Sains Malaysiana 47(6)(2018): 1319–1326
http://dx.doi.org/10.17576/jsm-2018-4706-29

A New Discordancy Test on a Regression for Cylindrical Data

(Ujian Ketakselanjaran Terbaru ke atas Regresi untuk Data Silinder)

NURUL HIDAYAH SADIKON, ADRIANA IRAWATI NUR IBRAHIM*, IBRAHIM MOHAMED & DHARINI PATHMANATHAN

ABSTRACT

A cylindrical data set consists of circular and linear variables. We focus on developing an outlier detection procedure for cylindrical regression model proposed by Johnson and Wehrly (1978) based on the k-nearest neighbour approach. The procedure is applied based on the residuals where the distance between two residuals is measured by the Euclidean distance. This procedure can be used to detect single or multiple outliers. Cut-off points of the test statistic are generated and its performance is then evaluated via simulation. For illustration, we apply the test on the wind data set obtained from the Malaysian Meteorological Department.

Keywords: Circular-linear; cylindrical data; k-nearest neighbour's distance; outlier

ABSTRAK

Data silinder adalah data yang mengandungi pemboleh ubah bulatan dan linear. Kami memberi tumpuan kepada pembangunan prosedur pengecaman nilai tersisih untuk model regresi silinder yang dicadangkan oleh Johnson dan Wehrly (1978) dengan menggunakan pendekatan jiran k-terdekat. Prosedur tersebut adalah berdasarkan nilai-nilai reja dengan jarak di antara dua reja diukur menggunakan jarak Euclidean. Prosedur ini boleh digunakan untuk mengesan nilai tersisih tunggal atau berbilang. Titik potongan untuk statistik ujian dijana dan prestasi bagi ujian tersebut dikaji secara simulasi. Untuk ilustrasi, kami menggunakan set data angin yang diperoleh daripada Jabatan Meteorologi Malaysia.

Kata kunci: Bulatan-linear; data silinder; jarak jiran k-terdekat; nilai tersisih

INTRODUCTION

In statistical modeling, regression analysis is one of the common methods used to investigate the relationship between variables. For the linear case, the theory of linear regression is readily available in various literature. As for the circular regression, it can be divided into different types according to the type of dependent of independent variables (Jammalamadaka & SenGupta 2001). Circular-circular regression is a type of regression when both the dependent and independent variables are circular; circular-linear regression is a regression when the linear variable depends on the independent circular variable while linear-circular regression is a type of regression when the circular variable depends on the linear variable. The regression for cylindrical data can be considered as the circular-linear regression or the linear-circular regression.

Johnson and Wehrly (1978) proposed a regression of a linear variate on other linear and circular variates in which the model follows closely the linear regression; the least-square method is used to find the parameter estimates. Then, SenGupta and Ugwuowo (2006) proposed three different models of circular-linear regression for multivariate data based on both circular and linear predictors. These models can be used to deal with both symmetric and asymmetric model forms. Qin et al. (2011) proposed a nonparametric regression model for circular-linear multivariate regressors using a kernel-weighted local linear method.

Outliers can affect the estimation of a regression model. In linear regression, many outlier detection methods have been proposed in the literature. For the case of a single outlier, Barnett and Lewis (1978) and Srikantan (1961) used residuals from the least square fit in their outlier detection procedures. Cook (1977) presented a new distance measure based on two maximum likelihood estimates using row-deletion approach. Srivastava and Rosen (1998) proposed a likelihood ratio test for detecting single outlier in multivariate regression models. For the case of multiple outliers, Hadi and Simonoff (1993) proposed procedures to detect outliers in univariate linear regression model. Barrett and Ling (1992) presented general classes of multivariate influence measure for a univariate regression based on Cook's influence measure. The outlier detection in circular regression mainly focuses on the circular-circular regression models. Abuzaid et al. (2011) and Ibrahim et al. (2013) extended the COVRATIO statistic that is used in linear regression to a circular-circular regression model. Abuzaid et al. (2013) and Rambli et al. (2016) proposed new outlier detection methods in the circular-circular regression models called mean circular error statistic by using row-deletion method. Rambli et al. (2016) transformed the residuals into linear scales using a trigonometric function while Abuzaid et al. (2013) used the circular distance between two circular observations. While different outlier detection procedures have been developed for linear and circular regression models, no such work has been done on the regression model for cylindrical data. Hence, we propose a new test of outlier detection in the regression model for cylindrical data.

