} + \overline{\beta_1} \frac{x_j x_k (y_j - y_k)}{x_j y_j - x_k y_k} + \frac{x_k y_j - x_j y_k}{x_j y_j - x_k y_k} = 0. \tag{4.3}$$

Denoting the left hand side of (4.2) as L_1 and the left hand side of (4.3) as L_2, the solution to (4.2) and (4.3) should satisfy $L_1 = L_2 = 0$. Now, $Re(L_1) = Re(L_2)$, $Im(L_1) = Im(L_2)$ and $Im(L_1) = 0$ gives a set of three linear equations in two unknowns, corresponding to the real and imaginary parts of β_1 which, under the assumption of continuity of x and y, has a solution with probability 0.

Thus, the probability of having a solution for $z_l = z_j = z_k$ is 0. Similarly, corresponding to four observations, the probability of $z_l = z_j = \overline{z_k} = \overline{z_r}$ is 0 as in this case there should exist a β_1 which is at the intersection of three lines, this again gives two unknowns in three linear equations having probability 0 under the assumption of continuity of x and y . Also, the solutions of $z_l = z_j = \overline{z_k}$ are continuous.

Thus, we find solutions for $z_l = z_j = \overline{z_k} = C$ for any $C \in \Omega$, and then we get the weight sets. Fixing the value of $C = 1$, first we solve $z_l = z_j = 1$ and find a unique solution for β_1. We get a unique solution with probability 1 because $z_l = z_j = 1$ gives two lines L_1' (which joins $-x_l$ and y_l) and L_2' (which joins $-x_j$ and y_j) and the solution β_1 will be the intersection point of these two lines. Now, for this particular value of β_1, we shall take all the other observations and find β_0 corresponding to each of the triplets $z_l = z_j = \overline{z_k}$. For each of these triplets, we shall find the two values of β_0 which give $\frac{z_l}{\beta_0} = \frac{\overline{z_k}}{\beta_0}$. Then, we order the residuals and take the set corresponding to the first h ordered indices. Next, based on these set of observations, we find the MLE of β_0, β_1. Then, among all these set of parameter values, the MTCE will be the one for which the sum of cosines is maximum.

Thus, we shall take all $\binom{n}{2}$ pairs of observations.

$$z_l = z_j = 1 \tag{4.4}$$

Taking two observations with indices l, j, we get a unique solution for $z_l = z_j = 1$, and then the equality of $z_l = z_j$ holds at this value of β_1 for all values of $\theta_0 \in [0, 2\pi)$. We shall thus take a third observation (with index k, say) and at this value of β_1, we solve for

$$z_l = z_j = \overline{z_k}. \tag{4.5}$$

Solving (4.5) for this particular value of β_1, we get two different values for β_0 for each triplet (l, j, k). Now, for each of these values of β_0, β_1, we order the observations and get the weight set. If $r_l(\beta_0, \beta_1) = r_{(h)}(\beta_0, \beta_1)$, then we choose this set of weights. We take the set of weights corresponding to each of these β_1 and β_0. Then, the maximum of subsets of set s which are needed will be $2(n-2)\binom{n}{2}$. Then, we find the subset for which the MLTC residual is maximum. Note that, by using the conjugate of z_k we have also considered the cases when $z_l = \overline{z_k}$.

Formally, the algorithm for this computation is mentioned below. Denoting any set of three observations by ν_s where, $1 \leq s \leq \binom{n}{2}$, we can write the algorithm as follows.

Algorithm 1:

Set $s = 1$, $(\theta_0, \beta_1) = (0, 1)$ and $MTCE_{\max} = -\infty$.

Step 1: Solve (4.4) for ν_s and find the β_1.

Step 2: For this β_1, take one new observation and find the corresponding θ_0 by taking $\frac{z_l}{\beta_0} = \frac{z_j}{\beta_0} = \frac{\overline{z_k}}{\beta_0}$.

Step 3: Hence, find the set of weights corresponding to this parameter for which jth observation can be the hth residual. The maximum number of such weights can be $2^3 - 1 = 7$.

Step 4: $MTCE_{\nu_s}$ = maximum of the residuals.

Step 5: if $MTCE_{\nu_s} > MTCE_{\max}$, then $MTCE_{\max} = MTCE_{\nu_s}$.

