Sains Malaysiana 48(7)(2019): 1557–1563
http://dx.doi.org/10.17576/jsm-2019-4807-25

Outlier Detection in Multiple Circular Regression Model using *DFFITc* Statistic

(Pengesanan Nilai Tersisih dalam Model Regresi Berkeliling Berganda menggunakan Statistik *DFFITc*)

NAJLA AHMED ALKASADI, SAFWATI IBRAHIM*, ALI H. M. ABUZAID, MOHD IRWAN YUSOFF,
HASHIBAH HAMID, LEOW WAI ZHE & AMELIA BT ABD RAZAK

ABSTRACT

This paper presents the identification of outliers in multiple circular regression model (MCRM), where the model studies the relationship between two or more circular variables. To date, most of the published papers concentrating on detecting outliers in circular samples and simple circular regression model with one independent circular variable. However, no related studies have been found for more than one independent circular variable. The existence of outliers could alert the sign and change the magnitude of regression coefficients and may lead to inaccurate model development and wrong prediction. Hence, the intention is to develop an outlier detection procedure using DFFITS *statistic for circular case. This method has been successfully used in multiple linear regression model. Therefore, the* DFFITc *statistic for circular variable has been derived. The corresponding critical values and the performance of the procedure are studied via simulations. The results of simulation studies show that the proposed statistic perform well in detecting outliers in MCRM using* DFFITc *statistic. The proposed statistic was applied to a real data for illustration purposes.*

Keywords: Circular data; circular regression model; DFFITS; *outlier*

ABSTRAK

Kertas ini membentangkan pengesanan nilai tersisih dalam model regresi berkeliling berganda (MCRM) dengan model tersebut mengkaji hubungan antara dua atau lebih pemboleh ubah berkeliling. Sehingga kini, kebanyakan kertas yang diterbitkan menumpukan ke atas pengesanan nilai tersisih dalam sampel berkeliling dan model regresi berkeliling ringkas untuk satu pemboleh ubah tak bersandar. Walau bagaimanapun, tiada kajian yang berkaitan telah dijumpai untuk lebih daripada satu pemboleh ubah berkeliling tak bersandar. Kewujudan nilai tersisih dapat memberi isyarat tanda dan mengubah perubahan magnitud pekali regresi dan mungkin menyebabkan pembangunan model yang tidak tepat dan ramalan yang salah. Oleh itu, objektif kajian adalah untuk membangunkan kaedah pengesanan nilai tersisih menggunakan statistik DFFITS *untuk kes berkeliling. Kaedah ini telah berjaya digunakan dalam model regresi linear berganda. Oleh itu, statistik* DFFITc *untuk pemboleh ubah berkeliling telah diterbitkan. Nilai genting sepadan dan prestasi prosedur dikaji melalui simulasi. Hasil kajian simulasi menunjukkan bahawa statistik yang dicadangkan menunjukkan prestasi yang baik dalam mengesan nilai tersisih di dalam MCRM menggunakan statistik* DFFITc. *Statistik yang dicadangkan diaplikasikan kepada data sebenar untuk tujuan ilustrasi.*

Kata kunci: Data berkeliling; DFFITS; *model regresi berkeliling; nilai tersisih*

INTRODUCTION

The multiple circular regression model is used to study the relationship between two or more circular variables as proposed by Ibrahim (2013). The model has interesting properties which are very close resemblance to that of the multiple linear regression models, including the sensitivity to the existence of outliers. Circular statistics are used in many different fields such as physics, medicine, oceanography, meteorology and biology. One of the most common problems in any statistical analysis is the existence of some unexpected observations, it is called outliers. Some studies have shown that outliers affect the performance of standard statistical methodology in modeling, diagnostic, and forecasting processes. The existence of outliers affects most of the statistical properties of the model (Abuzaid et al. 2009; Beckman & Cook 1983; Peña 1990).

The identification of outliers in circular data received great interest especially on the use the new methods, which were extended from the linear regression model to the simple circular regression model (Abuzaid et al. 2013, 2011; Hussin et al. 2013; Ibrahim et al. 2013; Rambli et al. 2015, 2012). Recently, Alkasadi et al. (2018, 2016) considered the problem of outliers in multiple circular regression model. There are few published works on the problem of outlier detection in multiple linear regression by using the *DFFITS* statistic, such as in Ampanthong & Suwattee (2009), Belsley et al. (1980), Wong (1992) and Zakaria et al. (2014).

In the literature, the methods of outliers' detection in linear case has been successfully used *DFFITS* statistic. However, there is no published work related to the detection of outliers on circular case using *DFFITS* statistic.