Thus in this article, the regression model for cylindrical data, in particular the Johnson-Wehrly (JW) circular-linear regression model, is discussed and a discussion on the k-nearest neighbour approach is given. Then, a new outlier detection method for circular-linear regression model based on the k-nearest neighbour approach is presented. Next, the cut-off points of the new statistic are calculated, and its performance is studied through simulation. An application of the new test of discordancy is later shown using real data set from the Malaysian Meteorological Department.

REGRESSION FOR CYLINDRICAL DATA

Johnson and Wehrly (1978) proposed three different regression models including a regression of a linear variable on other linear and circular variables. Herewith, we refer to the model as the JW circular-linear regression model. Consider the joint density $f(\boldsymbol{x}, \boldsymbol{\theta})$ such that

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = c \exp\left\{-\frac{1}{2}\boldsymbol{x}'\sum^{-1}\boldsymbol{x} + \boldsymbol{\lambda}'\sum^{-1}\boldsymbol{x} + \boldsymbol{a}(\boldsymbol{\theta})'\sum^{-1}\boldsymbol{x}\right\} \tag{1}$$

where c is a constant of integration and $\boldsymbol{a}(\boldsymbol{\theta})' = (a_1(\boldsymbol{\theta}), \ldots, a_q(\boldsymbol{\theta}))$ is given by

$$a_i(\boldsymbol{\theta}) = \sum_{j=1}^{p}\sum_{k=1}^{n} [u_{ijk}\cos(k\theta_j) + v_{ijk}\sin(k\theta_j)], \quad i = 1, \ldots, q, \tag{2}$$

$\boldsymbol{x} \in \mathbb{R}^q$, $\boldsymbol{\theta} \in [0, 2\pi)^p$, $\boldsymbol{\lambda} \in \mathbb{R}^q$, Σ^{-1} is positive definite while u_{ijk} and v_{ijk} are constant. Let us partition $\boldsymbol{x} = \left(\boldsymbol{x}_1' \middle| \boldsymbol{x}_2'\right)'$ and hence $\boldsymbol{\lambda}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{a}(\boldsymbol{\theta})$ accordingly. The model is constructed from the conditional distribution of $\boldsymbol{x}_1 = (x_1, x_2, \ldots, x_r)'$ given $\boldsymbol{x}_2$ and $\boldsymbol{\theta}$, where $\boldsymbol{x}_1$ is the dependent variable while $\boldsymbol{x}_2$ and $\boldsymbol{\theta}$ are the independent variables. The conditional distribution $f(\boldsymbol{x}_1 | \boldsymbol{x}_2, \boldsymbol{\theta})$ is the r-dimensional normal distribution with mean $\boldsymbol{\lambda}_1 + \sum_{12}\sum_{22}^{-1}[\boldsymbol{x}_2 - (\boldsymbol{\lambda}_2 + \boldsymbol{a}_2(\boldsymbol{\theta}))]$ and co-variance matrix $\sum_{11} + \sum_{12}\sum_{22}^{-1}\sum_{21}$. It can be shown that each component x_i of $\boldsymbol{x}_1$ has a variance not depending on the conditioning variables and a mean of the form

$$\beta_0 + \sum_{i=r+1}^{q}\beta_i x_i + \sum_{i=r+1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}\left[\gamma_{ijk}\cos\left(k\theta_j\right) + \delta_{ijk}\sin\left(k\theta_j\right)\right] \tag{3}$$

where $\beta_0, \beta_i, \gamma_{ijk}$ and δ_{ijk} are the coefficients which represent the relationship between the variables and k is the angular frequency. This model is basically reduced to a standard method of predicting a linear variable from a mixture of linear and circular variables.

In the next section, we use a simple form of the model given in (3) with one linear variable and one circular variable with the frequency $k = 1$. The model takes the form of

$$x_{1i} = \beta_0 + \beta_2 x_{2i} + \gamma\cos\theta_i + \delta\sin\theta_i + \epsilon_i, \quad i = 1,2,\ldots,n \tag{4}$$

$\epsilon_i \sim N(0, \sigma^2)$. The estimation of the parameters β_0, β_2, γ and δ can be obtained using the least square estimation method.

THE k-NEAREST NEIGHBOR APPROACH

We denote $d(x_i, x_1), d(x_i, x_2), \ldots, d(x_i, x_n)$ as the distances between the ith observation with the other observations while $d_{(1)}(x_i, x_1), d_{(2)}(x_i, x_2), \ldots, d_{(k)}(x_i, x_n)$ are the corresponding ordered distances. The first-nearest distance for the ith observation is defined as the smallest distance or the distance at the first position in the ordered distances given by,

$$L_{1i} = d_{(1)}(x_i, x_j), \; i, j = 1, 2, \ldots, n, \; i \neq j. \tag{5}$$

Note that L_{1i} gives a sequence of distances between consecutive observations on the p-dimensional surface. Hence, we can define L_{ki} as the k-NN distance for the ith observation to other points where,

$$L_{ki} = d_{(k)}(x_i, x_j), \; i, j = 1, 2, \ldots, n, \; i \neq j. \tag{6}$$

In the next section, we develop a new test of discordancy to detect outliers in JW circular-linear regression model using the statistic as given in (6), but on residuals, instead of observations.

