Outlier Detection in a Circular Regression Model Using COVRATIO Statistic

S. Ibrahim, A. Rambli, A. G. Hussin & I. Mohamed

Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia
Centre for Defence Foundation Studies, National Defence University of Malaysia, Kuala Lumpur, Malaysia


To cite this article: S. Ibrahim, A. Rambli, A. G. Hussin & I. Mohamed (2013): Outlier Detection in a Circular Regression Model Using COVRATIO Statistic, Communications in Statistics - Simulation and Computation, 42:10, 2272-2280

To link to this article: http://dx.doi.org/10.1080/03610918.2012.697239

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Outlier Detection in a Circular Regression Model Using \textit{COVRATIO} Statistic

S. IBRAHIM,\textsuperscript{1} A. RAMBLI,\textsuperscript{1} A. G. HUSSIN,\textsuperscript{2} AND I. MOHAMED\textsuperscript{1}

\textsuperscript{1}Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia
\textsuperscript{2}Centre for Defence Foundation Studies, National Defence University of Malaysia, Kuala Lumpur, Malaysia

In this article, we model the relationship between two circular variables using the circular regression models, to be called JS circular regression model, which was proposed by Jammalamadaka and Sarma (1993). The model has many interesting properties and is sensitive enough to detect the occurrence of outliers. We focus our attention on the problem of identifying outliers in this model. In particular, we extend the use of the \textit{COVRATIO} statistic, which has been successfully used in the linear case for the same purpose, to the JS circular regression model via a row deletion approach. Through simulation studies, the cut-off points for the new procedure are obtained and its power of performance is investigated. It is found that the performance improves when the resulting residuals have small variance and when the sample size gets larger. An example of the application of the procedure is presented using a real dataset.

\textbf{Keywords} Circular regression; \textit{COVRATIO}; Influential observation; Outlier; von Mises distribution.

\textbf{Mathematics Subject Classification} Primary 62H11; Secondary 62J20.

1. Introduction

Circular data are known to occur in various fields including biology, meteorology, and medicine. They refer to a set of observations measured by angles and distributed within $[0, 2\pi]$ or $[0^\circ, 360^\circ]$. Due to this circular property, specialized statistical techniques are required in order to perform statistical analysis on the data. Such techniques include the study of the relationship between two or more circular variables.

Discussion on the development of circular regression models began with Gould (1969), who predicted the mean direction of a circular response variable from a vector of linear covariates. Mardia (1972) extended the model by assuming each of the response variables...
say, $\theta_j$, $j = 1, \ldots, n$ to be independently distributed from the von Mises distribution with mean direction $\mu_j$ and common concentration parameter $\kappa$. For the case when both response and explanatory variables, $u$ and $v$ respectively, are circular, a few circular regression models have been developed using different approaches. The earliest model, due to Laycock (1975), was expressed as a multiple linear regression model with complex entries. Downs and Mardia (2002) proposed a model based on a one-to-one mapping between the independent angle $u$ and the mean of the dependent angle $v$ such that the locus of the points $(u, v)$ is a continuous closed curve winding once around a toroidal surface. Other models include those found in Hussin et al. (2004) and Kato et al. (2008). Another interesting model was proposed by Jammalamadaka and Sarma (1993) who used the conditional expectation of the vector $e^iv$ given $u$, which is the true meaning of regression to represent the relationship between $u$ and $v$. The properties of the models for the case of a single explanatory variable have been studied (see also Sec. 8.6 of Jammalamadaka and SenGupta, 2001).

The analysis of circular regression modeling is also affected by the occurrence of outliers. Abuzaid et al. (2008) showed that the occurrence of outliers changes the parameter estimates of the linear regression for circular variables (Hussin et al., 2004) when fitted on a dataset involving wind direction. In the linear regression case, extensive studies on the problem of outliers can be found in the literature (Belsley et al., 1980; Barnett and Lewis, 1994; Montgomery and Peck, 1992). In addition, various outlier detection procedures have been successfully employed, including those based on the COVRATIO statistic. Since the procedure uses a row deletion approach, the identified outliers are usually known as influential observations. In this article, we extend the use of the COVRATIO statistic to circular regression models. The procedure is expected to work with any circular regression model, but we will focus on the model proposed by Jammalamadaka and Sarma (1993). In the rest of the article, we would refer the model as the JS circular regression model.