In this paper, we will extend the statistic of *DFFITS* to detect outliers in multiple circular regression models (MCRM).

This article is organized as follows: First, we review the multiple circular regression model and explains the estimation of model parameters via the least squares method. Next, we demonstrate the proposed of *DFFITc* statistic for the MCRM. After that, we obtains the cut-off points and investigates the performance of the proposed statistic. Lastly, we discusses the detection of outliers in multivariate eye data for illustration purpose.

THE MULTIPLE CIRCULAR REGRESSION MODEL (MCRM)

The MCRM which was proposed by Ibrahim (2013) study the relationship between a dependent circular variable and one or more independent circular variables. In this paper, we only focuses on two independent circular variables; $\boldsymbol{U}_1$, and $\boldsymbol{U}_2$ with the dependent circular variable $\boldsymbol{V}$. The MCRM in terms of the conditional expectation, e^{iv} is given by,

$$E(e^{iv}|u_1, u_2) = \rho(u_1, u_2)e^{i\mu(u_1, u_2)} = g_1(u_1, u_2) + ig_2(u_1, u_2) \tag{1}$$

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 $\mu(u_1, u_2)$.

The parameters $\mu(u_1, u_2)$ may be estimated such that

$$\mu(u_1,u_2) = \hat{v} = \arctan^{*}\frac{g_2(u_1,u_2)}{g_1(u_1,u_2)}\begin{cases} \tan^{-1}\dfrac{g_2(u_1,u_2)}{g_1(u_1,u_2)} & \text{if } g_1(u_1,u_2) \geq 0 \\ \pi + \tan^{-1}\dfrac{g_2(u_1,u_2)}{g_1(u_1,u_2)} & \text{if } g_1(u_1,u_2) \leq 0 \\ \text{undefined} & \text{if } g_1(u_1,u_2) = g_2(u_1,u_2) = 0 \end{cases} \tag{2}$$

The values of $g_1(u_1, u_2)$ and $g_2(u_1, u_2)$ may be estimated using the following *trigonometric polynomials* of a suitable degree (m) as,

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

$$g_2(u_1,u_2) \approx \sum_{k,\ell=0}^{m} \gamma_{k\ell}\begin{bmatrix} E_{k\ell}\cos ku_1 \cos \ell u_2 + F_{k\ell}\cos ku_1 \sin \ell u_2 \\ +G_{k\ell}\sin ku_1 \cos \ell u_2 + H_{k\ell}\sin ku_1 \sin \ell u_2 \end{bmatrix} \tag{3}$$

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

Thus, and based on (3), there are two models as follow;

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

$$V_{2j} = \sin v_j = \sum_{k,l=0}^{m}\begin{pmatrix} E_{k\ell}\cos ku_1 \cos \ell u_2 + F_{k\ell}\cos ku_1 \sin \ell u_2 \\ +G_{k\ell}\sin ku_1 \cos \ell u_2 + H_{k\ell}\sin ku_1 \sin \ell u_2 \end{pmatrix} + \varepsilon_{2j} \tag{4}$$

for $i = 1, \ldots, n$ and $\varepsilon = (\varepsilon_1, \varepsilon_2)$ is the vector of random errors following a bivariate normal distribution with mean $\boldsymbol{0}$ and dispersion matrix $\boldsymbol{\Sigma}$. The parameters $A_{kl}, B_{kl}, C_{kl}, D_{kl}, E_{kl}, F_{kl}, G_{kl}$ and H_{kl}, where $kl = 0, 1, \ldots, m$, the standard errors as well as the dispersion matrix $\boldsymbol{\Sigma}$ can then be estimated using generalized least squares method.

We estimate the parameters of MCRM by using the least squares method. For a random sample of size n from (4), in order to ensure identifiability, it was assumed that $B_{00} = C_{00} = D_{00} = F_{00} = G_{00} = H_{00} = 0$.