OUTLIER DETECTION IN A REGRESSION MODEL FOR CYLINDRICAL DATA USING k-NN APPROACH

The new outlier detection for JW circular-linear model is constructed based on the k-NN approach when applied on the distances between the residuals. The residual is given by,

$$e_i = x_{1i} - \hat{x}_{1i}, \quad \mathrm{i} = 1, 2, \ldots, \mathrm{n}. \tag{7}$$

Given e_i and e_j, the Euclidean distance between the two residuals is defined as,

$$d(e_i, e_j) = |e_i - e_j|, \; i = 1, 2, \ldots, n. \tag{8}$$

Using the same k-NN approach given in the previous section, the k-NN distance for this case is given by,

$$L_{ki} = d_{(k)}(e_i, e_j), \; k = 1,2,3,\ldots, \quad i, j = 1,2,\ldots, n, \; i \neq j \tag{9}$$

Hence, the test statistic is

$$L_n^k = \max_i \left[L_{ki} \right] \tag{10}$$

where n is the sample size and k is the kth-nearest neighbour. The complete steps to detect the outlier in regression for cylindrical data are given as follows:

First, fit the circular-linear regression as given in (4); next, calculate the residuals as defined in (7); after that, choose any $k = 1,2,3,...$ for the k-nearest neighbour distance, then calculate the distances between residuals, L_{ki}, as given in (9); subsequently, define the test statistic L_n^k as given in (10); and lastly if the value of L_n^k exceeds the cut-off point, say a_L, then the ith observation corresponding to $\max_i \left[L_{ki} \right]$ is identified as an outlier.

We note that L_n^k statistic can also be used to detect a patch of outliers. For example, when $k = 1$, it can be used to detect an outlier while when $k = 2$, it can be used to detect a patch of 2 outliers. For multiple outliers, we usually need to repeat the L_n^k statistic iteratively for $k = 1,2,3,...$ until no outliers are detected.

CUT-OFF POINTS OF THE TEST STATISTIC

We design a simulation study for L_n^k statistic to obtain the cut-off points using the *R* statistical package based on the null hypothesis that there are no outliers present in the cylindrical data set. The generation of the cut-off points are based on the sample size n and residual standard deviation σ.

In our study, the cut-off points are generated from various values of n in the range of [10, 100] and σ = [0.05, 1]. We generate x_2 from Normal distribution N(5,2) and θ from von Mises distribution VM(π,2). Then, we generate the residuals of size n from N(0,σ). For each sample, we obtain the variable x_1 using (4) where the values of the JW model parameters are chosen to be $\beta_0 = 0.306$, $\beta_2 = 1$, γ = 1 and δ = 1. Next, we fit the JW model and compute the fitted values and resulting residuals. We then calculate {L_n^k, k = 1,2} statistic as given in (10). The process is repeated for 2000 times and the estimated cut-off points at 10%, 5% and 1% upper percentiles are collected.

The cut off-points of L_n^k statistic are tabulated in Table 1 for the case of a single outlier (k = 1) and two outliers (k = 2). It can be seen that for each sample size n, the cut-off points increase as the value of σ increases. On the other hand, the cut-off points are a decreasing function of sample size n.

THE PERFORMANCE OF L_n^k STATISTIC

THE PERFORMANCE OF L_n^1 STATISTIC FOR SINGLE OUTLIER

From Barnett and Lewis (1978) and David (1981), P1 = 1 - β is the power function where β is the type-II error; P3 is the probability that the contaminant point is an outlier and it is identified as discordant; and P5 is the probability that that the contaminant point is an outlier given that it is identified as discordant. A good test should have: high P1; high P5; and low P1-P3.

