Journal of Economics and Administrative Sciences Vol.28 (NO. 134) 2022, pp. 168-185

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

Comparison of Robust Circular S and Circular Least Squares Estimators for Circular Regression Model using Simulation

Huda Hadib Abbas
Post-Graduate Institute For Accounting& Financial Studies,
University of Baghdad
Hoda.Hadib201a@coadec.uobaghdad.edu.iq

Suhail Najim Abood
Department of Statistics,
College of Administration and Economics,
University of Baghdad
suhnaj2005@coadec.uobaghdad.edu.iq

Received:1/12/2022 **Accepted:13/12/2022** **Published: December / 2022**

Abstract

In this paper, the Monte-Carlo simulation method was used to comparing the robust circular S estimator with the circular Least squares method in the case of no outlier data and in the case of the presence of outlier in the data through two trends, the first is contaminant with high inflection points that represents contaminant in the circular independent variable, and the second the contaminant in the vertical variable that represents the circular dependent variable using three comparison criteria, the median standard error (Median SE), the median of the mean squares of error (Median MSE), and the median of the mean cosines of the circular residuals (Median A(k)). It was concluded that the method of least squares is better than the methods of the robust circular S method in the case that the data does not contain outlier values because that it was recorded the lowest mean criterion, mean squares error (Median MSE), the least median standard error (Median SE) and the largest value of the criterion of the mean cosines of the circular residuals A(K) for all proposed sample sizes (n=20, 50, 100). In the case of contaminant in the vertical data, it was found that the circular least squares method is not preferred at all contaminant rates and for all sample sizes, and the higher the percentage of contamination in the vertical data, the greater the preference of the validity of estimation methods, where the mean criterion of median squares of error (Median MSE) and criterion of median standard error (Median SE) decrease and the value of the mean criterion of the mean cosines of the circular residuals A(K) increases for all proposed sample sizes. In the case of contaminant at high lifting points, the circular least squares method is not preferred by a large percentage at all levels of contaminant and for all sample sizes, and that the higher the percentage of contaminant at the lifting points, the greater

the preference of the validity estimation methods, so that the mean criterion of mean squares of error (Median MSE) and criterion of median standard error (Median SE) decrease, and the value of the mean criterion increases for the mean cosines of the circular residuals A(K) and for all sample sizes.

Key words: Robustness, Circular Regression, S Estimator, Least Squares Estimator, Circular Statistics, Circular Data, and Outlier Observations.

1. Introduction

That is what occupies the mind of researchers in all fields of knowledge is the nature of the data. It is very rare for these data to be directly prepared for the use of appropriate statistical methods for estimation. When a case of violation of one of the conditions required for estimation or inaccuracy in the data occurs, it requires searching for different methods to treat those cases. Among the data that we encounter in many practical applications are the data that are measured in the form of angles in degrees or radians, such as the angles of bone fracture, the angles of curvature of the cornea, and others. The applications to circular variables have increased in the past two decades and diversified into many fields including biology, meteorology, and medicine. Although the first circular regression model dates back to (Gould, 1969) and after that, different types of these models have been proposed; however, outliers and the robust of circular regression models are still poorly studied. It used the methods of detecting outliers based on the simple circular regression model by extending the common methods of linear regression. In some cases, when performing the statistical analysis of a set of data, it may occur that some observations deviate or deviate from the largest amount of observations that exist with them, which are called "outliers" and, if they are included in the data, the traditional estimations fail to give estimates accurate information about the parameters of the statistical population from which those data were withdrawn due to breaching its basic conditions, so the use of statistical analyzes for data directions, ocean current directions, and the direction of bone fracture level. It is not possible to containing outlier values is a real problem and should be avoided. At the same time, we encounter data recorded in degrees or radial angles in different sets of scientific research, including daily wind, ocean current directions and the direction of the bone fracture level that cannot use the ordinary regression models with it, so it is necessary to use a model applied to these circular data. Various versions of these circular models have been proposed. But still, the problem of studying outliers and the robust circular regression model has not been taken into account so far, which is a real problem when we face the problem of the presence of outliers within the studied data. (Goud, 1969) predicted the directional mean of a circular response variable and a set of linear independent variables, and after it was proposed, many circular regression models (Alshqaq, et al, 2021, 1) and (Jammalamadaka & Sarma, 1993) were able to put forward the concept of circular regression for data representing directions measured as angles in a plane with reference to a constant sense of rotation and a fixed direction of zero, where they discussed some experimental methods and approximate tests to determine the degree of the regression equation and developed some numerical algorithms to find circular regression coefficients. In (2008) (Mohamed) and others used the Maximum Likelihood method to estimate the coefficients of the simple linear regression model and the multiple linear regression model when the variables are