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

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

$$\boldsymbol{V}^{(2)} = \boldsymbol{U}\boldsymbol{\lambda}^{(2)} + \boldsymbol{\varepsilon}^{(2)}. \tag{5}$$

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

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

$$\hat{\boldsymbol{\lambda}}^{(2)} = (\boldsymbol{U'U})^{-1}\boldsymbol{U'V}^{(2)}. \tag{6}$$

where $\boldsymbol{U}$ is the matrix of the combination of cosine and sine functions, such that

$$\underset{n\times(4m+1)}{\boldsymbol{U}} = \begin{bmatrix} \underset{n\times 1}{1} & \underset{n\times m}{\boldsymbol{CC}} & \underset{n\times m}{\boldsymbol{CS}} & \underset{n\times m}{\boldsymbol{SC}} & \underset{n\times m}{\boldsymbol{SS}} \end{bmatrix} \tag{7}$$

The covariance matrix of the residuals, $\boldsymbol{\Sigma}$ is estimated as follow

$$\hat{\boldsymbol{\Sigma}} = [n - 2(4m+1)]^{-1} R_0 \tag{8}$$

where $R_0 = (R_0(p, q))_{p,q=1,2}$ and $R_0(p,q) = \boldsymbol{V}^{(p)\prime}\boldsymbol{V}^{(q)} - \boldsymbol{V}^{(p)\prime}\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{V}^{(q)}$, is an unbiased estimation of $\boldsymbol{\Sigma}$ and m is a suitable degree (Ibrahim 2013).

DFFITc STATISTIC OF MCRM

One of the methods to identify outliers in linear regression is *DFFITS* statistic which measures the effect of deleting a given observation on the predicted or fitted values. Belsley et al. (1980) proposed *DFFITS* statistic which is defined as,

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{i(-i)}}{\sqrt{S^2_{(-i)} h_{ii}}} \tag{9}$$

for , $i = 1, 2, \ldots, n$, where $\hat{y}_i$ and $\hat{y}_{i(-i)}$ are the prediction for point i with and without point i included in the regression, respectively. $S_{(-i)}$ denotes the standard error estimated without the point i and h_{ii} is the leverage of the point. The *DFFITS* statistic is large if the data point has high leverage which leads h_{ii} to be close to unity. Belsley et al. (1980) suggested that any observation for which $|DFFITS_i| > 2\sqrt{k/n}$ indicates outliers, where k is the number of predictor variables and n is the sample size (Cousineau & Chartier 2010; Rousseeuw & Leroy 2005). This section extends the *DFFITS* to identify possible outliers in the MCRM.

An outlier detection procedure for the MCRM is developed using row deletion approach. In regression model, it is expected that the parameter estimates, variance of residuals, covariance matrix as well as the standard errors will be affected if an outlier exists in the data. In particular, we look at the effect of removing an observation on the fitted values, at the same time it will effects on standard error estimated and covariance matrix of residuals. The proposed of *DFFITc* statistic is given by,

$$DFFITc_{ji} = \frac{\hat{\boldsymbol{V}}_j - \hat{\boldsymbol{V}}_{j(-i)}}{\sqrt{S_{(-i)}{}^2 h_{jj}}} \quad (\text{mod } 2\pi) \tag{10}$$

where $\hat{\boldsymbol{V}}_j$ denotes the prediction from the full regression model for the ith observation, meanwhile $\hat{\boldsymbol{V}}_{j(-i)}$ denotes the prediction when the ith observation is deleted. $S_{(-i)}$ denotes the standard error which is estimated without the point i while h_{jj} is the jth diagonal element of $(\boldsymbol{U}'\boldsymbol{U})^{-1}$ where the matrix $\boldsymbol{U}$ is the combination of cosine and sine functions as given in Equation (7). The ith observation is identified as an outlier if the value of *DFFITc* exceeds the pre-specified cut-off point, which will be obtained in the following section.

CRITICAL VALUES OF *DFFITc* STATISTIC

A simulation study is carried out to obtain the cut-off points of the *DFFITc* statistic for different values of different sample sizes n = 20, 40, 60, 80, 100 and standard deviations and . For m = 1, ten coefficients are to be estimated; namely, $A_0, A_1, B_1, C_1, D_1, E_0, E_1, F_1, G_1$ and H_1. For simplicity, we set the true values of A_0 and E_0 of the MCRM of order m=1 to be zero, while the other eight parameters, namely, $A_1, B_1, C_1, D_1, E_1, F_1, G_1$ and H_1 are obtained by using the standard additive trigonometric polynomial equations cos $(a + u_1 + u_2)$ and sin $(a + u_1 + u_2)$. Then, these functions are expanded using standard additive trigonometric function. For example, when $a = 2$, we have cos $(2 + u_1 + u_2) = 0.4161$ cos u_1 cos u_2 –0.9093 cos u_1 sin u_2 –0.9093 sin u_1 cos u_2 +0.4161 sin u_1 cos u_2 and sin $(2 + u_1 + u_2)$ = 0.9093 cos u_1 cos u_2 –0.4161 cos u_1 sin u_2 –0.4161 sin u_1 cos u_2 –0.9093 sin u_1 sin u_2. Then, by comparing with (4), the true values of $A_1, B_1, C_1, D_1, E_1, F_1, G_1$ and H_1 to be 0.4161, –0.9093, –0.9093, 0.4161, 0.9093, –0.4161, –0.4161 and –0.9093, respectively, with A_0 and E_0 being zero. Similarly, we can also get different sets of true values by choosing different values of a (Ibrahim 2013).

