International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 11 (2018) pp. 9083-9090

Outliers Detection in Multiple Circular Regression Model via *DFBETAc* Statistic

Najla Ahmed Alkasadi
Student, Institute of Engineering Mathematics, Universiti Malaysia Perlis,
Pauh Putra Main Campus, 02600 Arau, Perlis, Malaysia.
Orcid Id: 0000-0003-4473-392X

Ali H. M. Abuzaid
Associate Professor, Department of Mathematics,
Faculty of Science, Al-Azhar University-Gaza, Palestine.
Orcid Id: 0000-0002- 6680-7371

***Safwati Ibrahim**
Senior Lecturer, Institute of Engineering Mathematics,
Universiti Malaysia Perlis, Pauh Putra Main Campus, 02600 Arau, Perlis, Malaysia.
Orcid Id: 0000-0001-9009-5880

Mohd Irwan Yusoff
Senior Lecturer, Center for Diploma Studies, S2-L1-26,
Kampus Uniciti Sungai Chuchuh, Universiti Malaysia Perlis 021000 Padang Besar (U), Perlis, Malaysia.
Orcid Id: 0000-0003-2250-0674

Abstract

In regression analysis, an outlier is an observation for which the residual is large in magnitude compared to other observations in the data set. The investigation on the identification of outliers in linear regression models can be extended to those for circular regression case. In this paper, we study the relationship between more than two circular variables using the multiple circular regression model, which is proposed by [13]. The model has precise enough and interesting properties to detect the occurrence of outliers. Here, we concentrate the attention on the problem of identifying outliers in this model. In particular, the extension of *DFBETAS* statistic which has been successfully used for the same purpose to this model via the row deleted approach. The cut-off points and the power of performance of the procedure are investigated via Monte Carlo simulation. It is found that the performance improves when the resulting residuals have small variance and when the sample size gets larger. The real data is applied for illustration purpose.

Keywords: circular regression model, DFBETA, outlier.

INTRODUCTION

One of the common problems in circular regression modelling is the existence of some unexpected observations which is called outliers, such observation affects the statistical inference. Thus, it is important to detect and assess these observations and estimate its impact on the proposed model. The existence of outliers in any dataset distorting the coefficients estimates in the regression [1].

For multiple linear regression given by $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y}$ is dependent variable, $\boldsymbol{X}$ is the explanatory variable and $\boldsymbol{\beta}$ is the slope of the line. The *DFBETAS* statistic was introduced by [11], where it is a measure of how much an observation has affected the estimate of a regression coefficient.

The *DFBETAS* statistic is given by

$$DFBETAS_{j.i} = \frac{\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{j(i)}}{\sqrt{s_i^2 \, c_{jj}}} \tag{1}$$

where $S_{(i)}$ is the standard error which is estimated without the *i*th observation, c_{jj} is the *j*th diagonal element of $(\boldsymbol{X'X})^{-1}$ and $\hat{\boldsymbol{\beta}}_{j(i)}$ is the *j*th regression coefficient which is obtained without the *i*th observation. Any observation with $\left|DFBETAS_{j.i}\right| > \frac{2}{\sqrt{n}}$ will be identified as an outlier [2, 3].

Not much study has been done on the problem of outliers and influential points in multiple circular regression models, but there is a set of hypothesis testing and graphical plots have been proposed to identify outliers in the simple circular regression model. Recently, the detection of the outliers in the linear and circular case received a great interest see [4, 5, 6, 7, 8, 9, 10, 11, 13].

Here, we extend this method for the multiple circular regression models. The interest here is how the deletion of any row affects the estimated coefficients since the *DFBETAS* statistic indicates how much the regression coefficient $\hat{\boldsymbol{\beta}}_j$ changes if the *i*th observation was deleted. It is defined as the difference between the regression coefficients calculated for all of the data and the regression coefficient calculated with the deleted observation, scaled by the standard error calculated with the observation deleted. Since there is no literature found on effect of outliers on coefficients of multiple circular regression models, thus this paper extends *DFBETAS* statistic to the circular regression case.

The rest of paper is organized as follows, Section 2 reviewed the multiple circular regression model, Section 3 propose the *DFBETAS* statistic to circular regression model and obtain its cut-off point and examines its performance, Section 4 discusses the analysis.

THE MULTIPLE CIRCULAR REGRESSION MODEL (MCRM)

The Multiple Circular Regression Model (MCRM) studies the relationship between a circular dependent and a set of independent circular variables, the model is proposed by [13]. Here, the MCRM only focuses for two independent circular random variables; $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ and dependent circular variable $\boldsymbol{V}$, in terms of the conditional expectation, e^{iv} given (u_1, u_2) as;

$$E\left(e^{iv} \mid u_1, u_2\right) = \rho(u_1, u_2)e^{i\mu(u_1,u_2)} = g_1(u_1, u_2) + ig_1(u_1, u_2) \qquad (2)$$