circular with a random error that has a Von-Mises distribution with zero mean and k variance. In (2013) A.H. Abuzaid and others proposed a new numerical statistic called mean circular error. In the same year (2013), (A. Ibrahim et al.) used the method of least squares (OLS) to estimate the parameters of the circular regression model of JS depending on the COVARATIO statistic and row-deletion method to identify and detect outliers in the JS model. In (2015) (Rambli et al.) used the COVARATIO statistics to identify and detect outliers in the Circular Regression Model proposed by (Downs & Mardia, 2002) and the performance of the method was evaluated by Monte-Carrois simulation experiments. In (2017) (A. Mahmood et al.) proposed a new statistic to determine the multiple outliers in the simple circular regression model based on calculating the circular robust distance between the circular residuals and the circular position parameter; the performance of the proposed method was evaluated using Monte-Carlo simulation. In the same year (2017), (Jha & Biswas) proposed the Robust Maximum Trimmed estimator RMTE, which was based on the cosine distance to estimate the Circular-Circular Regression Model proposed by before (Kato et al., 2008). In (2019), (Alkasadi et al.) identified the outliers in the Multiple Circular Regression Model (MCRM) that was proposed by (Ibrahim, 2013), which studies the relationship between two or more circular variables, as they extended the DFFITS statistic that was proposed by (Belsley, 1980) and tested the method using Monte Carlo simulation experiments. In (2020), (A. Mahmood et al.) suggested robust methods for estimating the parameters of the Circular Regression Model, which are the Maximum Weighted Likelihood Estimation MWLE method and the Maximum Trimmed Likelihood Estimation MTLE method. The two robust methods were compared using Monte-Carlo simulation using the statistical criteria of mean square error (MSE) for estimators, variance and amount of bias. In the year (2021), (Alshqaq et al.) expanded the M-estimation, Least-Trimmed squares LTS and Least-median squares LMS estimators from the traditional linear regression model to the JS Circular Regression Model, which is characterized by its good qualities and sensitivity in detecting outliers. This paper aims to estimate the JS Circular regression model by suing the circular Least Square estimation without outliers is data and comparing it with S circular robust estimation with outlier's data. This paper is the first article that deals with robustness of robust methods.

2. Materials and Methods

2.1 Circular Statistics (Pewsey et al., 2013, 16)

The term circular statistics refers to a specific branch of statistics that deals with data that can be represented as points on the circumference of the unit circle. This type of data is called circular data. The term circular data is used to distinguish it from linear data, which are often used in analytics. Since the support for circular data is the unit circle, while for linear data, the support is the real number line. Circular data enters various disciplines such as biology, medicine, image analysis, earth sciences, physics, political studies, and astronomy. In two-dimensional space, any point can be represented by either its Cartesian coordinates as (v, u) or with its polar coordinates as (r, θ), where r is the distance from the circumference of the circle to the origin. In circular analysis, only the direction is focused, so the vector r is considered to be of unit length (i.e. r = 1). Therefore, any point on the circle can be represented as (Sin (θ), Cos (θ)). Examples of circular data are directions measured using instruments such as a compass, protractor,

weather vane, or compass. It is usual to record such directions in degrees (degrees) or radians either clockwise or counterclockwise from the point of origin of the circle, which is referred to as the zero direction. What is required is to determine the location of the point and the direction and not, as in the data, on the real line; values to the left of the origin (0) are negative and values to the right are positive. For circular data, each angle is defined as a point on the circumference of the unit circle, just as each value of a linear variable is defined as a point on the real line. And as the absolute value of the linear variable increases, we move away from the origin. Therefore, in the real line, the value 360 is relatively close to the value 355 but is relatively far from the origin. But for cyclic variables, the angle 355 degrees corresponds to a point on the circumference of the unit circle close to that corresponding to 360 degrees, the angles 0 and 360 degrees define the exact same point. It is this cyclical nature of circular data that forces us to abandon traditional statistical methods designed for linear data and make us look for those that serve circular data and take into account the structure of this data.

2.2 Circular Data (G Pramesti, 2018, 2-3)

Most of measurements in any field are represented by a real number, but in fact in many variety fields any observation can be measured as a direction, for example wind direction, the direction of migratory birds as circular data that can be measured by a compass or clock, and the age of universe can also be considered, directions of living organisms and the direction of pollutants are considered as directional observations, and such data is referred to as directional data, where the direction can be represented as points on the circumference of the unit circle or unit vectors that connect the point of origin to these points; thus, the two-dimensional data is called circular data, and the observations in three dimensions are called spherical data. Circular data can be represented by the angle θ whose range is [0, 2π] or [-π, π]. The angle θ is periodic since =θ+2kπθ. In addition to the circular data, the random variable that has a value on the unit semicircle and which has the domain [-π/2, π/2] or [0, π] is called the axial data or the semi-circular data.

2.3 Outliers

Outlier data means the presence and interaction of different random mechanisms one of which produces the majority of the data, and the other is responsible for the occurrence of outliers. Even if we can model the majority of the data, it is not clear how to select trends for deviations from model assumptions caused by outliers. In this sense, the outlier in the data is assumed to be unexpected values, and the analysis cannot detect their random generation process. (Barnett and Lewis, 1994), (Davies and Gather, 1993), (Markatou et al.1998), and (Gather et al. , 2003) define the outliers are values taken from phenomena that receive a small probability under the assumed model (Farcomeni & Gareco, 2015, 4). Outliers or contaminants are defined as arbitrary points in nature that represent an observation or set of observations outside the normal pattern of the data set. (God's gift: 2005, 4) and that they are data points that are far away from the majority of other data points; they are observations that are not consistent with the rest of the set's data for any of the variables for a particular phenomenon or for a set of phenomena, the value of this observation may be large or it may be small. It is located at one end of the set of observations arranged ascending or descending, and that its outliers may in many cases be a natural issue associated with some

variables (Al-Yasiri: 2007, 6). It is generated in a different way from the method of generating the original data. These are the observations that appear inconsistent with the rest of the data set (Obikee & et. al., 2014, 536). The researchers (Rousseeum & Leroy) in 1987 defined it as observations in the regression that deviate from the bulk of the data (Hekimoglu & Erenoglu, 2013, 421), while researchers (Barnett & Lewis) in (1994) defined the outliers as (observations that appear inconsistent with the rest of the data set).