Then, the 10%, 5% and 1% upper percentiles of the maximum values of *DFFITc* are obtained. The full procedures to obtain the critical values are summarized as follows;

1. Generate the independent variables $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ of size n from von Mises distribution with mean π and concentration parameters 3 and 2 (VM(π , 3) and VM(π , 2)), respectively.
2. Generate ε_1 and ε_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 (4) .
3. Obtain the circular variable $v_j = \arctan\left(\frac{\boldsymbol{V}_{2j}}{\boldsymbol{V}_{1j}}\right), j = 1, \ldots, n$ using (2).
4. Fit the generated circular data using the MCRM to give the parameter estimates of $\hat{\boldsymbol{\lambda}} = \hat{A}_0, \hat{A}_1, \hat{B}_1, \hat{C}_1, \hat{D}_1, \hat{E}_0, \hat{E}_1, \hat{F}_1, \hat{G}_1$ and $\hat{H}_1$ as given in (6).
5. Exclude the ith row from the generated circular data, where $i = 1, \ldots, n$. For each i, repeat steps (4) for the reduced data set to obtain $\hat{\boldsymbol{\lambda}}_{j(-i)}$.
6. Compute $DFFITc_{ji}$ for each i from (10).
7. Specify the maximum value of $DFFITc_{ji}$.

The process is repeated 2000 times for each combination of sample size n and standard deviation (σ_1, σ_2) = (0.03, 0.03),(0.05, 0.05), (0.1, 0.1) and (0.3, 0.3).

Table 1 represents the critical values at 5% upper percentiles for different sample size n and standard deviation (σ_1, σ_2) at a=2. The others critical values can be obtained from the authors upon request. The results show that, for a fixed σ_1 and $\sigma_2 \geq \sigma_1$, the cut-off point increases as σ_1 gets larger. A similar pattern is observed for a fixed σ_2 where $\sigma_1 \geq \sigma_2$. This is because the residual error values will be small and fluctuated around the unit circle. Thus, for low leverage points, the values of *DFFITc* are expected to be small, whereas as the leverage goes to 1 the distribution of the *DFFITc* value enlarges infinitely. On the other hand, the cut-off points have a decreasing function of the sample size n.

TABLE 1. Cut-off points at 5% upper percentiles of *DFFITc* statistic at a=2

n	σ_1	σ_2 0.03	0.05	0.08	0.1	0.3
	0.03	0.0761	0.0809	0.0835	0.0850	0.0888
	0.05	0.0707	0.0738	0.0780	0.0782	0.0799
20	0.08	0.0795	0.0815	0.0817	0.0827	0.0864
	0.1	0.0804	0.0822	0.0848	0.0845	0.0859
	0.3	0.0720	0.0722	0.0723	0.0750	0.0765
	0.03	0.0475	0.0490	0.0552	0.0579	0.0695
	0.05	0.0489	0.0500	0.0531	0.0560	0.0704
40	0.08	0.0510	0.0521	0.0522	0.0535	0.0704
	0.1	0.0536	0.0536	0.0539	0.0543	0.0690
	0.3	0.0668	0.0677	0.0679	0.0698	0.0693
	0.03	0.0352	0.0381	0.0426	0.0445	0.0550
	0.05	0.0367	0.0370	0.0395	0.0421	0.0540
60	0.08	0.0388	0.0391	0.0391	0.0411	0.0538
	0.1	0.0389	0.0398	0.0403	0.0404	0.0525
	0.3	0.0487	0.0499	0.0500	0.0511	0.0514
	0.03	0.0275	0.0318	0.0347	0.0357	0.0516
	0.05	0.0289	0.0291	0.0328	0.0340	0.0513
80	0.08	0.0310	0.0314	0.0316	0.0321	0.0502
	0.1	0.0305	0.0313	0.0313	0.0318	0.0486
	0.3	0.0411	0.0419	0.0426	0.0431	0.0445
	0.03	0.0246	0.0273	0.0296	0.0312	0.0418
	0.05	0.0259	0.0263	0.0276	0.0276	0.0416
100	0.08	0.0263	0.0266	0.0272	0.0284	0.0411
	0.1	0.0250	0.0253	0.0282	0.0293	0.0418
	0.3	0.0298	0.0323	0.0365	0.0394	0.0421