Step 6: if, $s < \binom{n}{2}$, $s = s + 1$ else, End.

The maximum number of weights mentioned in Step 3 is 7 as three of the observations have the same residual. Thus, if only one of them can be taken in the first h residuals, then there are 3 possible set of weights. But when 2 of them can be taken, then there are again 3 possible set of weights. When 3 of them can be taken, then there is only one set of weights. Thus, the complexity of the new algorithm comes out to be $7(n-2)\binom{n}{2}$, which is of order $O(n^3)$, and hence is significantly less than $\binom{n}{h}$ for large values of n.

5 Simulation

In the simulation study, we first take a sample size of 20, which is close to the sample size of the data analysis example. We take two different concentration parameters for data generation when the error distribution is von Mises, and two different values of the parameter ρ when the error distribution is Wrapped Cauchy. Two cases are considered: 15% (i.e. 3) outliers and 30% (i.e. 6) outliers. The outliers are created in a way that when the proportion of outliers is p, the remaining $1-p$ proportion of observations are taken from the true regression model with $\beta_1 = 0.9, \beta_0 = 1$ while the rest of the responses are taken in the von Mises setup from the regression model with parameters as $\beta_0 = 1, \beta_1 = -0.9$ and $\kappa = 5$; and in the Wrapped Cauchy case the true regression model parameters are $\beta_0 = 1, \beta_1 = 0.9$ and the rest of the responses are taken with model parameters $\beta_0 = 1, \beta_1 = -0.9$ and $\rho = 0.9$. We then carry out the same computation with a sample size of 40 with 15% (i.e. 6) and 30% (i.e. 12) outliers, respectively. In all the cases, h is taken as $\left[\frac{n+4}{2}\right]$. The covariates in all the simulations are generated from circular uniform distribution.

Tables 1 and 2 show the mean values (and standard errors in parentheses) of the estimates of b_1, b_2 where $\beta_1 = b_1 + ib_2$, based on 1000 simulations. Also, the circular mean and circular dispersion is reported for the parameter θ_0, which represents the argument of β_0 such that $\beta_0 = e^{i\theta_0}$. The mean values and the standard errors of the number of outliers involved in estimating the regression parameters is also reported in the tables. Tables 1 and 2 also present the Maximum Likelihood Estimates of b_1, b_2, θ_0 from the standard circular-circular regression model based on 1000 simulations each of sample sizes 20 and 40. It can be seen from the tables that the MTCE estimates are closer in all the the cases to the true parameter than the MLE.

In Tables 3 and 4, we present the computational results for $\beta_1 = -0.6 + 0.3i$ and $\beta_0 = i$, the values close to the estimated values in data analysis. We again consider $n = 20, 40$, and the same κ and ρ-values as earlier for the von Mises and Wrapped Cauchy distributions, respectively. The contamination is done with parameters $\beta_0 = i, \beta_1 = 0.84 - 0.42i$ with the error distribution same as in 1 and 2. We see that our proposed MTCE outperform the MLE's.

To check the usefulness of our method in rejecting outliers in the calculation of the MTCE, let v be the number of outliers in the sample, and G be the number of outliers involved in the prediction of the estimates of the regression coefficients. Then, under the null hypothesis of no detection of outliers, G follows a hypergeometric distribution with mean $M_G = \frac{hv}{n}$ and variance $V_G = \frac{hv(n-v)(n-h)}{n^2(n-1)}$. As the number of replications (N) goes to infinity,

$$\sqrt{N}(\bar{G} - M_G) \sim N(0, V_G).$$

The mean values (and standard errors) of M_G are also reported in the Tables. The null hypothesis is rejected in all the simulations with $v = 3, 6$ at 0.001 level.

6 Data Analysis

In the Example 2 of Kato *et al.* (2008), circular-circular regression is carried out for the wind direction data, where the covariate is taken to be the wind direction at 6 a.m. and the response is taken as the wind direction at 12 noon for 21 consecutive days at the Milwaukee weather station. If the regression link function is used with von Mises error distribution, then it can be seen from Figure 2 that only two of the points, 5 and 17, can be considered as outliers. But, during the computation of the MLE, these points were also incorporated.