TABLE 1. The cut-off points of statistic for the JW distribution where $\beta_0 = 0.306$, $\beta_2 = 1$, $\gamma = 1$ and $\delta = 1$

n	Level of percentile	k = 1						k = 2					
		σ											
		0.08	0.2	0.5	0.8	1	2	0.08	0.2	0.5	0.8	1	2
10	10%	0.096	0.239	0.598	0.956	1.196	2.391	0.125	0.312	0.779	1.246	1.558	3.116
	5%	0.114	0.284	0.710	1.136	1.420	2.840	0.144	0.361	0.901	1.442	1.803	3.606
	1%	0.146	0.366	0.914	1.463	1.828	3.657	0.180	0.451	1.127	1.803	2.254	4.508
20	10%	0.094	0.235	0.587	0.939	1.174	2.349	0.119	0.299	0.747	1.195	1.494	2.987
	5%	0.111	0.277	0.691	1.106	1.383	2.765	0.135	0.338	0.845	1.352	1.689	3.379
	1%	0.144	0.361	0.902	1.443	1.804	3.608	0.165	0.413	1.033	1.652	2.065	4.130
30	10%	0.088	0.221	0.553	0.884	1.105	2.211	0.114	0.286	0.715	1.143	1.429	2.858
	5%	0.106	0.265	0.662	1.059	1.324	2.647	0.130	0.326	0.814	1.303	1.629	3.258
	1%	0.142	0.354	0.885	1.417	1.771	3.542	0.166	0.416	1.041	1.665	2.081	4.162
50	10%	0.084	0.210	0.525	0.839	1.049	2.098	0.108	0.270	0.675	1.079	1.349	2.698
	5%	0.103	0.258	0.644	1.031	1.288	2.576	0.125	0.313	0.783	1.252	1.565	3.130
	1%	0.143	0.356	0.891	1.426	1.782	3.565	0.161	0.403	1.006	1.610	2.013	4.026
80	10%	0.081	0.202	0.505	0.809	1.011	2.022	0.103	0.257	0.642	1.027	1.283	2.567
	5%	0.098	0.244	0.609	0.975	1.219	2.438	0.119	0.298	0.746	1.194	1.492	2.984
	1%	0.129	0.323	0.808	1.293	1.616	3.232	0.148	0.370	0.925	1.481	1.851	3.702
100	10%	0.082	0.204	0.510	0.817	1.021	2.042	0.101	0.253	0.631	1.010	1.263	2.526
	5%	0.095	0.238	0.595	0.952	1.190	2.381	0.116	0.290	0.725	1.160	1.449	2.899
	1%	0.125	0.312	0.780	1.249	1.561	3.121	0.149	0.373	0.932	1.492	1.864	3.729

The performance of L_n^k statistic is conducted using simulation method. To investigate the performance of L_n^1 statistic, the samples are generated from various samples n = [20, 100] from normal distribution, $x_2 \sim$ N(5,2) and von Mises distribution, $\theta \sim$ VM(π,2) with different values of σ in the range (0.2, 2). Using the generated data of x_2 and θ, the values of the response variable x_1 are obtained using (4) where the values of the JW model parameters β_0, β_2, γ and δ are the same as in the previous section. Then, the outlier is generated by altering,

$$x'_{1n} = x_{1n} + \Delta,$$

where $\Delta \geq 0$ is the contamination level. Next, the generated cylindrical data of x_1, x_2 and θ are fitted to JW circular-linear regression to find the estimates $\hat{\beta}_0$, $\hat{\beta}_2$, $\hat{\gamma}$ and $\hat{\delta}$. Then, we apply the L_n^1 statistic for the detection of outlier in each sample. If the value of the L_n^1 statistic is greater than the specified cut-off points, then we have correctly detected the outlier. The process is repeated for 2000 times and the values of P1, P3 and P5 are obtained.

The results for the samples when n = 20 and n = 100 are plotted in Figures 1 and 2, respectively. From both figures, the performance of L_n^1 statistic are almost similar. It can be seen that the performance of the L_n^1 statistic depends on the value of σ. The performance is better as the value of σ decreases. Hence, the performance is a decreasing function of σ. However, the performance is slightly better for n = 100 than for n = 20. When n is large, the distance between the residuals is expected to be shorter resulting in lower values of L_n^1 statistic as illustrated by the smaller cut-off points as shown in Table 1. Hence, when outlier occurs in large sample size, we detect the corresponding observation easier as its respective distance will be relatively longer compared to the case in smaller sample size.

We can see that the performance of P1 and P5 for different values of n shows similar behaviour. However, we note that, for large sample sizes, the curves are approaching 1 slightly faster. In addition, the differences between P1 and P3 are also approximately close to 0 but not shown here.