THE PERFORMANCE OF *DFFITc* STATISTIC

A simulation study is carried out to investigate the performance of *DFFITc* statistic for detecting outliers in the MCRM. Five different sample size are considered, n = 20, 40, 60, 80 and 100 with different value of (σ_1, σ_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:

$$v_d^* = v_d + \tau\pi \quad (\bmod 2\pi)$$

where v_d^* is the response value after contamination and τ is the degree of contamination in the range $0 \le \tau \le 1$. The generated data of $\boldsymbol{U}_1$, $\boldsymbol{U}_2$ and $\boldsymbol{V}$ are then fitted to obtain the parameter estimates 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 ith row from the sample, for i=1, …, n and refit the remaining data using (6). Then, the $DFFITc_{ji}$ is calculated. If the values of *DFFITc* is maximum and greater than the corresponding cut-off point, then the procedure has correctly detecte the outlier in the data. The process is carried out 5000 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.

Figure 1 illustrates the power of performance of *DFFITc* statistic for n=100 and four values of standard deviation (σ_1, σ_2) = (0.03, 0.03), (0.05, 0.05), (0.1, 0.1) and (0.3, 0.3). It is shown that the power of performance is an increasing function of contamination level provided that the standard deviation, and decreasing function. The power of performance depends on the level of contamination, τ, where the proposed statistic is able to detect almost all contamination points for $\tau > 0.3$.

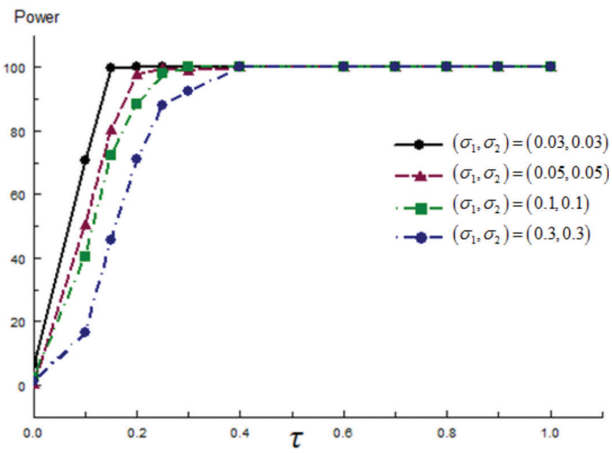

FIGURE 1. Power performance for *DFFITc* statistic for n=100

Figure 2 shows the performance of *DFFITc* statistic for and different sample sizes, n. For *DFFITc* statistic, the power of performance is an increasing function of sample size n. It is also clear and show good performance if the contaminated level is greater than 0.3, ($\tau \geq 0.3$). Similar results were obtained for fixed values of $(\sigma_1, \sigma_2) = (0.1, 0.1)$ and different sample sizes n as shown in Figure 3. These results are supported by Alkasadi et al. (2018) and Ibrahim et al. (2013).

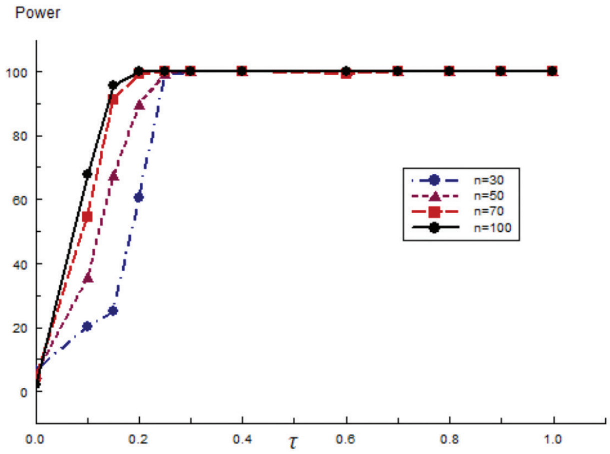

FIGURE 2. Power performance for *DFFITc* statistic for $(\sigma_1,\sigma_2)=(0.05,0.05)$