In this present paper, instead of MLE, we apply the MTCE discussed in the paper for this dataset. The MTCE for $\beta_1 = -0.528 + 0.278i$ and the MTCE for $\beta_0 = 0.309 + 0.951i$. On the basis of this, we plot the *circular distance*, given by $1 - \cos(\theta_{y_{j,p}} - \theta_{y_j})$, which is *one minus cosine residual.* This plot shows that the points $5, 7, 12, 17, 20$ are the outliers. Also,

Table 1: Different types of estimates of the parameters (with standard errors in parentheses) when angular error follows von Mises distribution with parameter κ. Here $b_1 = 0.9$, $b_2 = 0$, $\theta_0 = 0$.

				Estimated Values (s.e.) of			
n	κ	v	Estimator	b_1	b_2	θ	M_G
20	5	3	MTCE	0.776 (0.011)	0.000 (0.015)	0.001 (0.143)	0.363 (0.017)
			MLE	0.335 (0.022)	0.042 (0.019)	-0.098 (0.587)	
	5	6	MTCE	0.767 (0.009)	0.025 (0.015)	-0.017 (0.142)	0.699 (0.025)
			MLE	0.218 (0.029)	-0.026 (0.023)	-0.002 (0.650))	
	4	3	MTCE	0.779 (0.009)	-0.019 (0.014)	0.033 (0.136)	0.357 (0.017)
			MLE	0.290 (0.022)	0.018 (0.020)	-0.032 (0.629)	
	4	6	MTCE	0.670 (0.021)	-0.016 (0.027)	0.012 (0.212)	0.831 (0.028)
			MLE	0.201 (0.025)	0.002 (0.025)	-0.042 (0.653)	
40	5	6	MTCE	0.846 (0.004)	0.008 (0.011)	-0.006 (0.012)	0.356 (0.018)
			MLE	0.465 (0.020)	0.003 (0.016)	0.007 (0.431)	
	5	12	MTCE	0.834 (0.006)	0.007 (0.012)	-0.014 (0.087)	0.620 (0.027)
			MLE	0.392 (0.021)	0.004 (0.018)	-0.005 (0.458)	
	4	6	MTCE	0.837 (0.006)	-0.016 (0.012)	0.014 (0.087)	0.391 (0.020)
			MLE	0.463 (0.024)	-0.024 (0.017)	0.050 (0.455)	
	4	12	MTCE	0.817 (0.006)	-0.005 (0.012)	0.012 (0.099)	0.760 (0.031)
			MLE	0.366 (0.021)	0.009 (0.017)	-0.018 (0.459)	

Table 2: Different types of estimates of the parameters (with standard errors in parentheses) when angular error follows Wrapped Cauchy distribution with parameter ρ. Here $b_1 = 0.9$, $b_2 = 0$, $\theta_0 = 0$.

				Estimated Values (s.e.) of			
n	ρ	v	Estimators	b_1	b_2	θ	M_G
20	0.65	3	MTCE	0.704 (0.019)	0.006 (0.041)	-0.014 (0.228)	0.461 (0.021)
			MLE	0.655 (0.017)	-0.032 (0.021)	0.005 (0.245)	
	0.65	6	MTCE	0.454 (0.034)	-0.011 (0.037)	-0.015 (0.443)	1.811 (0.053)
			MLE	0.098 (0.034)	0.017 (0.033)	-0.070 (0.795)	
	0.9	3	MTCE	0.854 (0.005)	-0.005 (0.008)	0.004 (0.047)	0.199 (0.013)
			MLE	0.373 (0.022)	0.002 (0.018)	0.433 (0.560)	
	0.9	6	MTCE	0.823 (0.007)	-0.000 (0.010)	-0.003 (0.074)	0.498 (0.020)
			MLE	0.341 (0.023)	-0.015 (0.020)	0.039 (0.591)	
40	0.65	6	MTCE	0.821 (0.009)	-0.002 (0.013)	0.001 (0.112)	0.484 (0.024)
			MLE	0.285 (0.023)	-0.007 (0.020)	-0.031 (0.020)	
	0.65	12	MTCE	0.802 (0.022)	0.013 (0.029)	-0.018 (0.232)	1.719 (0.071)
			MLE	0.145 (0.034)	0.008 (0.033)	-0.123 (0.747)	
	0.9	6	MTCE	0.888 (0.001)	0.001 (0.005)	-0.002 (0.014)	0.187 (0.013)
			MLE	0.436 (0.022)	0.000 (0.016)	-0.001 (0.495)	
	0.9	12	MTCE	0.879 (0.002)	-0.009 (0.006)	0.009 (0.023)	0.490 (0.022)
			MLE	0.433 (0.022)	-0.100 (0.016)	0.016 (0.491)	