With that view in mind, this article is organized as follows: Section 2 reviews the JS circular regression model. Section 3 looks at the method of estimating the parameters using the least squares method and the covariance matrix of the JS circular regression model. Section 4 presents the definition of the COVRATIO statistic. In Section 5, a simulation study is conducted to investigate the sampling behavior of the statistic. We then study the power performance of the procedure for detecting influential observation in Section 6. Finally, in Section 7 we apply the procedure to the wind dataset given in Hussin et al. (2004).

2. The JS Circular Regression Model

Jammalamadaka and Sarma (1993) proposed a regression model for two circular random variables $U$ and $V$ in terms of the conditional expectation of $e^{iv}$ given $u$ such that

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

where $e^{iv} = \cos v + i \sin v$, $\mu(u)$ represents the conditional mean direction of $v$ given $u$ and $\rho(u)$ the conditional concentration parameter for some periodic function $g_1(u)$ and $g_2(u)$. Equivalently, we may write

$$E(\cos v | u) = g_1(u) \quad (2)$$
$$E(\sin v | u) = g_2(u).$$

COVRATIO Statistic for JS

2273
We may then predict \( v \) such that

\[
\mu(u) = \hat{v} = \arctan \left( \frac{g_2(u)}{g_1(u)} \right)
\]

\[
\begin{align*}
\frac{\tan^{-1} \frac{g_2(u)}{g_1(u)}}{\pi + \tan^{-1} \frac{g_2(u)}{g_1(u)}} & \quad \text{if } g_1(u) > 0 \\
\text{undefined} & \quad \text{if } g_1(u) = g_2(u) = 0.
\end{align*}
\]

The difficulty of non-parametrically estimating \( g_1(u) \) and \( g_2(u) \) leads us to approximate them by using suitable functions, taking into account the fact that they are both periodic with period \( 2\pi \). The approximations used are the trigonometric polynomials of suitable degree \( m \) of the form

\[
\begin{align*}
g_1(u) & \approx m \sum_{k=0}^{m} (A_k \cos ku + B_k \sin ku) \\
g_2(u) & \approx m \sum_{k=0}^{m} (C_k \cos ku + D_k \sin ku).
\end{align*}
\]

We therefore have the following models:

\[
\begin{align*}
\cos v & = \sum_{k=0}^{m} (A_k \cos ku + B_k \sin ku) + \varepsilon_1, \\
\sin v & = \sum_{k=0}^{m} (C_k \cos ku + D_k \sin ku) + \varepsilon_2
\end{align*}
\]

where \( \varepsilon = (\varepsilon_1, \varepsilon_2) \) is the vector of random errors following the normal distribution with mean vector 0 and unknown dispersion matrix \( \Sigma \). The parameters \( A_k, B_k, C_k, \) and \( D_k \), where \( k = 0, 1, \ldots, m \), the standard errors as well as the dispersion matrix \( \Sigma \) can then be estimated.

3. Estimation of the JS Circular Regression Parameters

Jammalamadaka and Sarma (1993) proposed the following generalized least squares estimation method for JS circular regression models: Let \((u_1, v_1), \ldots, (u_n, v_n)\) be a random circular sample of size \( n \). From Eq. (5), the observational equations can be written as

\[
\begin{align*}
V_{1j} & = \cos v_j = \sum_{k=0}^{m} (A_k \cos ku_j + B_k \sin ku_j) + \varepsilon_{1j} \\
V_{2j} & = \sin v_j = \sum_{k=0}^{m} (C_k \cos ku_j + D_k \sin ku_j) + \varepsilon_{2j},
\end{align*}
\]