2.4 Classes of Outlier Observations: (Lukman et al., 2015, 55)

In (2015), researchers (Lukman, et al.) classified the outliers into three categories of outliers in the V-axis (vertical or vertical outliers) and outliers values in both directions (axis of the explanatory variable, axis of the dependent variable V). Anomalies in the axis of the explanatory variable U are called High Leverage Points (HLPs), and they are called by this name because they bend or tilt the regression model line towards it and thus cause more serious problems than the vertical anomalies in the V axis.

2.5 Jammalamadaka & Sarma Circular regression model (JS Model)

(Jammalamadaka and Sarma, 1993) proposed a circular regression model for two circular random variables U and V to predict V by U in the context of the conditional prediction of the vector e^{iv} given u as follows:

$$\begin{aligned} E(e^{iv}|u) &= \rho(u)e^{i\mu(u)} \\ &= g_1(u) + ig_2(u), \qquad \dots (1) \end{aligned}$$

where:

$$e^{iv} = \cos(v) + i\sin(v) \qquad \dots (2)$$

$\mu(u)$ Represents the directional conditional mean of v given u,
$\rho(u)$. The conditional concentration parameter of the periodic functions, $g_1(u)$, $g_2(u)$ which is written as:

$$E(\cos(v|u)) = g_1(u) \qquad \dots (3)$$

$$E(\sin(v|u)) = g_2(u) \qquad \dots (4)$$

Then v can be predicted as:

$$\begin{aligned} \mu(u) &= \hat{v} \\ &= \arctan * \frac{g_2(u)}{g_1(u)} \\ &= \begin{cases} \arctan\frac{g_2(u)}{g_1(u)} & \text{if } g_1(u) > 0 \\ \pi + \arctan\frac{g_2(u)}{g_1(u)} & \text{if } g_1(u) < 0 \\ \text{undefined} & \text{if } g_1(u) = g_2(u) = 0 \end{cases} \qquad \dots (5) \end{aligned}$$

Considering that $g_1(u) = \cos(v_j)$ and $g_1(u) = \sin(v_j)$ are periodic functions within the interval 2π. Therefore, these functions are approximated using suitable functions, so the approximation will be using a trigonometric polynomial of a suitable degree m, according to two models similar to the regression model:

$$\begin{aligned} V_{1j} &= g_1(u) \\ &= \cos(v_j) \simeq \sum_{k=0}^{m}\left(A_k\cos(ku_j) + B_k\sin(ku_j)\right) + \varepsilon_{1j} \qquad \dots (6) \end{aligned}$$

$$\begin{aligned} V_{2j} &= g_2(u) \\ &= \sin(v_j) \simeq \sum_{k=0}^{m}\left(C_k\cos(ku_j) + D_k\sin(ku_j)\right) + \varepsilon_{2j} \qquad \dots (7) \end{aligned}$$

For j=1,….,n , $\varepsilon = (\varepsilon_1, \varepsilon_2)$ is a random error vector that has a bivariate normal distribution with a mean vector equal to zero and an unknown covariance matrix Σ, and that (A_k, B_k, C_k, D_k) are the parameters of the two models, as $k = 0,1,2,\dots,m$. Standard errors and unknown variance matrix Σ can be estimated assuming that $B_0 = D_0 = 0$ to ensure that the model can be defined.
(Alshqaq2, et al., 2021, 2) (S. IBRAHIM et al., 2013), (Jammalamadaka and Sarma, 1993).
If k=0, then:

$$V_{1j} = g_1(u) = \cos v_j \simeq A_0 \cos 0u_j + B_0 \sin 0u_j + \varepsilon_{1j} \simeq A_0 + \varepsilon_{1j} \quad \dots (8)$$

$$V_{2j} = g_2(u) = \sin v_j \simeq C_0 \cos 0u_j + D_0 \sin 0u_j + \varepsilon_{2j} \simeq C_0 + \varepsilon_{2j} \quad \dots (9)$$

If k=1, then:

$$V_{1j} = g_1(u) = \cos v_j \simeq A_0 \cos 0u_j + B_0 \sin 0u_j + A_1 \cos 1u_j + B_1 \sin 1u_j + \varepsilon_{1j} \simeq A_0 + A_1 \cos u_j + B_1 \sin u_j + \varepsilon_{1j} \quad \dots (10)$$

$$V_{2j} = g_2(u) = \sin v_j \simeq C_0 \cos 0u_j + D_0 \sin 0u_j + C_1 \cos 1u_j + D_1 \sin 1u_j + \varepsilon_{2j} \simeq C_0 + C_1 \cos u_j + D_1 \sin u_j + \varepsilon_{2j} \quad \dots (11)$$

If k=2, then:

$$V_{1j} = g_1(u) = \cos v_j \simeq A_0 \cos 0u_j + B_0 \sin 0u_j + A_1 \cos 1u_j + B_1 \sin 1u_j + A_2 \cos 2u_j + B_2 \sin 2u_j + \varepsilon_{1j}$$
$$\simeq A_0 + A_1 \cos u_j + B_1 \sin u_j + A_2 \cos 2u_j + B_2 \sin 2u_j + \varepsilon_{1j} \quad \dots (12)$$

$$V_{2j} = g_2(u) = \sin v_j \simeq C_0 \cos 0u_j + D_0 \sin 0u_j + C_1 \cos 1u_j + D_1 \sin 1u_j + C_2 \cos 2u_j + D_1 \sin 2u_j + \varepsilon_{2j}$$
$$\simeq C_0 + C_1 \cos u_j + D_1 \sin u_j + C_2 \cos 2u_j + D_2 \sin 2u_j + \varepsilon_{2j} \quad \dots (13)$$