THE PERFORMANCE OF L_n^2 STATISTIC FOR TWO OUTLIERS

To investigate the performance of L_n^2 statistic, the samples are generated from various samples n = [20,100] from normal distribution, $x_2 \sim$ N(5,2) and von Mises distribution, $\theta \sim$VM(π,5) with different values of σ in the range (0.2, 2). The values of the response variable x_1 are obtained using the same method as for the L_n^1 statistic case. Then, the outlier is generated by altering,

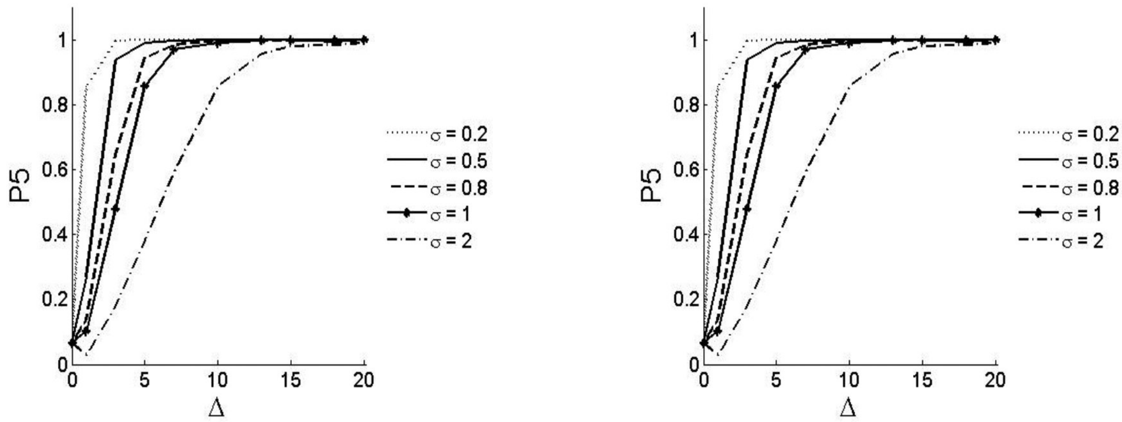

FIGURE 1. Sampling behaviour of the statistic for different values of σ when n = 20

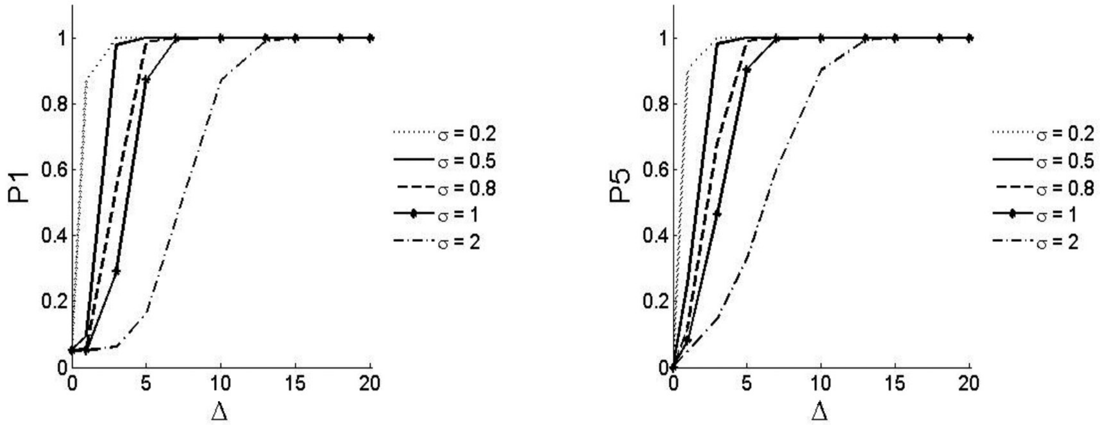

FIGURE 2. Sampling behaviour of the statistic for different values of σ when n = 100

$$x'_{1n} = x_{1n} + \Delta$$

$$x'_{1n-1} = x_{1n-1} + \Delta$$

where $\Delta \geq 0$ is the contamination level. Then, similar procedure as the performance of L_n^1 is used.

The performance of L_n^2 statistic when n = 50 and n = 100 are given in Figures 3 and 4, respectively. Generally, the performance of L_n^2 statistic shows a similar behaviour of L_n^1 statistic.

PRACTICAL EXAMPLE

We now apply the JW circular-linear regression model on a real data set. For a practical example, we use wind direction, wind speed (in m/s) and temperature (°C) data taken from the Malaysian Meteorological Department, which were measured in Bayan Lepas, Penang, in January 2005 at a pressure of 850 Hpa at 12:00 am. The data are given in Table 2.

The parameter estimates of the JW circular-linear regression model are $\hat{\beta}_0$ = 5.260, $\hat{\beta}_2$ = - 0.124, $\hat{\gamma}$ = 0.309 and $\hat{\delta}$ = 1.486 with the fitted is given by,

$$\hat{x}_1 = 0.526 - 0.124\, x_2 + 0.309 \cos\theta + 1.486 \sin\theta.$$

From Figures 5-7, it can be seen that there is an observation that is located far away from the rest of the data. This shows that a possible outlier is present in the data set. We apply the proposed outlier detection method for circular-linear regression using L_n^1 statistic.