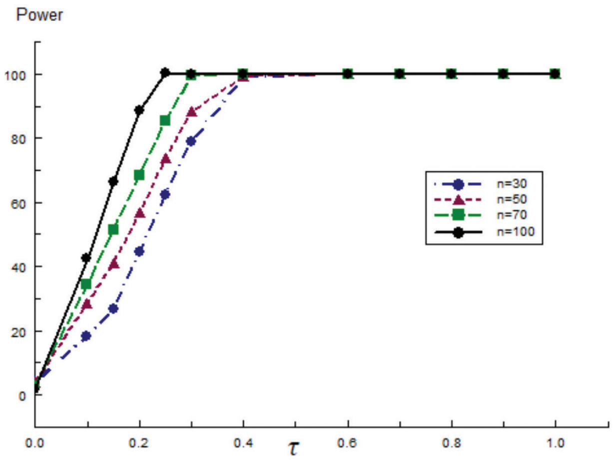

FIGURE 3. Power performance for *DFFITc* statistic for $(\sigma_1,\sigma_2)=(0.1,0.1)$

PRACTICAL EXAMPLE: MULTIVARIATE EYE DATA

The multivariate eye data consist of 23 observations of glaucoma patients recorded using Optical Coherence Tomography at the University Malaya Medical Centre, Malaysia for three angles v, u_1, u_2 (Alkasadi et al. 2018 & Ibrahim 2013).

Thus, the MCRM is used to fit the multivariate eye data. The least squares parameter estimates of MCRM are given as follow;

$$\hat{A}_0 = -1.076, \hat{A}_1 = 7.0890, \hat{B}_1 = -11.6852, \hat{C}_1 = 2.9691, \hat{D}_1 = -1.4526, \hat{E}_0 = 3.1246, \hat{E}_1 = -9.9634, \hat{F}_1 = 16.5369, \hat{G}_1 = -4.2270, \hat{H}_1 = 2.2351, \hat{\sigma}_1 = 0.17 \text{ and } \hat{\sigma}_2 = 0.14.$$

The MCRM of multivariate eye data respect to $\hat{g}_1(u_1, u_2)$ and $\hat{g}_2(u_1, u_2)$ are given by

$$\hat{g}_1(u_1, u_2) = -1.076 + 7.0890 \cos u_1 \cos u_2 - 11.6852 \cos u_1 \sin u_2 + 2.9691 \sin u_1 \cos u_2 - 1.4526 \sin u_1 \sin u_2$$

$$\hat{g}_2(u_1, u_2) = 3.1246 - 9.9634 \cos u_1 \cos u_2 + 16.5369 \cos u_1 \sin u_2 - 4.2270 \sin u_1 \cos u_2 + 2.2351 \sin u_1 \sin u_2$$

Figure 4 illustrates the Q-Q plot for residuals. The corresponding plot of ε_1 shows that almost all points are adjacent to the straight line excluding only two points positioned at the upper right of the plot (observations number 1 and 23). Meanwhile, plot of ε_2 also showed that almost all points are adjacent to the straight line, excluding only one point positioned at the upper right of the plot (observation number 1).

By applying *DFFITc* statistic on the regression model of multivariate eye data in order to detect any possible outliers, we obtain the cut-off point equals to 0.08134. Figure 5 shows only one observation is above the specific cut-off point, which is observation number one.

Table 2 presents the effect of removing of detected outlier on the parameters estimates. Upon excluding observation 1 from the multivariate eye data set changes the value of the parameters estimates where the standard

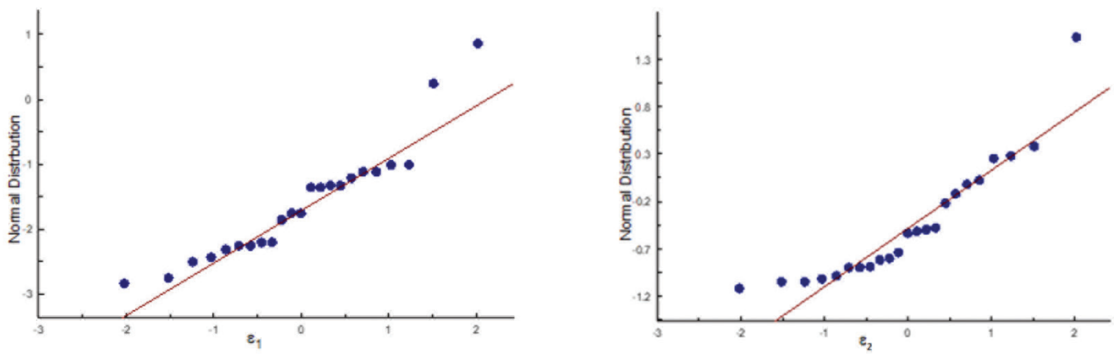

FIGURE 4. The Q-Q plot for residuals of fitted MCRM for eye data (n =23)