Then, the parameters $\mu(u_1, u_2)$ and $\rho(u_1, u_2)$ may be estimate such that

$$\mu(u_1,u_2) = \hat{v} = \begin{cases} \arctan \dfrac{g_2(u_1,u_2)}{g_1(u_1,u_2)} & \text{if} \quad g_1(u_1,u_2) \geq 0 \\ \pi + \arctan \dfrac{g_2(u_1,u_2)}{g_1(u_1,u_2)} & \text{if} \quad g_1(u_1,u_2) \leq 0 \\ \text{undefined} & \text{if} \quad g_1(u_1,u_2) = g_2(u_1,u_2) = 0, \end{cases} \qquad (3)$$

where $\mu(u_1, u_2)$ is the conditional mean direction of v given u_1 and u_2, and $\rho(u_1, u_2)$ is the conditional concentration towards $\mu(u_1, u_2)$.

Consequently, the values of $g_1(u_1, u_2)$ and $g_2(u_1, u_2)$ can be estimated using the following trigonometric polynomials of a suitable degree (m) as follows

$$g_1(u_1,u_2) \approx \sum_{k,\ell=0}^{m} \gamma_{k\ell} \begin{bmatrix} A_{k\ell}\cos ku_1 \cos \ell u_2 + B_{k\ell}\cos ku_1 \sin \ell u_2 \\ + C_{k\ell}\sin ku_1 \cos \ell u_2 + D_{k\ell}\sin ku_1 \sin \ell u_2 \end{bmatrix}$$

$$g_2(u_1,u_2) \approx \sum_{k,\ell=0}^{m} \gamma_{k\ell} \begin{bmatrix} E_{k\ell}\cos ku_1 \cos \ell u_2 + F_{k\ell}\cos ku_1 \sin \ell u_2 \\ + G_{k\ell}\sin ku_1 \cos \ell u_2 + H_{k\ell}\sin ku_1 \sin \ell u_2 \end{bmatrix} \qquad (4)$$

where $\gamma_{kl} = \begin{cases} \frac{1}{4} & \text{for} \quad k = l = 0 \\ \frac{1}{2} & \text{for} \quad k > 0,\ l = 0 \quad \text{and} \quad \text{for} \quad k = 0,\ l > 0. \\ 1 & \text{for} \quad k > 0, l > 0 \end{cases}$

Hence, according to equation (4), there are two observational-models as follows

$$\boldsymbol{V}_{1j} = \cos v_j = \sum_{k,l=0}^{m} \begin{pmatrix} A_{k\ell}\cos ku_1 \cos \ell u_2 + B_{k\ell}\cos ku_1 \sin \ell u_2 \\ + C_{k\ell}\sin ku_1 \cos \ell u_2 + D_{k\ell}\sin ku_1 \sin \ell u_2 \end{pmatrix} + \varepsilon_{1j}$$

$$\boldsymbol{V}_{2j} = \sin v_j = \sum_{k,l=0}^{m} \begin{pmatrix} E_{k\ell}\cos ku_1 \cos \ell u_2 + F_{k\ell}\cos ku_1 \sin \ell u_2 \\ + G_{k\ell}\sin ku_1 \cos \ell u_2 + H_{k\ell}\sin ku_1 \sin \ell u_2 \end{pmatrix} + \varepsilon_{2j} \qquad (5)$$

for j = 1, … , n and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$ is the random error vector following normal distribution with mean $\boldsymbol{0}$ and unknown dispersion matrix $\boldsymbol{\Sigma}$. The parameters are estimated by the least squares method for m=1 and p=2. In order to ensure identifiability, it was assumed that A_0 and E_0 are set to be zero.

Subsequently, $\boldsymbol{V}^{(1)}$ and $\boldsymbol{V}^{(2)}$ were written in the matrix form as

$$\boldsymbol{V}^{(1)} = \boldsymbol{U}\boldsymbol{\lambda}^{(1)} + \boldsymbol{\varepsilon}^{(1)} \qquad (6)$$
$$\boldsymbol{V}^{(2)} = \boldsymbol{U}\boldsymbol{\lambda}^{(2)} + \boldsymbol{\varepsilon}^{(2)}$$

Thus, the least squares estimation turns out to be given by

$$\hat{\boldsymbol{\lambda}}^{(1)} = (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{V}^{(1)} \qquad (7)$$
$$\hat{\boldsymbol{\lambda}}^{(2)} = (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{V}^{(2)}$$

where $\boldsymbol{U}$ is the matrix of the combination of cosine and sine function. The covariance matrix of the residuals, $\boldsymbol{\Sigma}$ is estimated as follow

$$\hat{\boldsymbol{\Sigma}} = \left[n - 2\left(2^{p} m + 1\right)\right]^{-1} R_0$$

where $R_0(p,q) = \boldsymbol{V}^{(p)\prime}\boldsymbol{V}^{(q)} - \boldsymbol{V}^{(p)\prime}\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{V}^{(q)}$,

$\underset{n\times(4m+1)}{\boldsymbol{U}} = \left[\underset{n\times 1}{1} \quad \underset{n\times m}{\boldsymbol{CC}} \quad \underset{n\times m}{\boldsymbol{CS}} \quad \underset{n\times m}{\boldsymbol{SC}} \quad \underset{n\times m}{\boldsymbol{SS}}\right]$ and $R_0 = \left(R_0(p,q)\right)_{p,q=1,2}$.