The root mean squared error (RMSE) for this data set is 2.547 and value of test statistic L_{31}^1 = 8.143. Knowing that n = 31 and $\hat{\sigma}$ = 2.547, the corresponding cut-off point is 3.301. Clearly, the value of the L_{31}^1 statistic for observation 1 is greater than the cut-off point. Hence, the observation is identified as an outlier. We apply again the procedure on the reduced data set by removing observation 1 and no outlier is detected.

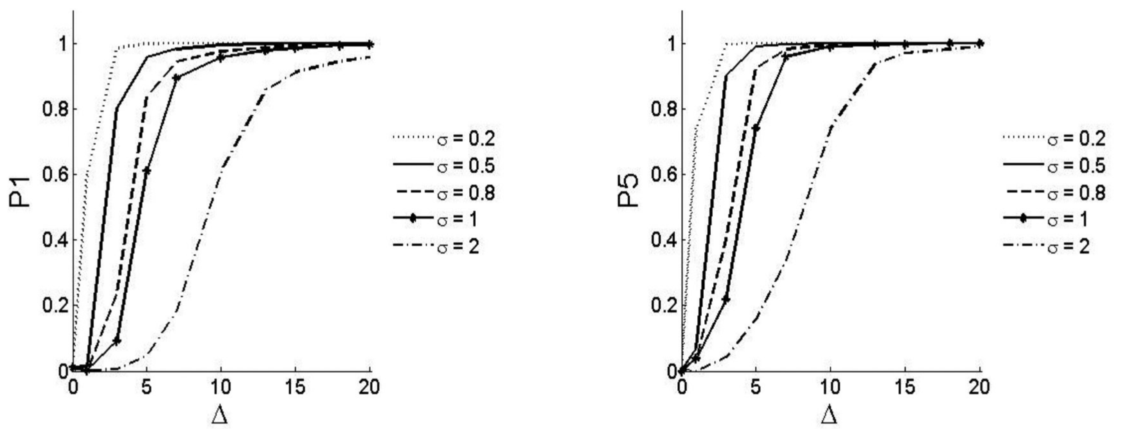

FIGURE 3. Sampling behaviour of the statistic for different values of σ when n = 20

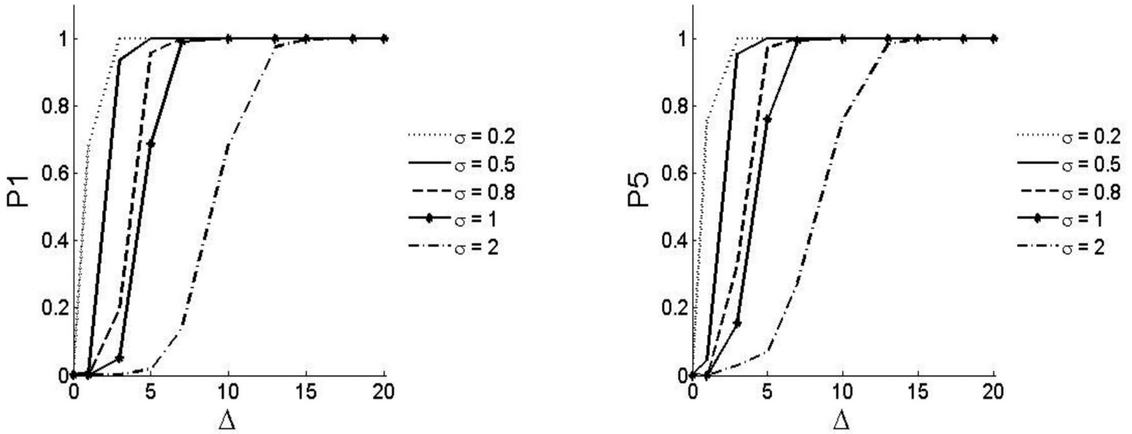

FIGURE 4. Sampling behaviour of the statistic for different values of σ when n = 100

TABLE 2. The wind data

Wind Speed (m/s)	Temperature (˚C)	Wind direction (˚)	Wind Speed (m/s)	Temperature (˚C)	Wind direction (˚)
14.9	17.6	85	2.1	19.0	125
5.1	18.0	85	1.5	17.8	185
4.6	18.2	140	1.0	17.4	190
6.2	18.0	100	0.5	17.2	70
3.6	18.2	135	4.6	16.6	135
1.5	17.6	310	2.6	17.2	125
2.1	18.4	340	3.6	18.2	90
4.6	18.2	120	2.1	16.6	200
4.6	17.6	130	3.6	18.0	5
5.1	17.4	120	2.6	17.2	30
4.6	19.0	150	3.1	17.2	165
2.6	17.6	80	3.1	18.6	260
1.0	18.4	205	4.6	18.0	325
0.5	18.6	60	3.6	17.4	325
5.1	18.4	110	2.6	17.8	345
3.1	18.0	125			