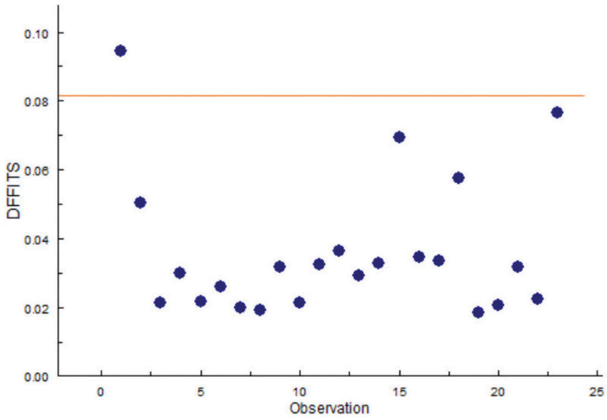

FIGURE 5. The values of *DFFITc* statistic to the multivariate eye data (n =23)

error becomes smaller for all parameters estimate. The concentration parameter increases from 0.96 to 0.992 upon deleting observation number one. Figure 6 plot shows that the quantiles are close to the straight line indicating the best fit for the data.

CONCLUSION

A new outlier detection statistic for multiple circular model was proposed by extending the *DFFITc* statistic from the multiple linear regression model and based on row deletion approach.

The cut-off points are obtained and the power of performance were investigated via simulation study. The statistic showed a very good performance in identifying prospective outlier in the MCRM even for lower level of contamination. The application of the proposed statistics on the multivariate eye data revealed one outlier which is consistent to the findings of Alkasadi et al. (2018, 2016).

ACKNOWLEDGEMENTS

Authors are most grateful and would like to thank the reviewers for their valuable suggestions, which led to an improvement of the article. The authors are also gratefully

TABLE 2. MCRM parameters estimates for full and reduced multivariate eye data

Parameter estimates	Full data	Standard error	Reduced data (case 1 deleted)	Standard error
$\hat{A}_0$	-1.071	(0.037)	-1.097	(0.036)
$\hat{A}_1$	7.089	(0.050)	7.244	(0.042)
$\hat{B}_1$	-11.685	(0.053)	-12.192	(0.052)
$\hat{C}_1$	2.969	(0.027)	3.021	(0.025)
$\hat{D}_1$	-1.452	(0.036)	-1.471	(0.036)
$\hat{E}_0$	3.124	(0.039)	3.176	(0.030)
$\hat{E}_1$	-9.963	(0.058)	-9.954	(0.042)
$\hat{F}_1$	10.536	(0.037)	10.537	(0.029)
$\hat{G}_1$	-4.227	(0.047)	-4.052	(0.035)
$\hat{H}_1$	2.235	(0.036)	2.026	(0.026)
$\hat{\sigma}_1$	0.136	(0.285)	0.120	(0.110)
$\hat{\sigma}_2$	0.115	(0.230)	0.112	(0.133)
$\hat{\rho}$	0.96	-	0.992	-

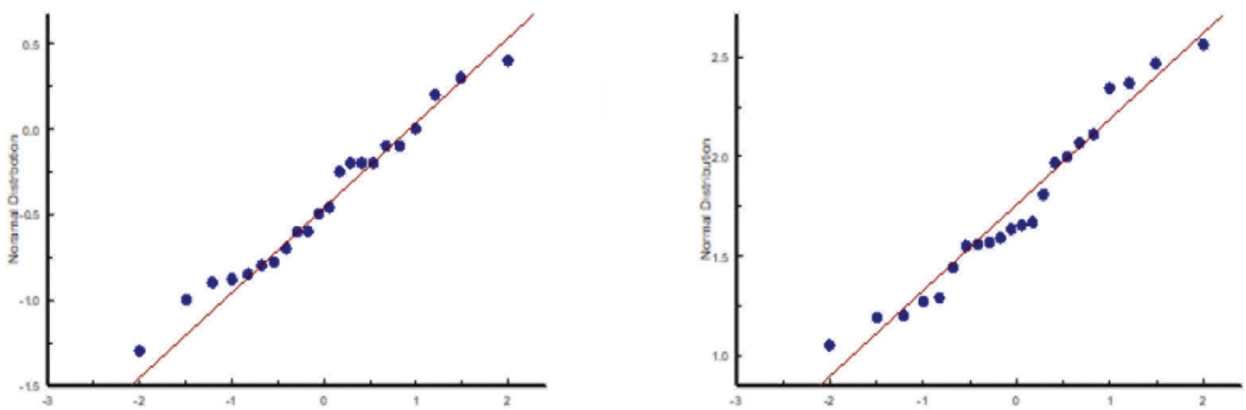

FIGURE 6. Q-Q plot of circular residuals after removing observation number 1 (n =22)

acknowledge the financial support from the Ministry of Higher Education, Malaysia (FRGS/1/2019/STG06/UNIMP/02/7).