DFBETAc$_{j,i}$ of Multiple Circular Regression Model

The extension of *DFBETAS* statistic in equation (1) to the circular case is formatted as follows

$$DFBETAc_{j,i} = \frac{\hat{\boldsymbol{\lambda}}_j - \hat{\boldsymbol{\lambda}}_{j(i)}}{\sqrt{s^2_{(i)} C_{jj}}} \quad (8)$$

where $\hat{\boldsymbol{\lambda}}_j$ is the estimating parameter of the full data and $\hat{\boldsymbol{\lambda}}_{j(i)}$ is the corresponding estimating parameter after the *i*th observation is deleted, $S_{(i)}$ the estimated standard error without the *i*th observation is deleted, and C_{jj} is the *j*th diagonal element of $(\boldsymbol{U}'\boldsymbol{U})^{-1}$.

(i) Cut-off Points of DFBETAc$_{j,i}$ Statistic

The simulation study is carried out to obtain the cut-off points of the test statistic for each combination of sample sizes *n* and standard deviation (σ_1, σ_2). Specifically, a sets of random errors from the bivariate Normal distribution with mean vector $\boldsymbol{0}$ for various combination of (σ_1, σ_2) in the range of [0.05, 0.3] and *n* in the range [20, 150]. The complete steps to obtain the cut-off points are described below:

Step 1. Generate a variable $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ of size *n* from *VM* (π , 3) and *VM* (π , 2) respectively.

Step 2. Generate $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ of size *n* from $\boldsymbol{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma_1 & 0\\0 & \sigma_2\end{pmatrix}\right)$. For a fixed *a*=2, obtain the true values of $\lambda = A_0, A_1, B_1, C_1, D_1, E_0, E_1, F_1, G_1$ and H_1. Here, let the true values of A_0 and E_0 to be zero. Then, calculate $\boldsymbol{V}_{1j}$ and $\boldsymbol{V}_{2j}$, $j=1,\ldots,n$ using equation (5) .

Step 3. Obtain the circular variable $v_j = \arctan\left(\frac{\boldsymbol{V}_{2j}}{\boldsymbol{V}_{1j}}\right)$, $j=1,\ldots,n$ using equation (3).

Step 4. Fit the generated circular data using the MCRM to give the parameter estimates of $\hat{\lambda} = \hat{A}_0, \hat{A}_1, \hat{B}_1, \hat{C}_1, \hat{D}_1, \hat{E}_0, \hat{E}_1, \hat{F}_1, \hat{G}_1$ and $\hat{H}_1$.

Step 5. Exclude the *i*th row from the generated circular data, where *i* = 1,…,*n*. For each *i*, repeat steps (4) for the reduced data set to obtain $\hat{\boldsymbol{\lambda}}_{j(i)}$.

Step 6. Compute $|DFBETAc_{j,i}|$ for each *i* from equation (8).

Step 7. Specify the maximum value of $|DFBETAc_{j,i}|$.

The process is repeated 2000 times for each combination of sample size n and standard deviation (σ_1, σ_2). Then the 5% upper percentiles of the maximum values of $|DFBETAc_{j,i}|$ are calculated and used as the cut-off points of the proposed procedure. The 1% and 10% upper percentiles are available from the authors. Table 1 gives the cut-off points of 5% percentile of the parameters estimation $\hat{\lambda} = \hat{A}_0, \hat{A}_1, \hat{B}_1, \hat{C}_1, \hat{D}_1, \hat{E}_0, \hat{E}_1, \hat{F}_1, \hat{G}_1$ and $\hat{H}_1$ for different *n* and standard deviation $(\sigma_1,\sigma_2)=(0.3,0.3)$ at *a* = 2.

The result shows that, the cut-off points present an increasing trend as σ_2 gets larger for fixed σ_1 and $\sigma_2 \geq \sigma_1$. The same trend is seen when σ_2 is fixed and $\sigma_1 \geq \sigma_2$. On the other hand, the cut-off points are increasing function of the sample size *n*.