For j=1, 2, …, n & k=1 then:

$$V_{11} = g_1(u) = \cos v_1 = A_0 + A_1 \cos u_1 + B_1 \sin u_1 + \varepsilon_{11}$$

$$V_{12} = g_1(u) = \cos v_2 = A_0 + A_1 \cos u_2 + B_1 \sin u_2 + \varepsilon_{12}$$

$$\vdots$$

$$V_{1n} = g_1(u) = \cos v_n = A_0 + A_1 \cos u_n + B_1 \sin u_n + \varepsilon_{1n}$$

And:

$$V_{21} = g_2(u) = \sin v_1$$

$$= C_0 + C_1 \cos u_j + D_1 \sin u_j + \varepsilon_{2j}$$
$$V_{22} = g_2(u)$$
$$= \sin v_2$$
$$= C_0 + C_1 \cos u_2 + D_1 \sin u_2 + \varepsilon_{22}$$
$$\vdots$$
$$V_{2n} = g_2(u)$$
$$= \sin v_n$$
$$= C_0 + C_1 \cos u_n + D_1 \sin u_n + \varepsilon_{2n}$$

2.6 Circular Least Squares Estimation

Let $(u_1, v_1), \ldots, (u_n, v_n)$ a circular random sample of size n as follows:

$$V^{(1)} = (V_{11}, \ldots, V_{1n})'$$
$$V^{(2)} = (V_{21}, \ldots, V_{2n})'$$
$$\varepsilon^{(1)} = (\varepsilon_{11}, \ldots, \varepsilon_{1n})'$$
$$\varepsilon^{(2)} = (\varepsilon_{21}, \ldots, \varepsilon_{2n})'$$

$$U_{nx(2m+1)} = \begin{bmatrix} 1 & \cos u_1 & \cos mu_1 & \sin u_1 & \vdots & \sin mu_1 \\ 1 & \cos u_2 & \cos mu_2 & \sin u_2 & \vdots & \sin mu_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos u_n & \cos mu_n & \sin u_n & \vdots & \sin mu_n \end{bmatrix} \quad \ldots (14)$$

For the purpose of simplification, the parameter vector of models (6) and (7) will be named as follows:

$$\lambda^{(1)} = (A_0, A_1, \ldots, A_m, B_1, \ldots \ldots . B_m)' \quad \ldots (15)$$
$$\lambda^{(2)} = (C_0, C_1, \ldots, C_m, D_1, \ldots \ldots . D_m)' \quad \ldots (16)$$

Therefore, equations (15) and (16) can be written in matrix form as follows:

$$V^{(1)} = U\lambda^{(1)} + \varepsilon^{(1)} \quad \ldots (17)$$
$$V^{(2)} = U\lambda^{(2)} + \varepsilon^{(2)} \quad \ldots (18)$$

The least squares estimates are as follows:

$$\hat{\lambda}^{(1)} = \min \sum_{i=1}^{n} (V_j^{(1)} - U\lambda^{(1)})^2 \quad \ldots (19)$$
$$\hat{\lambda}^{(2)} = \min \sum_{j=1}^{n} (V_j^{(2)} - U\lambda^{(2)})^2 \quad \ldots (20)$$

From equations (19) and (20), we get the estimates of circular least squares by deriving the two formulas relative to the parameters to be estimated and equating the first derivative to zero, the following results:

$$\hat{\lambda}^{(1)} = (U'U)^{-1} U' V^{(1)} \quad \ldots (21)$$
$$\hat{\lambda}^{(2)} = (U'U)^{-1} U' V^{(2)} \quad \ldots (22)$$

2.7 Robustness in JS model

The problem of the presence of outliers is one of the common problems in statistical analysis, which is defined as observations that differ greatly from the rest of the observations in the data set. (Ibrahim et al., 2013) found robustness in the JS regression model and concluded that the JS model is sensitive to the presence of anomalies and is likely to have effective effects on the least squares estimates of the JS model. (Ibrahim et al. 2013, 2275). Therefore, robustness can be studied in the circular regression model JS from two directions:

i. Circular Vertical Outliers

They are the outliers in the dependent variable V, so if V_{1j} is replaced by $V^*_{1j} = Z_1 V_{1j}$ and V_{2j} with $V^*_{2j} = Z_2 V_{2j}$ which leads to:

$$V_{1j} = Z_1^{-1} V^*_{1j} \qquad \dots (23)$$

$$V_{2j} = Z_2^{-1} V^*_{2j} \qquad \dots (24)$$

The circular regression in equation (1) can be written as:

$$Z_1^{-1} V^*_{1j} = \cos(v_j) = \sum_{k=0}^{m} \left(A_k \cos(ku_j) + B_k \sin(ku_j)\right) + \varepsilon_{1j} \qquad \dots (25)$$

$$Z_2^{-1} V^*_{2j} = \sin(v_j) = \sum_{k=0}^{m} \left(D_k \cos(ku_j) + C_k \sin(ku_j)\right) + \varepsilon_{2j} \qquad \dots (26)$$