REFERENCES

Abuzaid, A.H., Hussin, A.G. & Mohamed, I.B. 2013. Detection of outliers in simple circular regression models using the mean circular error statistic. *Journal of Statistical Computation and Simulation* 83(2): 269-277.

Abuzaid, A., Mohamed, I., Hussin, A.G. & Rambli, A. 2011. COVRATIO statistic for simple circular regression model. *Chiang Mai International Journal of Science and Technology* 38(3): 321-330.

Abuzaid, A.H., Mohamed, I.B. & Hussin, A.G. 2009. A new test of discordancy in circular data. *Communications in Statistics-Simulation and Computation* 38(4): 682-691.

Alkasadi, N.A., Abuzaid, A.H., Ibrahim, S. & Yusoff, M.I. 2018. Outliers detection in multiple circular regression models via *DFBETAc* statistic. *International Journal of Applied Engineering Research* 13(11): 9083-9090.

Alkasadi, N.A., Ibrahim, S., Ramli, M.F. & Yusoff, M.I. 2016. A comparative study of outlier detection procedures in multiple circular regression. *AIP Conference Proceedings* 1775(1): 030032.

Ampanthong, P. & Suwattee, P. 2009. A comparative study of outlier detection procedures in multiple linear regression. *In Proceedings of the International MultiConference of Engineers and Computer Scientists* Volume 1.

Beckman, R.J. & Cook, R.D. 1983. Outlier.......... s. *Technometrics* 25(2): 119-149.

Belsley, D.A., Kuh, E. & Welsch, R.E. 1980. *Regression Diagnostic: Identifying Influential Data and Sources of Collinearity.* New York: John Wiley & Sons.

Cousineau, D. & Chartier, S. 2010. Outliers detection and treatment: A review. *International Journal of Psychological Research* 3(1): 58-67.

Hussin, A.G., Abuzaid, A.H., Ibrahim, A.I.N. & Rambli, A. 2013. Detection of outliers in the complex linear regression model. *Sains Malaysiana* 42(6): 869-874.

Ibrahim, S. 2013. *Some Outlier Problems in a Circular Regression Model*. PhD Thesis, University of Malaya (Unpublished).

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

Peña, D. 1990. Influential observations in time series. *Journal of Business & Economic Statistics* 8(2): 235-241.

Rambli, A., Yunus, R.M., Mohamed, I. & Hussin, A.G. 2015. Outlier detection in a circular regression model. *Sains Malaysiana* 44(7): 1027-1032.

Rambli, A., Ibrahim, S., Abdullah, M.I., Mohamed, I. & Hussin, A.G. 2012. On discordance test for the wrapped normal data. *Sains Malaysiana* 41(6): 769-778.

Rousseeuw, P.J. & Leroy, A.M. 2005. *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.

Wong, C. 1992. *Diagnostic and Influence Measures in Linear Regression*. PhD Thesis. Simon Fraser University (Unpublished).

Zakaria, A., Howard, N.K. & Nkansah, B.K. 2014. On the detection of influential outliers in linear regression analysis. *American Journal of Theoretical and Applied Statistics* 3(4): 100-106.

Najla Ahmed Alkasadi & Safwati Ibrahim*
Institute of Engineering Mathematics
Universiti Malaysia Perlis
Pauh Putra Main Campus
02600 Arau, Perlis Indera Kayangan
Malaysia

Ali H.M. Abuzaid
Department of Mathematics
Faculty of Science
Al-Azhar University-Gaza
Palestine

Mohd Irwan Yusoff
Center for Diploma Studies
S2-L1-26, Kampus Uniciti Sungai Chuchuh
Universiti Malaysia Perlis
02100 Padang Besar (U), Perlis Indera Kayangan
Malaysia

Hashibah Hamid
School of Quantitative Sciences
College of Arts & Sciences
Universiti Utara Malaysia (UUM)
06010 UUM Sintok, Kedah Darul Aman
Malaysia

Leow Wai Zhe & Amelia Bt Abd Razak
School of Electrical System Engineering
Universiti Malaysia Perlis
Pauh Putra Main Campus
02600 Arau, Perlis Indera Kayangan
Malaysia

*Corresponding author; email: isafwati@gmail.com

Received: 16 October 2018
Accepted: 3 May 2019