(ii) The Performance of the DFBETAc Statistic

A simulation study is carried out to investigate the performance of the *DFBETAc* statistic for detecting outliers in the multiple circular regression model (2) based on equation (8). Four different sample size are considered, *n* = 30, 50, 70 and 100 with different value of $(\sigma_1,\sigma_2)=(0.03,0.03),(0.05,0.05),(0.1,0.1)$ and $(0.3,0.3)$. The observation at position *d*, say v_d, is contaminated as follows:

$$v_d^* = v_d + \tau\pi \mod (2\pi),$$

where v_d^* is the response value after contamination and τ is the degree of contamination in the rang $0 \leq \tau \leq 1$. The generated data of $\boldsymbol{U}_1$, $\boldsymbol{U}_2$ and $\boldsymbol{V}$ are then fitted to give the parameter estimate of $\hat{A}_0, \hat{A}_1, \hat{B}_1, \hat{C}_1, \hat{D}_1, \hat{E}_0, \hat{E}_1, \hat{F}_1, \hat{G}_1$ and $\hat{H}_1$. Consequently, exclude the *i*th row from the sample, for *i* = 1, …, *n* and refit the remaining data using equation (5). Then, the $|DFBETAc_{j.i}|$ is calculated. If the values of $|DFBETAc_d|$ is maximum and greater than the corresponding cut-off point, then the procedure has correctly detected the outlier in the data. The process is carried out 2000 times. The power of performance of the procedure is then examined by computing the percentage of the correct detection of the contamination observation at point *d*.

The simulation results are plotted in Figures 1-2. Figure 1 illustrates the power of performance of the *DFBETAc* detection method for *n* = 100 and four different values of $(\sigma_1,\sigma_2)=(0.03,0.03),(0.05,0.05),(0.1,0.1)$ and $(0.3,0.3)$. It can be seen that the performance of the procedure is a decreasing function of σ_1 and σ_2. This is expected as $\boldsymbol{V}_{1j}$ and $\boldsymbol{V}_{2j}$ in equation (5) will fluctuate closer to the horizontal axis when σ_1 and σ_2 are closer to zero, and hence, better chance to detect the outlier even when τ is small.

Table 1. The 5% upper percentiles of $\left|DFBETAc_{j,i}\right|$ statistic for $(\sigma_1,\sigma_2)=(0.3,0.3)$ at a = 2

sample size, *n*	A_0	A_1	B_1	C_1	D_1	E_0	E_1	F_1	G_1	H_1
20	3.72	2.55	2.55	2.43	1.91	4.45	2.77	3.65	2.92	2.09
30	3.82	2.67	2.92	2.44	1.94	4.53	2.83	3.54	2.98	2.24
40	4.22	2.96	3.06	2.44	2.18	4.56	2.92	3.58	2.93	2.16
50	4.26	3.00	3.01	2.63	2.29	4.88	3.13	3.59	3.17	2.23
60	4.84	3.36	3.08	2.54	2.33	5.05	3.21	3.62	3.00	2.41
70	4.93	3.37	3.08	2.75	2.36	5.13	3.37	3.64	3.40	2.42
80	5.34	3.77	3.08	2.71	2.62	5.18	3.47	3.75	3.47	2.41
90	5.43	3.80	3.18	2.90	2.69	5.55	3.60	3.82	3.69	2.96
100	5.51	3.93	3.26	3.05	2.71	5.70	3.62	3.87	3.96	2.57
120	5.57	3.97	3.39	3.08	2.83	5.71	3.71	3.99	4.10	2.68
150	5.69	4.20	4.06	3.22	2.88	5.89	4.12	4.07	4.20	2.89

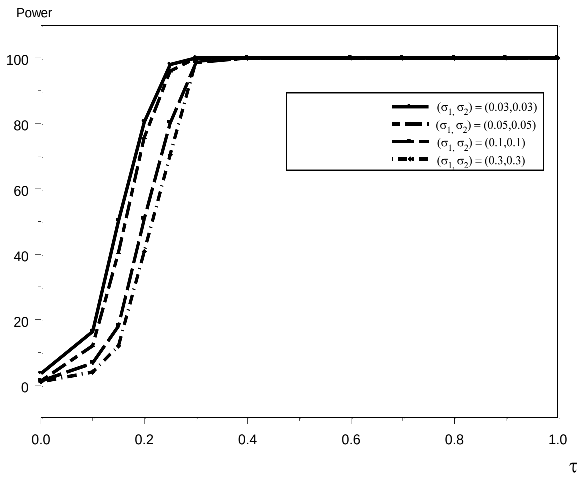


Figure 1. Power performance of *DFBETAc* statistic at n=100

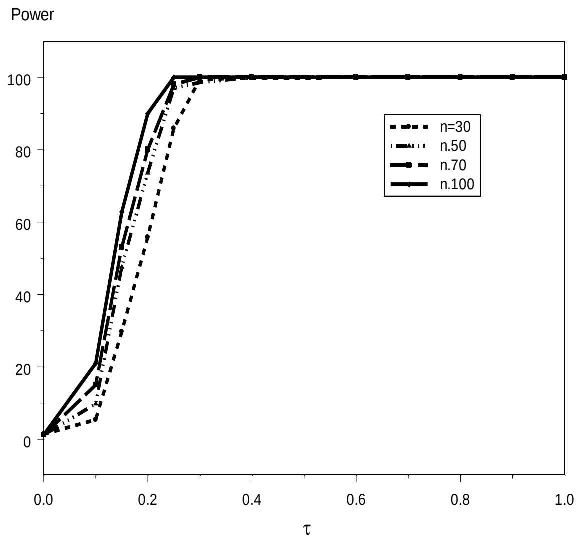


Figure 2. Power performance of *DFBETAc* statistic for $(\sigma_1,\sigma_2)=(0.1,0.1)$