Source: Malaysian Meterological Department

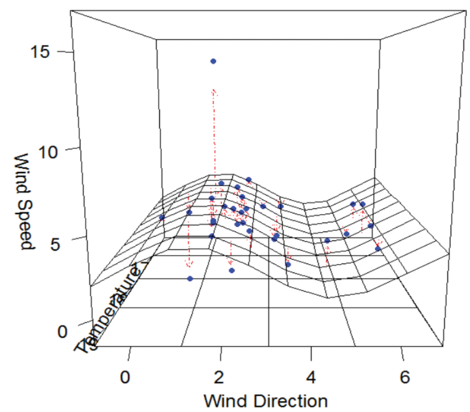

FIGURE 5. The regression plot of the wind data

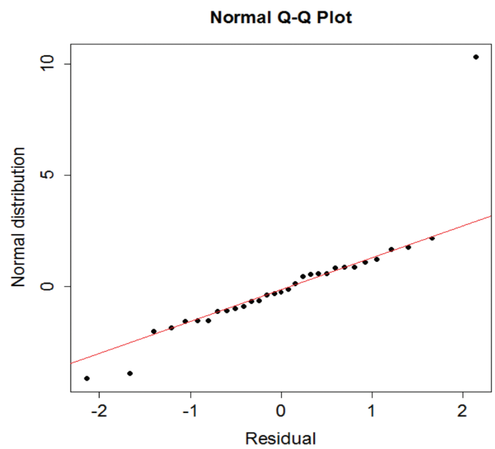

FIGURE 7. Q-Q normal plot of the residuals

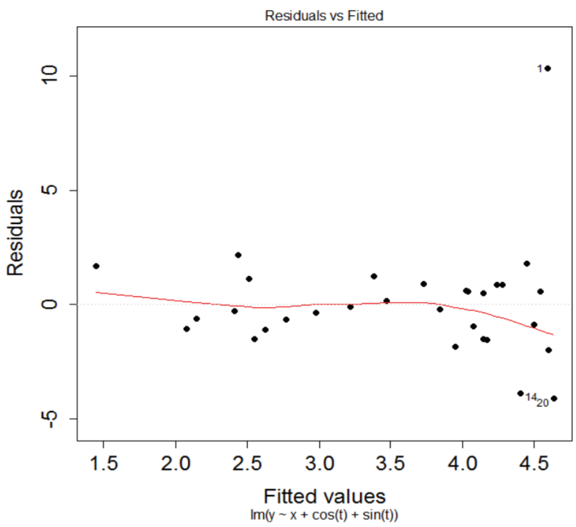

FIGURE 6. The residual vs fitted plot for wind data

TABLE 3. The summary of the effect of outlier removal from the wind data set

Parameters	Full data	Data after removing observation 1
$\hat{\beta}_0$	5.260	-0.138
$\hat{\beta}_2$	-0.124	0.170
$\hat{\gamma}$	0.309	-0.033
$\hat{\delta}$	1.486	0.807
$\hat{\sigma}$	2.547	1.509

The removal of observation 1 from the data set notably changes the values of $\hat{\beta}_0$, $\hat{\beta}_2$, $\hat{\gamma}$, $\hat{\delta}$ and $\hat{\sigma}$. The results are shown in Table 3 and Figures 8-10. Thus, the removal of observation 1 gives a better model fitting to the data set.