for \( j = 1, \ldots, n \). Assume that \( B_0 = D_0 = 0 \) to ensure identifiability. The observational Eq. (6) can then be summarized as

\[
\begin{align*}
V^{(1)} & = (V_{11}, \ldots, V_{1n})' \\
V^{(2)} & = (V_{21}, \ldots, V_{2n})' \\
\varepsilon^{(1)} & = (\varepsilon_{11}, \ldots, \varepsilon_{1n})' \\
\varepsilon^{(2)} & = (\varepsilon_{21}, \ldots, \varepsilon_{2n})'
\end{align*}
\]
\[ U_{n \times (2m+1)} = \begin{bmatrix} 1 & \cos u_1 & \cdots & \cos mu_1 & \sin u_1 & \cdots & \sin mu_1 \\ 1 & \cos u_2 & \cdots & \cos mu_2 & \sin u_2 & \cdots & \sin mu_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos u_n & \cdots & \cos mu_n & \sin u_n & \cdots & \sin mu_n \end{bmatrix} \]

and
\[
\lambda^{(1)} = (A_0, A_1, \ldots, A_m, B_1, \ldots, B_m)' \]
\[
\lambda^{(2)} = (C_0, C_1, \ldots, C_m, D_1, \ldots, D_m)' .
\]

Consequently, they can now be written in the matrix form as
\[
V^{(1)} = U\lambda^{(1)} + \epsilon^{(1)}
\]
\[
V^{(2)} = U\lambda^{(2)} + \epsilon^{(2)}.
\]

The least squares estimates turn out to be
\[
\hat{\lambda}^{(1)} = (U'U)^{-1}U'V^{(1)}
\]
\[
\hat{\lambda}^{(2)} = (U'U)^{-1}U'V^{(2)}.
\]

Next, we want to estimate the covariance matrix \( \Sigma \) using the least squares theory. Let
\[
R_0(p,q) = V^{(p)}V^{(q)} - V^{(p)}U(U'U)^{-1}U'V^{(q)}
\]
where \( M = U(U'U)^{-1}U' \) and \( R_0 = (R_0(p,q))_{p,q=1,2} \). Then, it can be shown that
\[
\hat{\Sigma} = [n - 2(2m + 1)]^{-1}R_0
\]
is an unbiased estimate of \( \Sigma \). The standard errors of the estimators can then be found, see Jammalamadaka and Sarma (1993) for more details. The estimated covariance matrix \( \hat{\Sigma} \) will be used in the next section to identify possible influential observations in the data.

4. The COVRATIO Statistic

Belsley et al. (1980) used the row deletion approach to investigate the impact of deleting one row at a time on the estimated coefficients, fitted values, residuals and covariance matrix of linear regression models. Here, we are interested in the corresponding effects on the covariance matrix \( \hat{\Sigma} \) of the JS circular regression model. The procedure we develop looks at the effect of deleting one row from the circular regression data on the ratio of the estimated covariance matrix using all available observations over the estimated covariance matrix when the \( j \)th observation is deleted. We retain the name COVRATIO statistic for this ratio.

Let \(| \text{COV} | \) be the determinant of the covariance matrix for the full dataset, and \(| \text{COV}_{(-j)} | \) be the determinant of the covariance matrix for the dataset that excludes the \( j \)th row. Then the COVRATIO statistic is denoted and defined by
\[
\text{COVRATIO}_{(-j)} = \frac{| \text{COV} |}{| \text{COV}_{(-j)} |}.
\]

As in the linear case, we will use the \(| \text{COVRATIO}_{(-j)} | - 1 | \) as the test statistic of the detection procedure. Any observation with \(| \text{COVRATIO}_{(-j)} | - 1 | \) exceeding the cut-off point will be identified as an influential observation.
5. Cut-off Points of the Test Statistic

We carry out a simulation study to obtain the cut-off points of the test statistic for different sample sizes $n$ and different values of standard deviation $\sigma_1$ and $\sigma_2$. Specifically, we generate sets of random errors from the Normal distribution with mean vector $\theta$ for various combination of $(\sigma_1, \sigma_2)$ in the range of $[0.03, 0.3]$ and $n$ in the range $[10, 150]$. We consider the special case of JS Circular regression model when $m = 1$, so that six coefficients are to be estimated; $A_0, C_0, A_1, B_1, C_1$, and $D_1$.