Table 3: Different types of estimates of the parameters (with standard errors in parentheses) when angular error follows von Mises distribution with parameter κ. Here $b_1 = -0.6$, $b_2 = 0.3$, $\theta_0 = \pi/2$.

n	κ	v	Estimator	Estimated Values (s.e.) of			
				b_1	b_2	θ	M_G
20	5	3	MTCE	-0.558 (0.008)	0.246 (0.011)	1.530 (0.079)	0.345 (0.018)
			MLE	-0.438 (0.014)	0.023 (0.012)	1.583 (0.140)	
	5	6	MTCE	-0.535 (0.009)	0.282 (0.010)	1.542 (0.067)	0.764 (0.032)
			MLE	-0.242 (0.036)	0.263 (0.065)	1.572 (0.263)	
	4	3	MTCE	-0.574 (0.007)	0.263 (0.010)	1.536 (0.069)	0.277 (0.016)
			MLE	-0.432 (0.016)	0.213 (0.012)	1.569 (0.158)	
	4	6	MTCE	0.535 (0.014)	0.268 (0.013)	1.543 (0.082)	0.861 (0.033)
			MLE	-0.204 (0.039)	0.143 (0.033)	1.568 (0.320)	
40	5	6	MTCE	-0.604 (0.006)	0.241 (0.008)	1.511 (0.044)	0.387 (0.021)
			MLE	-0.516 (0.006)	0.263 (0.005)	1.573 (0.039)	
	5	12	MTCE	-0.580 (0.004)	0.282 (0.007)	1.543 (0.026)	0.900 (0.034)
			MLE	-0.345 (0.027)	0.191 (0.025)	1.563 (0.152)	
	4	6	MTCE	-0.604 (0.006)	0.241 (0.008)	1.501 (0.044)	0.387 (0.021)
			MLE	-0.516 (0.006)	0.263 (0.005)	1.573 (0.039)	
	4	12	MTCE	-0.573 (0.013)	0.262 (0.007)	1.523 (0.030)	1.053 (0.038)
			MLE	-0.353 (0.025)	0.138 (0.020)	1.564 (0.144)	

Table 4: Different types of estimates of the parameters (with standard errors in parentheses) when angular error follows Wrapped Cauchy distribution with parameter ρ. Here $b_1 = -0.6$, $b_2 = 0.3$, $\theta_0 = \pi/2$.

n	ρ	v	Estimators	Estimated Values (s.e.) of			
				b_1	b_2	θ	M_G
20	0.65	3	MTCE	-0.513 (0.015)	0.224 (0.017)	1.535 (0.122)	0.500 (0.022)
			MLE	-0.413 (0.018)	0.202 (0.016)	1.565 (0.269)	
	0.65	6	MTCE	-0.402 (0.029)	0.244 (0.031)	1.536 (0.290)	1.969 (0.059)
			MLE	-0.405 (0.017)	0.197 (0.017)	1.561 (0.280)	
	0.9	3	MTCE	-0.599 (0.003)	0.279 (0.005)	1.560 (0.021)	0.156 (0.012)
			MLE	-0.430 (0.014)	0.248 (0.012)	1.607 (0.225)	
	0.9	6	MTCE	-0.575 (0.005)	0.285(0.006)	1.563 (0.031)	0.424 (0.025)
			MLE	-0.426 (0.014)	0.236 (0.013)	1.592 (0.239)	
40	0.65	6	MTCE	-0.593 (0.007)	0.245 (0.009)	1.499 (0.043)	0.591 (0.025)
			MLE	-0.367 (0.017)	0.186 (0.017)	1.575 (0.288)	
	0.65	12	MTCE	-0.495 (0.016)	0.245 (0.016)	1.554 (0.094)	2.300 (0.085)
			MLE	-0.208 (0.031)	0.146 (0.031)	1.565 (0.398)	
	0.9	6	MTCE	-0.618 (0.002)	0.269 (0.005)	1.556 (0.017)	0.146 (0.013)
			MLE	-0.436 (0.014)	0.232 (0.013)	1.586 (0.219)	
	0.9	12	MTCE	-0.608 (0.002)	0.284 (0.004)	1.556 (0.010)	0.465 (0.022)
			MLE	-0.407 (0.019)	0.209 (0.019)	1.587 (0.225)	