The estimators of squares in the presence of outliers can be obtained as follows:

$$\hat{\lambda}^{(1)}(U, V^*_{1j}) = Z_1^{-1} \hat{\lambda}^{(1)}(U, V_{1j}) \qquad \dots (27)$$

$$\hat{\lambda}^{(2)}(U, V^*_{2j}) = Z_2^{-1} \hat{\lambda}^{(2)}(U, V_{2j}) \qquad \dots (28)$$

ii. Circular Leverage Points

They are the outliers in the independent variable U. If U is replaced by $U^* = ZU$, then:

$$\hat{\lambda}^{(1)}(U^*, V_{1j}) = Z_1^{-1} \hat{\lambda}^{(1)}(U, V_{1j}) \qquad \dots (29)$$

$$\hat{\lambda}^{(2)}(U^*, V_{2j}) = Z_1^{-1} \hat{\lambda}^{(1)}(U, V_{2j}) \qquad \dots (30)$$

(Alshqaq2, et al., 2021,3) (Alshqaq, 2021, 2-3)

2.8 Circular Robust S-Estimator

The researcher (Yohal, 1987) suggested estimations related to the measurement lab ($\hat{\sigma}_{MAD}$) for estimators (M). These estimations were called S estimators. The estimation process under this method is based on the standard residuals of the M method, and it has the same approximate properties as the M estimators. In this method, the standard deviation of the residuals is used to overcome the weakness resulting from the use of the median. As estimates of the coefficients of the circular regression model achieve the following objective function:

$$\hat{\hat{\lambda}}_S = \min_{\hat{\lambda}^{(p)}} \hat{\sigma}_{MAD}\left(r_1(\hat{\lambda}_{(p)s}), \dots, r_n(\hat{\lambda}_{(p)s})\right) \; ; i = 1, 2, \dots, n \qquad \dots (31)$$

Determining the least estimated fortuitous value of the residual standard deviation ($\hat{\sigma}_{MAD}$) that fulfills the objective function:

$$F = \min\left(\sum_{i=1}^{n} \rho \frac{(V_i - U\lambda^{(p)})}{\hat{\sigma}_s}\right) \quad ; p = 1, 2 \qquad \dots (32)$$

where:

$$\hat{\sigma}_s = \sqrt{\frac{1}{n\tau} \sum_{i=1}^{n} w_i . r^2{}_n(\hat{\lambda}_{(p)s})} \quad \text{for iteration} > 1 \qquad \dots (33)$$

where $\tau = 0.199$

$$w_i = w_\sigma(u_i) = \frac{\rho(u_i)}{u_i} \qquad \dots (34)$$

An initial value of the standard deviation of the residuals is calculated according to the following formula:

$$\hat{\sigma}_s = \frac{\text{median}|r_i(\hat{\lambda}_{(p)s}) - \text{median} r_i(\hat{\lambda}_{(p)s})|}{0.6745} \quad \dots (35)$$

And the solution to the objective function by equation (41) i.e. obtaining s estimates of the coefficients of the circular regression model, we get it through partial differentiation with respect to those parameters and equalizing them to zero as follows:

$$\sum_{i=1}^{n} \psi U \frac{(V_i - U\lambda^{(p)})}{\hat{\sigma}_s} = 0 \ , p=1, 2 \quad \dots (36)$$

ince ψ is the derivative of the function ρ and defined by the formula:

$$\psi = \begin{cases} \frac{r_i(\hat{\lambda}_{(p)s})}{\sigma_s}\left(1 - \left(\frac{\frac{r_i(\hat{\lambda}_{(p)s})}{\sigma_s}}{c}\right)\right)^2 & , |r_i(\hat{\lambda}_{(p)s})| \le c \ \text{Tukey's Bisquare} \\ 0 & |r_i(\hat{\lambda}_{(p)s})| > c \end{cases} \quad \dots (37)$$

$$\psi = \begin{cases} \frac{r_i(\hat{\lambda}_{(p)s})}{\sigma_s} & |r_i(\hat{\lambda}_{(p)s})| < c \\ c \sin\left(\frac{r_i(\hat{\lambda}_{(p)s})}{\sigma_s}\right) & |r_i(\hat{\lambda}_{(p)s})| \ge c \end{cases} \ , \ \text{Huber} \quad \dots (38)$$

To find the values of the estimated parameters from the set of equations (29) and (30), the IRLS method is based on the weights (w_i) shown in the following formula, which represents the Tukey's Bisquare function:

$$w_i = \begin{cases} \begin{cases} 1 & r_i(\hat{\lambda}_{(p)s})| \le c \\ \frac{c}{|r_i(\hat{\lambda}_{(p)s})|} & r_i(\hat{\lambda}_{(p)s})| > c \end{cases} & , \text{for iteration} = 1 \\ \frac{\rho}{\left(\frac{r_i(\hat{\lambda}_{(p)s})}{\sigma_s}\right)^2} & , \text{for iteration} > 1 \end{cases} \quad \dots (39)$$

The steps of the S method are as follows:

i. **Calculating the estimations of the parameters of the circular regression model by the ordinary least squares method.**

ii. **Calculating residuals $r_i(\hat{\lambda}_{(p)s})$.**

iii. **Calculate the value of $\hat{\sigma}_s$ according to equation (35).**

iv. **Calculating weights according to equation (37) Calculating estimations of $\hat{\lambda}_{(p)s}$s by WLS method Repeat steps (2-5) until we get similar values.**

The estimators of S are more impervious than the estimators of M because the estimators of S have a convergent bias and convergent variance less than the estimators of M in the case of data contamination.