For illustrative purposes, we consider a very simple model, where $v = a + u$. For example, when $a = 2$, we have $\cos (2 + u) = -0.4161 \cos u - 0.9093 \sin u$ and $\sin (2 + u) = 0.9093 \cos u - 0.4161 \sin u$. Then by comparing with Eq. (6), the true values of $A_1, B_1, C_1$, and $D_1$ are $-0.4161, -0.9093, 0.9093$, and $-0.4161$, respectively, with $A_0$ and $B_0$ being zero. Similarly, we can get different sets of true values by choosing different values of $a$. Here, we consider the values of $a = -3, -2, -1, 1, 2, 3$ and $6$. In order to assess the performance of the COVRATIO statistic, for this simple shift-model, we need to find the critical values.

The complete steps to obtain the cut-off points are described below:

Step 1. Generate a variable $U$ of size $n$ from $VM(\pi, 2)$.

Step 2. Generate $\varepsilon_1$ and $\varepsilon_2$ of size $n$ from $N\left(\begin{pmatrix} 0 \\ \sigma_1 \\ 0 \\ \sigma_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sigma_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \sigma_2 \end{pmatrix} \right)$. For a fixed $a$, obtain the true values of $A_1, B_1, C_1$, and $D_1$. The true values of $A_0$ and $C_0$ being zero here.

Step 3. Obtain the variable $V$ using Eq. (6).

Step 4. Fit the generated circular data to the JS circular regression model to give the parameter estimates $\hat{A}_0, \hat{A}_1, \hat{B}_1, \hat{C}_0, \hat{C}_1$, and $\hat{D}_1$.

Step 5. Calculate $|COV|$ from Eq. (8).

Step 6. Exclude the $j$th row from the generated sample, where $j = 1, \ldots, n$. For each $j$, repeat steps 3–5 for the reduced dataset to obtain $|COV(j)|$.

Step 7. Compute $COVRATIO(-j)$ and then obtain the value of $|COVRATIO(-j)|$ for each $j$.

Step 8. Specify the maximum value of $|COVRATIO(-j)|$.

The process is carried out 500 times for each combination of sample size $n$ and standard deviations $(\sigma_1, \sigma_2)$. Then the 5% upper percentiles of the maximum values of $|COVRATIO(-j)|$ 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 for $a = 2$. The result shows that, for fixed $\sigma_1$ and $\sigma_2 \geq \sigma_1$, the cut-off points show an increasing trend as $\sigma_2$ gets larger. The same trend is seen when $\sigma_2$ is fixed and $\sigma_1 \geq \sigma_2$. On the other hand, the cut-off points are a decreasing function of the sample size $n$. Similar results are observed for other values of $a$ and are not reported here. They can be obtained from the authors upon request.

6. The Power Performance of the COVRATIO Statistic

A simulation study is used to investigate the power performance of the $|COVRATIO(-j)|$ statistic. Four different sample sizes are considered, namely $n = 30, 50, 70,$ and $100$. The same procedure employed in Section 5 is used to generate the dataset. The observation at position $d$, say $v_d$, is then contaminated as follows:

$$v_d^* = v_d + \lambda \pi \text{ mod}(2\pi)$$
Table 1  
The 5% upper percentiles of the $|COVRATIO_{(-j)} - 1|$ statistic