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 11 (2018) pp. 9083-9090

PRACTICAL EXAMPLE

The multivariate eye data containing of 23 observations of glaucoma patients recorded using Optical Coherence Tomography at the University Malaya Medical Centre for three angles, namely the angle of the eye (v), the posterior corneal curvatures angles (u_1) and the posterior corneal curvature when length of the perpendicular is fixed to 2 mm angels (u_2) (See [10,13]). The least squares parameters estimates are $\hat{A}_0 = -1.071$, $\hat{A}_1 = 7.089$, $\hat{B}_1 = -11.685$, $\hat{C}_1 = 2.969$, $\hat{D}_1 = -1.452$, $\hat{E}_0 = 3.124$, $\hat{E}_1 = -9.963$, $\hat{F}_1 = 16.536$, $\hat{G}_1 = -4.227$, $\hat{H}_1 = 2.235$, $\hat{\sigma}_1 = 0.1$, $\hat{\sigma}_2 = 0.1$ and $\hat{\rho} = 0.987$. and thus the fitted model gives $\hat{g}_1(u_1, u_2)$ and $\hat{g}_2(u_1, u_2)$ are as follow

$$\hat{g}_1(u_1,u_2) = -1.076 + 7.089\cos u_1 \cos u_2 - 11.685 \cos u_1 \sin u_2 + 2.969 \sin u_1 \cos u_2 - 1.452 \sin u_1 \sin u_2$$

$$\hat{g}_1(u_1,u_2) = 3.124 - 9.963\cos u_1 \cos u_2 + 16.536 \cos u_1 \sin u_2 - 4.227 \sin u_1 \cos u_2 + 2.235 \sin u_1 \sin u_2$$

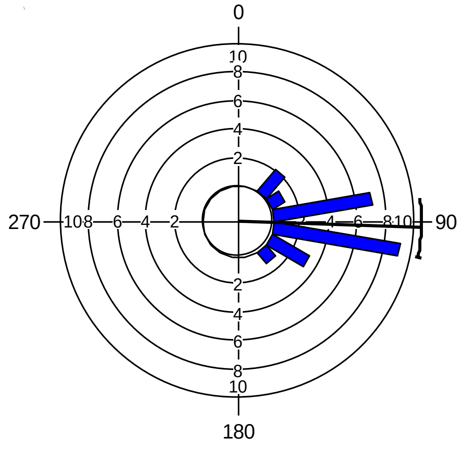
Further, the prediction of $\hat{v}_j$ given as

$$\hat{v} = \arctan^* \frac{3.1246 - 9.9634 \cos u_1 \cos u_2 + 16.5369 \cos u_1 \sin u_2 - 4.2270 \sin u_1 \cos u_2 + 2.2351 \sin u_1 \sin u_2}{-1.076 + 7.0890 \cos u_1 \cos u_2 - 11.6852 \cos u_1 \sin u_2 + 2.9691 \sin u_1 \cos u_2 - 1.4526 \sin u_1 \sin u_2}$$

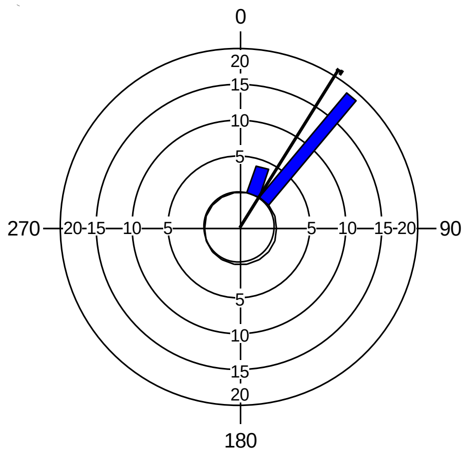
and the concentration parameter ρ toward $\mu(u)$ using equation (3) is given by $\hat{\rho}(\boldsymbol{u}) = \sqrt{\frac{1}{n}\sum_{j=1}^{n} \hat{\rho}^2(\boldsymbol{u}_j)} = 0.987$

which suggest the data seem to be highly concentrated.

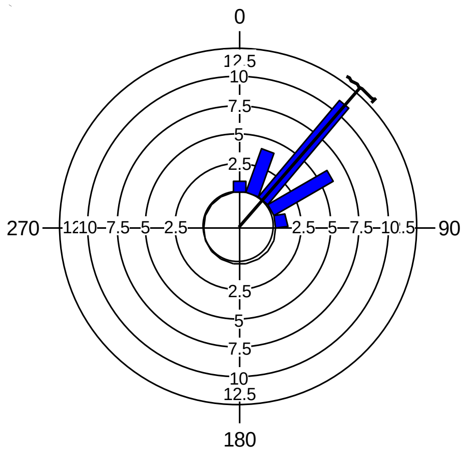
The goodness of fit test is performed by using Akaike information criteria (*AICC*) given by [14] .The *AICC* for MCRM is -51.06. This is supported by diagnostic plot. Figure 3 shows the simple circular histograms for multivariate eye data measured by three different angles. The posterior corneal curvatures angles concentrated around 90^o, the posterior corneal curvature when length of the perpendicular is fixed to 2 mm angels around 37^o, while the angle of the eye is more concentrated around 45^o. While, Figure 4 (a) and (b) show the Q-Q plots for the residuals from two observational-models of the MCRM. The plot for ε_1 shows that most of the points are closer to the straight line except two points at the top right. Meanwhile, plot of ε_2 shows the points also relatively close to the straight line except one point at the right top of the plot. These points are corresponding to the outliers that might be existing in the data. They will be dealt with using numerical statistic, *DFBETAc* statistic in the next Section.