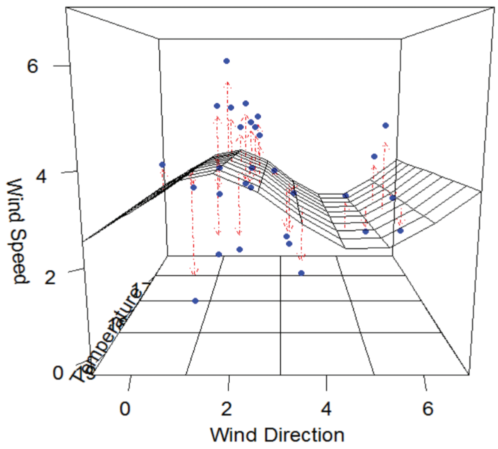

FIGURE 8. The regression plot of the wind data without observation 1

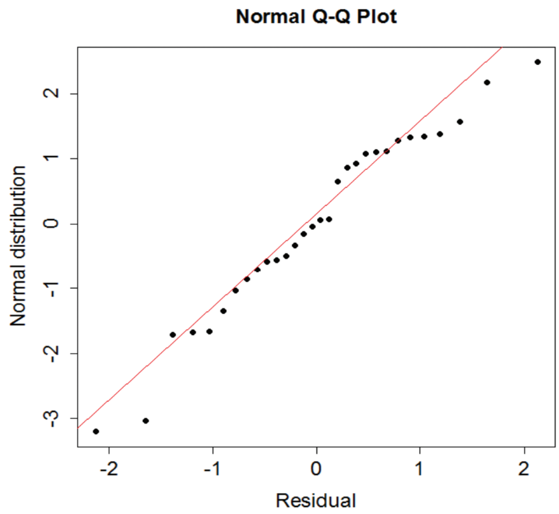

FIGURE 10. Q-Q normal plot of the residuals without observation 1

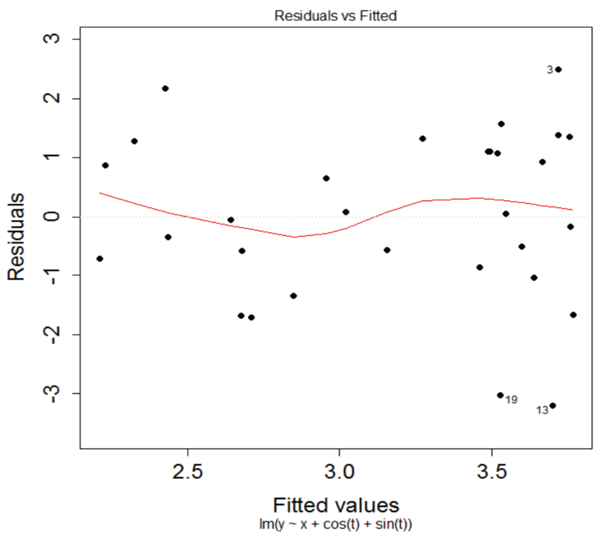

FIGURE 9. The residual vs fitted plot for wind data without observation 1

CONCLUSION

In this paper, we propose a new method of outlier detection in the JW circular-linear regression based on the k-nearest neighbour distance. The proposed test statistic performs well in detecting single and multiple outliers. Although we consider only the JW circular-linear regression model, the L_n^k statistic can be extended to other circular-linear regression models, with their corresponding simulated cut-off points.

ACKNOWLEDGEMENTS

This work was financially supported by the Ministry of Higher Education, Malaysia under the Fundamental Research Grant Scheme No: FP037-2014B.

REFERENCES

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

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

Barrett, B.E. & Ling, R.F. 1992. General classes of influence measures for multivariate regression. *Journal of the American Statistical Association* 87(417): 184-191.

Barnett, V. & Lewis, T. 1978. *Outliers in Statistical Data*. 2nd ed. New York: John Wiley & Sons.

Cook, R.D. 1977. Detection of influential observations in linear regression. *Econometrics* 19(1): 15-18.

David, H.A. 1981. *Order Statistic*. New York: Wiley.

Hadi, A.S. & Simonoff, J.S. 1993. Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association* 88(424): 1264-1272.

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

Jammalamadaka, S.R. & SenGupta, A. 2001. *Topics in Circular Statistics*. World Scientific.

Johnson, R.A. & Wehrly, T.E. 1978. Some angular-linear distributions and related regression models. *J. Am. Stat. Assoc*. 73(363): 602-606.

Qin, X., Jiang, S.Z. & Xiao, D.Y. 2011. A nonparametric circular-linear multivariate regression model with a rule-of-thumb bandwidth selector. *Computers and Mathematics with Applications* 62(8): 3048-3055.

Rambli, A., Ali, A., Mohamed, I. & Hussin, A.G. 2016. Procedure for detecting outliers in circular regression model. *PLoS ONE* 11(4): e0153074.

SenGupta, A. & Ugwuowo, F. 2006. Asymmetric circular-linear multivariate regression models with applications to environmental data. *Environmental and Ecological Statistics* 13(3): 299-309.

Srikantan, K.S. 1961. Testing for the single outlier in a regression models. *Sankhya Series A* 23(3): 251-260.

Srivastava, M.S. & Rosen, D. 1998. Outliers in multivariate regression models. *Journal of Multivariate Analysis* 65(2): 195-208.

Institute of Mathematical Sciences
University of Malaya
50603 Kuala Lumpur
Malaysia

*Corresponding author; email: adrianaibrahim@um.edu.my

Received: 12 May 2017
Accepted: 6 February 2018