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$0.05$</th>
<th>$0.08$</th>
<th>$0.1$</th>
<th>$0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.05</td>
<td></td>
<td>0.9203</td>
<td>0.9397</td>
<td>0.9379</td>
<td>0.9614</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td></td>
<td>0.9253</td>
<td>0.9340</td>
<td>0.9366</td>
<td>0.9579</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td></td>
<td>0.9301</td>
<td>0.9293</td>
<td>0.9416</td>
<td>0.9630</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td></td>
<td>0.9434</td>
<td>0.9382</td>
<td>0.9226</td>
<td>0.9534</td>
</tr>
<tr>
<td>30</td>
<td>0.05</td>
<td></td>
<td>0.5974</td>
<td>0.6818</td>
<td>0.6847</td>
<td>0.8000</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td></td>
<td>0.6213</td>
<td>0.6604</td>
<td>0.6662</td>
<td>0.7930</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td></td>
<td>0.6289</td>
<td>0.6262</td>
<td>0.6472</td>
<td>0.6856</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td></td>
<td>0.6850</td>
<td>0.6933</td>
<td>0.7131</td>
<td>0.7850</td>
</tr>
<tr>
<td>50</td>
<td>0.05</td>
<td></td>
<td>0.4388</td>
<td>0.4974</td>
<td>0.5120</td>
<td>0.6453</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td></td>
<td>0.4390</td>
<td>0.4678</td>
<td>0.4917</td>
<td>0.6226</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td></td>
<td>0.4517</td>
<td>0.4618</td>
<td>0.4706</td>
<td>0.6094</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td></td>
<td>0.4613</td>
<td>0.5694</td>
<td>0.5632</td>
<td>0.6184</td>
</tr>
<tr>
<td>80</td>
<td>0.05</td>
<td></td>
<td>0.3183</td>
<td>0.3456</td>
<td>0.3680</td>
<td>0.5132</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td></td>
<td>0.3373</td>
<td>0.3303</td>
<td>0.3428</td>
<td>0.4915</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td></td>
<td>0.3589</td>
<td>0.3418</td>
<td>0.3449</td>
<td>0.4882</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td></td>
<td>0.4414</td>
<td>0.4355</td>
<td>0.4331</td>
<td>0.4321</td>
</tr>
<tr>
<td>100</td>
<td>0.05</td>
<td></td>
<td>0.2731</td>
<td>0.3141</td>
<td>0.3361</td>
<td>0.5155</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td></td>
<td>0.2846</td>
<td>0.2891</td>
<td>0.3013</td>
<td>0.5051</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td></td>
<td>0.2930</td>
<td>0.2872</td>
<td>0.2961</td>
<td>0.4892</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td></td>
<td>0.3661</td>
<td>0.3603</td>
<td>0.3545</td>
<td>0.4106</td>
</tr>
<tr>
<td>130</td>
<td>0.05</td>
<td></td>
<td>0.2144</td>
<td>0.2440</td>
<td>0.2570</td>
<td>0.4529</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td></td>
<td>0.2186</td>
<td>0.2320</td>
<td>0.2414</td>
<td>0.4360</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td></td>
<td>0.2273</td>
<td>0.2288</td>
<td>0.2406</td>
<td>0.4144</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td></td>
<td>0.3450</td>
<td>0.3360</td>
<td>0.3265</td>
<td>0.3596</td>
</tr>
</tbody>
</table>

where $v_j^*$ is the value after contamination and $\lambda$ is the degree of contamination in the range $0 \leq \lambda \leq 1$. The generated data are then fitted to give the parameter estimates $\hat{A}_0, \hat{A}_1, \hat{B}_1, \hat{C}_0, \hat{C}_1$, and $\hat{D}_1$ and $|COV|$ is calculated using Eq. (8). Next, we exclude the $j$th row from sample for $j = 1, \ldots, n$ and refit the remaining data using Eq. (6). The $COVRATIO_{(-j)}$ is then calculated. If the values of $|COVRATIO_{(-j)} - 1|$ is maximum and greater than the specified cut-off point, then we say that the procedure has correctly detected the outlier in the data. The process is carried out 500 times. The power performance of the procedure is then examined by computing the percentage of the correct detection of the contaminated observations at position $d$.