(Ahmed et al., 2020) (ALMetwally and ALMmongy, 2018, 57-58) (Susanti et al., 2014, 7-8)

3. Discussion of Results

The Monte-Carlo Simulation method was adopted to estimate the parameters of the circular regression model of JS using the robust circular S estimator and compared them with the circular Least squares method using three comparison criteria, the median standard error (Median SE), the median mean squared error (Median MSE), and the mean of the mean cosine of the circular residuals (Median A(k)) to arrive at the best estimates. For the purpose of testing the estimation methods, the following steps were taken, First, model identification for the purpose of simplification and more clarification of the circular regression model, the case of the model will be considered when k=1, so we will get the model in equations (10) and (11).

For the sake of simplicity, A_0, C_0 will be considered equal to zero while the default values of A_1, B_1, C_1, D_1 are given by the trigonometric polynomial equations added to a cos (a+v) and sin (a+u) when a=2, So that:

$$cos(2+u) = -0.0416cos(u) - 0.9093sin(u)$$

$$sin(2+u) = 0.9093cos(u) - 0.0416sin(u)$$

The vector default values for parameters A_1, B_1, C_1, D_1 are (-0.04161, -0.09093, 0.09093, -0.04161) respectively.

Second, random errors generating where e will consider random errors that are uncorrelated ε_{1j} ، ε_{2j} and have a bivariate normal distribution with $\underline{\mu}=\underline{0}$ and the variances $\sigma^2{}_1$ and $\sigma^2{}_2$ 2 are 0 respectively.

Third, circular random variables generating were generated in the circular regression model of JS, which are the independent random variable U and the dependent circular random variable V from the Von-Misses distribution with mean π and a density parameter equal to 2 that is: $V\sim vM(\pi, 2)$. Figures (1), (2) and (3) show the generated circular data. (Alshqaq2, et al., 2021, 4)

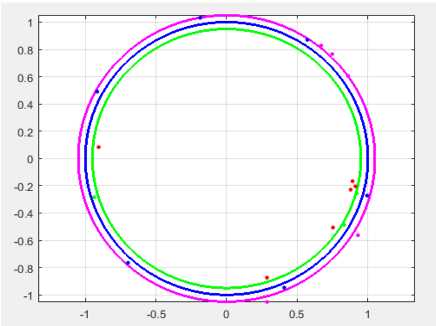

Figure (1) the generated circular data with a size (n=20)

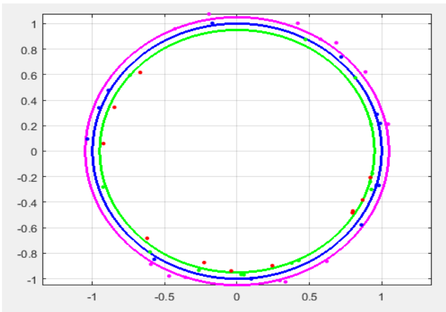

Figure (2) the generated circular data with a size (n=50)

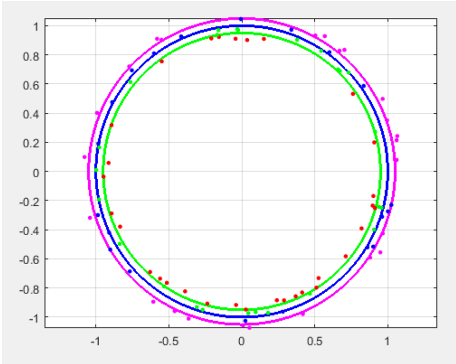

Figure (3) the generated circular data with a size (n = 100)

Fourth, Samples generating with size n=20, 50 and 100 will be generated to know the behavior of estimation methods at different samples. Fifth, for the purpose of verifying the robustness of the estimators against outlier in the values of the circular variables u and v, six outlets rates will be selected (5%, 10%, 20%, 30%, 40%, & 50%) according to the following scenarios:

i. **Using the data of the two variables without outliers and applying the estimation methods with the original data.**
ii. **Outlier of the dependent circular random variable values (vertical data) V only by the observation site d and then v_d and the polluting is as follows:**

$$v_d^* = v_d + \delta\pi \qquad \text{… (40)}$$

where: v_d^* are the values of the dependent variable v after pollution, δ is the percentage of pollution, since $0 \le \delta \le 1$.

iii. **Outlier of the values of the independent circular random variable (inflection points) u only by the observation site d and u_d is polluted by obtaining different percentages of the original data at location d using the Von-misses distribution with an average of 2π and a density parameter of 6 instead of the original data generated in above where $u_d \sim vM(2\pi, 6)$.**

Finally, for the purpose of identifying the priority of the estimation methods that were used in estimating the parameters of the JS circular regression model, three criteria were used, as follows: (Alshqaq2, et al., 2021, 5)

- **Median of Square Error:**

The square standard error median criterion was used for the model parameters according to the following formula:

$$M(SE)(\hat{\lambda}_j) = \sqrt{\frac{\sum_{j=1}^{s}(\hat{\lambda}_{ij} - \bar{\lambda})^2}{s}}\ ; \ j = 1, 2, \dots, 6 \qquad \text{… (41)}$$

where: $\bar{\lambda} = \frac{\sum_{j=1}^{s} \hat{\lambda}_{ij}}{s}$

- **Median of Mean Square Error:**

The median of mean square error criterion was used for the model parameters according to the following formula:

$$M(MSE) = \frac{\varepsilon_{ij}}{s}\ ; \ j = 1, 2, \dots, 6 \qquad \text{… (42)}$$

- **Mean of Cos of Circular Residuals:**

The median standard of cosines of circular residuals was used according to the following formula:

$$A(K) = \frac{\sum_{j=1}^{s} \cos(\varepsilon_{ij})}{s} \; ; \; j = 1, 2, \dots, 6 \qquad \dots (43)$$

From Table (1), it is clear that the Least Squares method was better than the methods of the Circular Fortress S method in the event that the data did not contain polluting values, as it recorded the lowest Median criterion, Mean Squares Error (Median MSE), the lowest Median Standard Error (Median SE) and the largest value of a Median Criterion, The Average Cosine of Circular residuals A(K) for all hypothetical sample sizes (n=20, 50, 100) when values of a=2.