(a) The posterior corneal curvatures angles

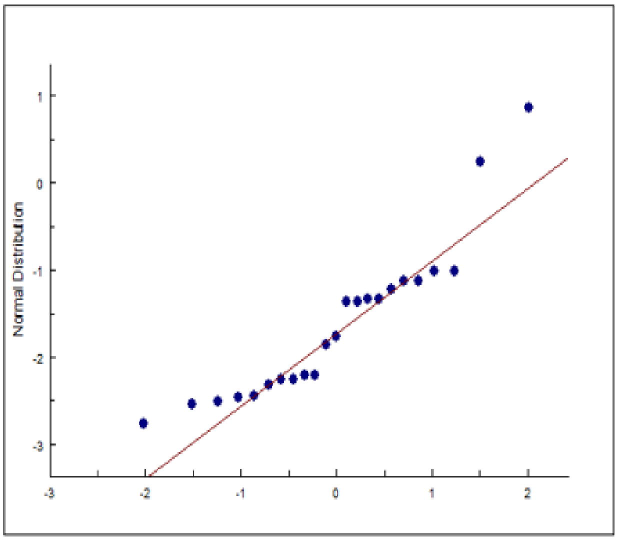


(b) The posterior corneal curvature when length of the perpendicular is fixed to 2 mm

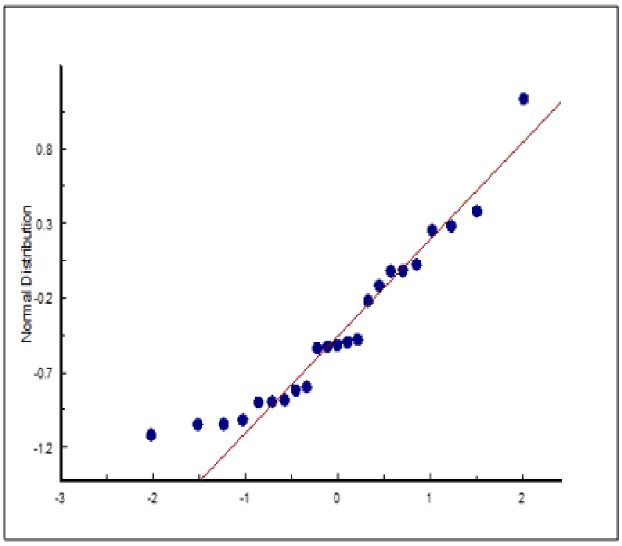


(c) The angle of the eye

Figure 3. Simple Circular Histogram of eye data



(a) ε_1



(b) ε_2

Figure 4. Q-Q plot for residuals for contaminated data

(i) *DFBETAc* Statistic

The *DFBETAc* statistic is applied to detect the possible outliers in the multivariate eye data. Since the number of multivariate eye data is 23, the cut-off point for this data is generated again and using the simulation program. The value of the parameter estimates for the 5% upper percentiles cut-off point for eye data is presented in Table 2.

Table 3 presents the values of estimated parameters and the last two columns in the table give the number parameter estimates which exceed the corresponding cut-off points. The observations number 1, observation number 15 and observation number 23 are indicated as outliers, because they are exceeded the percentage of 0.4.

(ii) The Effect of Outliers on the Parameter Estimates

In order to assess the detection of outliers, three observations; with numbers 1, 15 and 23 are identified and deleted. Then, the MCRM was refitted to get the parameter estimates. Table 4 summarizes the effect of excluding the outliers on the parameter estimates. The removal of observations with numbers 1, 15 and 23 significantly changes some of the estimated parameters of MCRM, where the parameter estimates of $\hat{A}_1, \hat{C}_1, \hat{E}_0, \hat{F}_1, \hat{H}_1, \hat{\sigma}_1$ and $\hat{\sigma}_2$ in clean data are smaller than the contaminated data. Furthermore, most of the value of standard error for parameter estimates are smaller than contaminated except for $\hat{C}_1$ and $\hat{D}_1$. On the other hand, the concentration parameter, $\hat{\rho}$ also increases from 0.987 to 0.992.