We find that the power performance of the procedure is decreasing as $\sigma_1$ and $\sigma_2$ get smaller. This is expected as $V_{1j}$ and $V_{2j}$ in Eq. (6) will fluctuate closer to the horizontal axis when $\sigma_1$ and $\sigma_2$ are closer to zero, and hence, there is a better chance to detect the outlier even when $\lambda$ is small. On the other hand, we find that the performance is an increasing
function of $n$ and becomes similar when $n$ is large enough. Complete results are available from the authors.

7. Practical Example

We consider the wind direction data given in Hussin et al. (2004). A total of 129 observations of wind direction (in radians) are recorded over a period of 22.7 days along the Holderness coastline (the Humberside coast of the North Sea, United Kingdom) by using two different instruments, i.e., an HF radar system and an anchored wave buoy.

The spoke plot of the wind direction data is shown in Fig. 1. The correlation value is 0.8317, which implies a positive and strong correlation between the two instruments. Only two pairs of observations result in the straight lines crossing the inner circle of the plot. One line cuts diametrically across the inner circle (observation 38) while the other only cuts it in a short chord (observation 111). These are candidates for influential observations.

We fit the JS regression model on the data. The parameter estimates are

\[ \hat{A_0} = 0.0674, \quad \hat{A_1} = 0.7559, \quad \hat{B_1} = -0.0948, \quad \hat{C_0} = -0.047, \]
\[ \hat{C_1} = -0.1049, \quad \hat{D_1} = 0.9762, \quad \hat{\sigma_1} = 0.0818, \quad \hat{\sigma_2} = 0.053 \]

and thus the fitted model gives $g_1(u)$ and $g_2(u)$ to be

\[ \hat{g}_1(u) = 0.0674 + 0.7559 \cos u - 0.0948 \sin u \]
\[ \hat{g}_2(u) = -0.047 + 0.1049 \cos u + 0.9762 \sin u. \]

![Figure 1. The spoke plot of the wind direction data. (Color figure available online).](image-url)
Thus, model $E(e^{i\hat{\nu}|u}) = \rho(u)e^{i\hat{\mu}(u)} = g_1(u) + ig_2(u)$ is obtained such that

$$
\mu(u) = \hat{v}_j = \arctan\left(-\frac{0.047 + 0.1049 \cos u_j + 0.9762 \sin u_j}{0.0674 + 0.7559 \cos u_j - 0.0948 \sin u_j}\right), \quad j = 1, \ldots, n,
$$

$$
\hat{\rho} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \hat{\rho}^2(u_j)} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left[g_1^2(u_j) + g_2^2(u_j)\right]} = 0.9322
$$

Now, we apply the COVRATIO statistic to detect any possible influential observations in the wind direction data. The determinant of the covariance matrix for the full dataset $|COV|$ is 0.0043 and the corresponding values of $|COVRATIO(-j) - 1|$ are then calculated. From simulated cut-off values for this context, we find observation 38 is highly significant. It is further illustrated by Fig. 2, which shows the value for observation number 38 is different from the others.

Thus the proposed COVRATIO statistic for the JS circular regression model has successfully identified only one observation as an influential observation compared to two in Abuzaied et al. (2008). The Akaike information criteria (AICC) formula, defined in Lund (1999), gives the AICC value for the JS circular regression model as $-369.32$, which is lower than that of the simple linear circular regression model considered in Abuzaied et al. (2008), which is $-362.44$. Therefore, the JS circular regression model provides a better fit to the data and has included observation 111 as part of the underlying structure of the data.
model. This is further supported by considering the spoke plot in Fig. 1, where the line for observation 111 only slightly cuts the inner circle.

8. Conclusions
In this article, we developed a procedure based on the COVRATIO statistic to identify influential observations in JS circular regression models. The cut-off points are obtained and the power performance examined through simulation studies for a simple model. We show that the sample size and dispersion of the residuals determine the level of cut-off points. The procedure also shows a good performance in identifying influential observations in JS circular regression models.

Acknowledgments
The authors are most grateful to the Associate Editor and referees for their thorough reading and valuable suggestions, which led to a substantial improvement of the article.

References