From Table (1), it is clear that the Least Squares method was better than the methods of the Circular Fortress S method in the event that the data did not contain polluting values, as it recorded the lowest Median criterion, Mean Squares Error (Median MSE), the lowest Median Standard Error (Median SE) and the largest value of a Median Criterion. The Average Cosine of Circular residuals A(K) for all hypothetical sample sizes (n=20, 50, 100) when values of a=2.

Table (1) The results of analysis of the estimations used in the absence of contamination in the data

n	Criteria	Estimates		Best
		CLS	CRS	
20	Median MSE	1.13234	3.42334	CLS
	Median SE	4.74544	6.98755	CLS
	Median A(K)	0.99746	0.96894	CLS
50	Median MSE	1.11211	3.41256	CLS
	Median SE	4.73344	6.98741	CLS
	Median A(K)	0.99855	0.96994	CLS
100	Median MSE	1.10313	3.31678	CLS
	Median SE	4.72156	6.97464	CLS
	Median A(K)	0.99943	0.97877	CLS

From Table (2) in the case of pollution in the vertical data and when the value of a=2 it is clear that the Circular Least Squares method is not preferred for all pollution rates and for all sample sizes. The higher percentage of contamination in the vertical data, the greater the preference of the resilient estimation methods, so that the mean criterion of Mean Squares Error (Median MSE), and criterion of Median Standard Error (Median SE) decrease, and the value of the mean criterion of the Mean Cosines of the circular residuals A(K) increases for all default sample sizes.

Table (2): The results of analysis of the estimators used in the case of contamination in the vertical data when a = 2

N	Criteria	Estimates		Best	n	Criteria	Estimates		Best
		CLS	CRS				CLS	CRS	
5% Vertical					10% Vertical				
20	Median MSE	4.5636	4.4455	CRS	20	Median MSE	4.78943	4.21333	CRS
	Median SE	7.5633	7.3347	CRS		Median SE	7.93543	7.12252	CRS
	Median A(K)	0.9234	0.9789	CRS		Median A(K)	0.80355	0.97312	CRS
50	Median MSE	4.4124	4.4256	CRS	50	Median MSE	4.77133	4.21113	CRS
	Median SE	7.5134	7.3112	CRS		Median SE	7.91452	7.11355	CRS
	Median A(K)	0.9279	0.9796	CRS		Median A(K)	0.81382	0.97151	CRS
100	Median MSE	4.3346	4.3355	CRS	100	Median MSE	4.75138	4.11194	CRS
	Median SE	7.4168	7.3111	CRS		Median SE	7.83456	6.99676	CRS
	Median A(K)	0.9345	0.9888	CRS		Median A(K)	0.84327	0.98671	CRS
20% Vertical					30% Vertical				
20	Median MSE	5.5364	4.4243	CRS	20	Median MSE	9.78444	5.46633	CRS
	Median SE	8.5633	7.6675	CRS		Median SE	8.34843	7.33491	CRS
	Median A(K)	0.7859	0.9666	CRS		Median A(K)	0.55636	0.96786	CRS
50	Median MSE	5.4532	4.418	CRS	50	Median MSE	9.77453	5.44133	CRS
	Median SE	8.5525	7.7537	CRS		Median SE	8.31343	7.31224	CRS
	Median A(K)	0.7753	0.9765	CRS		Median A(K)	0.58355	0.95343	CRS
100	Median MSE	5.4433	4.4173	CRS	100	Median MSE	9.66766	5.41624	CRS
	Median SE	8.5324	7.7834	CRS		Median SE	8.23134	7.12313	CRS
	Median A(K)	0.7653	0.9783	CRS		Median A(K)	0.77454	0.95112	CRS

40% Vertical					50% Vertical				
20	Median MSE	9.9978	4.3326	CRS	20	Median MSE	9.96345	5.33433	CRS
	Median SE	8.6776	6.2247	CRS		Median SE	8.65322	5.22333	CRS
	Median A(K)	0.5556	0.9457	CRS		Median A(K)	0.44144	0.94114	CRS
50	Median MSE	9.9877	4.3115	CRS	50	Median MSE	9.87755	5.32144	CRS
	Median SE	8.6647	6.2146	CRS		Median SE	8.56444	5.12355	CRS
	Median A(K)	0.7656	0.9513	CRS		Median A(K)	0.65444	0.95556	CRS
100	Median MSE	9.9779	4.2363	CRS	100	Median MSE	9.66776	5.31455	CRS
	Median SE	8.6545	6.2115	CRS		Median SE	8.46476	5.11067	CRS
	Median A(K)	0.7735	0.9588	CRS		Median A(K)	0.66114	0.96443	CRS

From Table (3), in the case of pollution at the lifting points, the Circular Least Squares method is not preferred by a large percentage at all levels of pollution and for all sample sizes. And that the higher the percentage of pollution at the lifting points, the greater the preference of the resilient estimation methods, so that the mean criterion of Mean Squares Error (Median MSE) and criterion of Median Standard Error (Median SE) decrease, and the value of the mean criterion increases for the Mean Cos of the circular residuals A(K) and for all default sample sizes.