Figure 5 gives the Q-Q plots of the resulting residuals corresponding to the observational-models of the MCRM after removing those outliers from the multivariate eye data set. The plot shows all the points are close to the straight line comparing to the Figure 4, the Q-Q plots of resulting residuals corresponding to the observational before deleted the observations numbers 1, 15 and 23. That is denoting that the proposed method is the best fit for the data. Therefore, the standard errors for all parameters estimation become smaller after removing the outliers. This indicates a more accurate estimation and represent the method is perform well.

Table 2. The cut-off point value for multivriate eye data

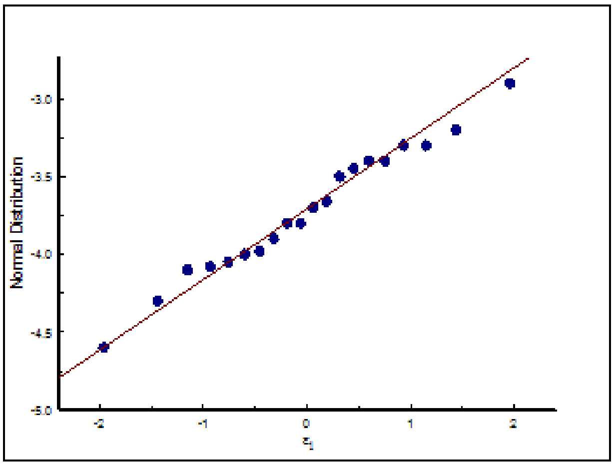
Parameter Estimate	A_0	A_1	B_1	C_1	D_1	E_0	E_1	F_1	G_1	H_1
Cut-off point value	3.82	2.63	2.69	2.35	2.05	4.10	2.78	3.22	2.81	2.13

Table 3. The values of the $\left|DFBETAc_{j,i}\right|$ statistic for multivariate eye data, n = 23 and the number influenced parameters

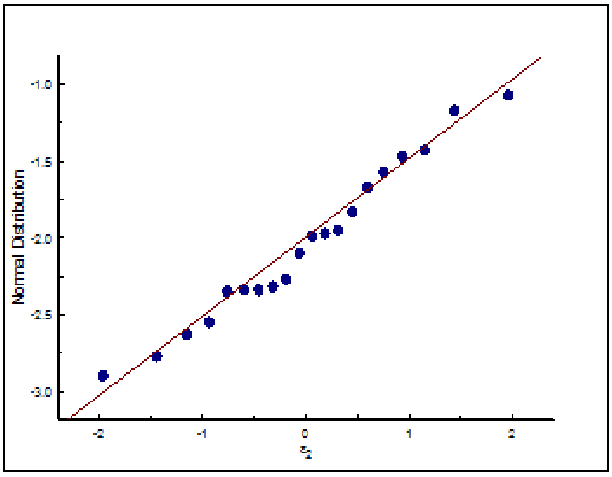
Observation	Parameter Estimate										Influenced parameters	
	A_0	A_1	B_1	C_1	D_1	E_0	E_1	F_1	G_1	H_1	Number	Percentage
1	4.69	2.89	2.39	8.69	7.47	5.74	2.99	0.29	10.64	9.14	8	0.8
2	3.93	1.33	0.85	2.93	0.38	5.53	1.88	1.20	4.12	0.53	3	0.3
3	0.21	0.06	0.04	0.16	0.02	0.27	0.08	0.05	0.21	0.02	0	0
4	0.22	0.07	0.04	0.08	0.10	4.60	1.39	0.85	1.72	2.06	1	0.1
5	0.40	0.14	0.09	0.41	0.07	2.09	0.70	0.46	2.10	0.35	0	0
6	0.85	0.36	0.25	1.38	0.65	1.12	0.48	0.33	1.83	0.86	0	0
7	0.21	0.01	0.03	0.41	0.54	0.22	0.01	0.03	0.43	0.55	0	0
8	0.15	0.01	0.01	0.24	0.07	0.11	0.01	0.01	0.17	0.05	0	0
9	2.40	0.83	0.53	1.63	0.25	2.48	0.86	0.55	1.68	0.26	0	0
10	0.12	0.16	0.09	0.12	0.02	0.04	0.06	0.03	0.05	0.01	0	0
11	0.44	0.22	0.15	0.71	0.35	0.53	0.27	0.19	0.86	0.42	0	0
12	2.38	0.74	0.47	2.22	0.30	3.48	1.08	0.68	0.47	0.43	0	0
13	1.17	0.22	0.10	0.37	1.20	1.14	0.22	0.10	0.36	1.17	0	0
14	1.51	0.65	0.44	1.72	0.46	1.63	0.68	0.45	1.78	0.48	0	0
15	7.19	2.61	1.61	0.75	4.09	4.51	1.64	1.01	3.25	2.57	5	0.5
16	1.69	0.45	0.26	0.34	0.97	1.96	0.52	0.30	0.39	1.12	0	0
17	1.61	0.45	0.30	1.20	0.18	0.95	0.27	0.18	0.71	0.11	0	0
18	0.41	0.15	0.17	2.61	2.56	0.52	0.19	0.21	3.32	3.25	4	0.4
19	0.09	0.12	0.10	0.56	0.46	0.10	0.13	0.10	0.57	0.47	0	0
20	0.29	0.03	0.01	0.25	0.04	0.30	0.03	0.01	0.26	0.04	0	0
21	1.40	0.54	0.36	1.52	0.39	1.47	0.57	0.38	1.60	0.41	0	0
22	0.47	0.19	0.12	0.06	0.28	0.87	0.34	0.22	0.11	0.52	0	0
23	7.91	3.23	2.15	6.48	2.91	4.65	3.05	0.97	2.91	2.75	8	0.8