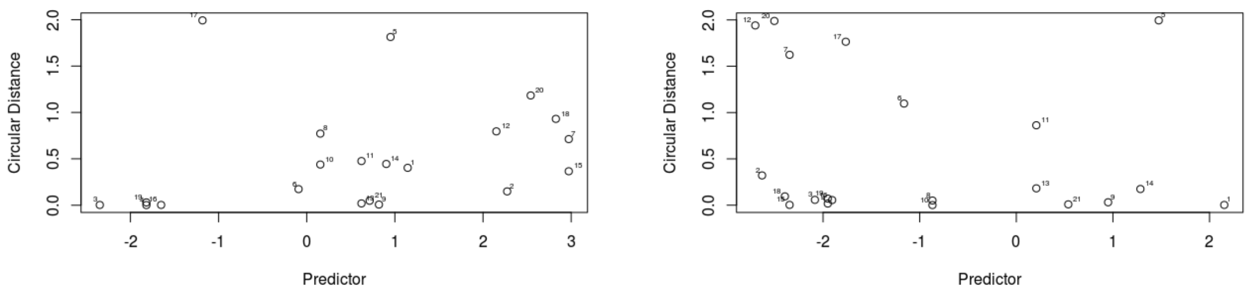

Figure 2: Circular Distance versus Predictors. (Left): Plot based on circular-circular regression model and (Right:) Plot based on robust circular-circular regression model.

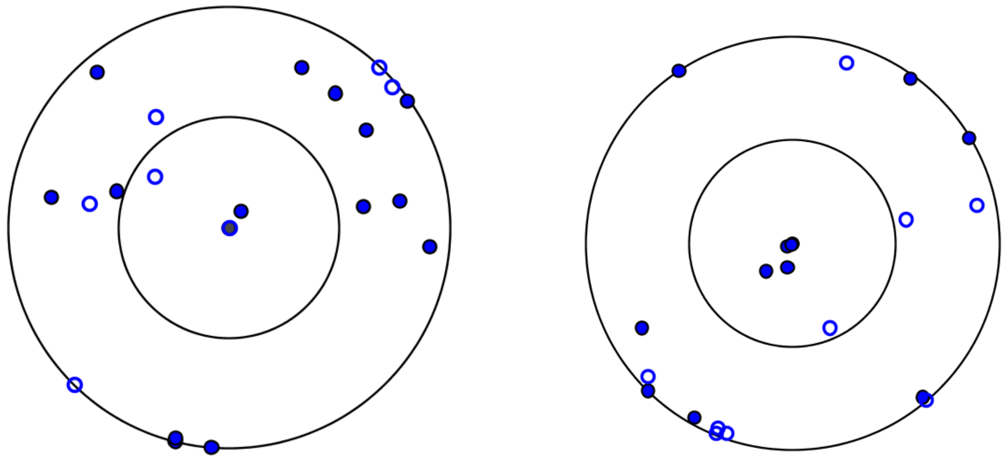

Figure 3: Donutplots. (Left:) Donutplot for the predicted values with circular-circular regression model and responses; (Right:) Donutplot with predicted values based on robust circular-circular regression model and responses.

the MTCE is robust against outliers, and hence the other points are fitted much better than as shown in the case of using circular-circular regression model. The circular distance versus predictors (predicted mean direction) for both the cases is shown in Figure 2.

In Figure 3, we present the donut plots, introduced by Jha and Biswas (2016), showing the complex numbers $(1 + cos(\theta_{y,e} - \theta_y))y_e$ where y_e is the complex number corresponding to the predicted points and y is the complex number corresponding to the observed point. Thus, for a good fit, the observations are on the circumference of the outer circle while a bad fit results in observations near the origin. It is clear that the proposed robust method works better than the standard non-robust fitting in terms of the fit of the data.

7 Concluding Remarks

Although presence of outliers has been observed in circular data (Otieno and Anderson-Cook, 2006), even in the regression setup (Kato *et al.*, 2008), the present paper is perhaps the first attempt to objectively discuss the robustness in terms of breakdown point and hence, the need to address the problem in case of circular random variables. The robustness is discussed for the circular-circular regression using the useful Möbius model. To overcome the problem, then we provided the cosine distance based Maximum Trimmed Cosine Estimator in place of the standard MLE or the Maximum Trimmed Likelihood Estimator. The computational algorithm provided to find the MTCE is also different from the usual linear regression setup as the geometry of the circular-circular regression is used to find the most appropriate weighted set. The MTCE works nicely to identify the outliers and also to provide robust fit of the data.

References

Agostinelli, C. (2007). Robust estimation for circular data. *Computational Statistics and Data Analysis* 51, 5867-5875.

Bednarski, T. and Clarke, B. (1993). Trimmed likelihood estimation of location and scale of the normal distribution. *Australian Journal of Statistics* 35, 141-153.

Busse, P. and Trociñska A. (1999). Evaluation of orientation experiment data using circular statistic doubts and pitfalls in assumptions. *Ring* 21, 107-130.

Chang, T., Stock, J. and Molnar, P. (1990). The rotation group in plate tectonics and the representation of uncertainties of plate reconstructions. *Geophysics Journal International* 101,649-661.

Cuesta-Albertos, J.A., Matrán, C. and Mayo-Iscar, A. (2008). Trimming and likelihood: robust location and dispersion estimation in the elliptical model. *The Annals of Statistics* 36, 2284-2318.

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann.* (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

Fisher, N.I. (1982). Robust estimation of the concentration parameter of Fisher's distribution on the sphere. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 31, 152-154.

Ibrahim, S., Rambli, A, Hussin, A.G. and Mohamed, I. (2013). Outlier detection in a circular regression model using COVRATIO statistic. *Communications in Statistics-Simulation and Computation* 42 (10), 2272-2280

Jha, J. and Biswas, A. (2016). Multiple circular-circular regression. *Statistical Modelling.* In Press.

Kato, S. and Eguchi, S. (2014). Robust estimation of location and concentration parameters for the von Mises-Fisher distribution. *Statistical Papers*, 1-30.

Kato, S., Shimizu, K. and Shieh, G.S. (2008). A circular-circular regression model. *Statistica Sinica* 18, 633-645.

Klouda, K. (2015). An exact polynomial time algorithm for computing the least trimmed squares. *Computational Statistics and Data Analysis* 84, 27-40.

Ko, D. (1992). Robust estimation of the concentration parameter of the von Mises-Fisher distribution. *The Annals of Statistics* 20, 917-928.

Laha, A. and Mahesh, K. (2012). Sb-robust estimator for the concentration parameter of circular normal distribution. *Statistical Papers* 53, 457-467.

Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics.* John Wiley and Sons, Chichester.

Müller, C.H. and Neykov, N. (2003). Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *Journal of Statistical Planning and Inference* 116, 503-519.

Neykov, N.M. and Neytchev, P. (1990). A robust alternative of the maximum likelihood estimators. *COMPSTAT90 – Short Communications, Dubrovnik, Yugoslavia*, pp. 99-100.

Otieno, B.S. and Anderson-Cook, C.M. (2006). Measures of preferred direction for environmental and ecological circular data. *Environmental and Ecological Statistics* 13, 311-324.

Protheroe, R. J. (1985). A new statistic for the analysis of circular data in gamma-ray astronomy. *NASA. Goddard Space Flight Center 19th Intern. Cosmic Ray Conf.*, Vol. 3, p 485-488.

Rousseeuw, P.J. (1984). Least median of squares regression. *Journal of American Statistical Association* 79, 871-880.

Rousseeuw, P.J. (1985). *Multivariate estimation with high breakdown point.* Mathematical Statistics and Applications, Vol. B (Bad Tatzmannsdorf, 1983), 283-297, Reidel, Dordrecht.

Vandev, D. and Neykov, N. (1993). *New Directions in Statistical Data Analysis and Robustness*, chapter *Robust maximum likelihood in the Gaussian case*, pages 257–264. Birkhauser Verlag.

Vandev, D.L. and Neykov, N.M. (1998). About regression estimators with high breakdown point. *Statistics* 32, 111-129.

Zubairi, Y.Z., Hussain, F. and Hussain, A.G. (2008). An alternative analysis of two circular variables via graphical representation: an application to the Malaysian wind data. *Computer and Information Science* 1, 3-8.