Table 4. *DFBETAc* statistic of multivariate eye data without observations number 1, 15 and 23 , n=20

Parameter estimates	Contaminated data	Standard error	Clean data (case1,15 and 23 deleted)	Standard error
$\hat{A}_0$	-1.0716	1.0818	0.8369	0.3518
$\hat{A}_1$	7.089	3.2265	1.5022	1.4217
$\hat{B}_1$	-11.685	4.9832	-2.6209	2.4508
$\hat{C}_1$	2.9691	1.2998	-0.9197	1.3060
$\hat{D}_1$	-1.4526	1.7862	1.1060	1.8805
$\hat{E}_0$	3.1246	0.9154	0.6566	0.2144
$\hat{E}_1$	-9.9634	2.7301	-1.8945	0.8665
$\hat{F}_1$	16.5369	4.2167	3.29208	1.4937
$\hat{G}_1$	-4.227	1.0998	0.1303	0.7960
$\hat{H}_1$	2.2351	1.5114	-0.1061	1.1461
$\hat{\sigma}_1$	0.136	0.285	0.103	0.079
$\hat{\sigma}_2$	0.115	0.230	0.107	0.094
$\hat{\rho}$	0.987	-	0.992	-



(a) ε_1



(b) ε_2

Figure 5. Q-Q plot for residuals without observations numbers 1, 15 and 23 for contaminated data

CONCLUSION

This paper proposed the *DFBETAc* statistic by extending the *DFBETAS* statistic in linear to the multiple circular regression model, where the model shows appropriate features as the linear case. The cut-off points and power of performance are obtained via simulation. The proposed method was applied on eye data, and three possible outliers have been identified which are similar to observations detected by dataset of outliers as found in [10].

REFERENCES

[1] D.R. Bacon. "A Maximum Likelihood Approach to Correlational Outlier Identification." *Multivariate Behavioral Research* 30.2 (1995): 125-148.

[2] D.C. Montgomery, A.P. Elizabeth, and G.G. Vining. "Introduction to Linear Regression Analysis". Vol. 821. John Wiley & Sons, 2012.

[3] O. Torres-Reyna. "Linear regression using Stata. In Statistic Handout", Princeton University, 2007

[4] A.H. Abuzaid, A.G. Hussin, and I. Mohamed. "Identifying Single Outlier in Linear Circular Regression Model Based on Circular Distance." *Journal of Applied Probability and Statistics*3, no. 1, 2008.

[5] A. Rambli, A., I. Mohamed, A.H. Abuzaid, and A.G. Hussin. "Identification of influential observations in circular regression model." In *Proceedings of the Regional Conference on Statistical Sciences*, 2010.

[6] A.H. Abuzaid, I. Mohamed, A.G. Hussin, and A. Rambli. "COVRATIO statistic for simple circular regression model." *Chiang Mai J. Sci* 38, no. 3, 2011.

[7] A.G. Hussin, A.H. Abuzaid, A.I.N. Ibrahim, and A. Rambli. "Detection of outliers in the complex linear regression model." *Sains Malaysiana* 42, no. 6, 2013.

[8] S. Ibrahim, A. Rambli, A.G. Hussin, and I. Mohamed. "Outlier detection in a circular regression model using COVRATIO statistic." *Communications in Statistics-Simulation and Computation* 42, no. 10, 2013.

[9] A. Rambli, R. M. Yunus, I. Mohamed, and A. G. Hussin. "Outlier Detection in a Circular Regression Model." *Sains Malaysiana*, 44, no. 7, 2015.

[10] N.A. Alkasadi, S. Ibrahim, M.F. Ramli, and M.I Yusoff. "A Comparative Study of Outlier Detection Procedures in Multiple Circular Regression." In *AIP Conference Proceedings*, vol. **1775(**1), "AIP Publishing", 2016.

[11] D.A. Belsley, E. Kuh, and R.E. Welsch, "Regression Diagnostics: Identifying Influential Data and Sources of Collinearity". New York: John Wiley & Sons, 1980.

[12] S.R. Jammalamadaka, and Y.R. Sarma. "Circular Regression. In: Matsusita, K", ed. *Statistical Science and Data Analysis*. Utrecht: VSP, 1993.

[13] S. Ibrahim. "Some Outlier Problems in a Circular Regression Model", PhD Thesis, University of Malaya, 2013.

[14] U. Lund, "Least circular distance regression for directional data", *Journal of Applied Statistics*, **26** (6), 723-733, 1999.