Table (3) Results of the analysis of estimators used in the case of contamination in patch points when a = 2

N	Criteria	Estimates		Best	n	Criteria	Estimates		Best
		CLS	CRS				CLS	CRS	
5% HLPs					10% HLPs				
20	Median MSE	3.39896	2.98756	CRLS	20	Median MSE	5.34133	2.95564	CRLS
	Median SE	7.86444	6.54564	CRLS		Median SE	8.78555	6.51795	CRLS
	Median A(K)	0.85644	0.95135	CRLS		Median A(K)	0.81134	0.93465	CRLS
50	Median MSE	3.37564	2.97757	CRLS	50	Median MSE	5.32445	2.81487	CRLS
	Median SE	7.83564	6.53455	CRLS		Median SE	8.68766	6.42144	CRLS
	Median A(K)	0.86786	0.95855	CRLS		Median A(K)	0.83668	0.94387	CRLS
100	Median MSE	3.35356	2.96644	CRLS	100	Median MSE	5.31141	2.71245	CRLS
	Median SE	7.81235	6.52235	CRLS		Median SE	8.62241	6.41808	CRLS
	Median A(K)	0.89245	0.96612	CRLS		Median A(K)	0.84421	0.95167	CRLS
20% HLPs					30% HLPs				
20	Median MSE	7.53755	4.54644	CRLS	20	Median MSE	8.88976	3.58977	CRLS
	Median SE	9.77865	5.53244	CRLS		Median SE	9.97897	4.67666	CRLS

	Median A(K)	0.78654	0.95667	CRLS		Median A(K)	0.75433	0.96665	CRLS
50	Median MSE	7.51223	4.53553	CRLS	50	Median MSE	8.85776	3.56755	CRLS
	Median SE	9.64447	5.52255	CRLS		Median SE	9.97446	4.53354	CRLS
	Median A(K)	0.79963	0.96101	CRLS		Median A(K)	0.75223	0.96755	CRLS
100	Median MSE	7.49665	4.52343	CRLS	100	Median MSE	8.84436	2.67657	CRLS
	Median SE	9.63324	5.51575	CRLS		Median SE	9.95575	4.45455	CRLS
	Median A(K)	0.81575	0.96212	CRLS		Median A(K)	0.74855	0.96929	CRLS
40% HLPs					50% HLPs				
20	Median MSE	8.97777	2.54444	CRLS	20	Median MSE	9.98777	2.45455	CRLS
	Median SE	9.99686	3.56555	CRLS		Median SE	10.6766	3.66464	CRLS
	Median A(K)	0.73435	0.96754	CRLS		Median A(K)	0.70557	0.98455	CRLS
50	Median MSE	8.86755	2.45436	CRLS	50	Median MSE	9.96576	2.44433	CRLS
	Median SE	9.96863	3.55977	CRLS		Median SE	10.5565	3.63544	CRLS
	Median A(K)	0.75132	0.96223	CRLS		Median A(K)	0.72781	0.98534	CRLS
100	Median MSE	8.76469	2.41133	CRLS	100	Median MSE	9.94565	2.43244	CRLS
	Median SE	9.95656	3.54633	CRLS		Median SE	10.4566	3.61322	CRLS
	Median A(K)	0.76345	0.96145	CRLS		Median A(K)	0.76341	0.98654	CRLS

4. Conclusions

i. The preference of the circular least squares method over the robust estimation methods (CTLS, S) when there is no contamination in the data and for all the assumed sample sizes and for all the different values of the assumed models.

ii. The least squares method is not preferred in the event that there are vertical polluted values or at the lifting points, and its lack of preference increases as the percentage of contamination increases in the data and for all assumed sample sizes and for all the different values of the assumed models.

iii. Convergence of the three comparison criteria (Median MSE, Median SE, A(K)) for robust estimation methods when there is contamination in vertical data and lifting points and for all assumed sample sizes and at all different values of the assumed models.

iv. The higher the pollution percentage, the higher the preference for robust estimation methods for all assumed sample sizes and for all different values of the assumed models.

v. The method of the robust circular S estimator is superior to the method of truncated circular least squares. The methods of estimation upon contamination in the vertical variable are superior.

vi. The method of circular truncated least squares is superior to other estimation methods for pollution at high inflection points.

References

1. Ahmed Dheyab Ahmed, Baydaa Ismael Abdulwahhab, Ebtisam Karim Abdulah, (2020), "A Comparison Among Different Methods for Estimating Regression Parameters with Autocorrelation Problem under Exponentially Distributed Error",[3] Computer Center, College of Administrations & Economics, University of Baghdad, Iraq, [*]Corresponding author: ahmedthieb@yahoo.com.
2. Al-Sarraf, Mustafa Nizar, the narrator, Asma Ghaleb, Ismail, Gharfan Kamal, “Statistical Estimation”, Al-Jazeera Office for Printing and Publishing, Baghdad, Deposit No. in the House of Documents and Documents in Baghdad (1588) for the year 2016, first edition.
3. Alshqaq, Shokrya Saleh, (2021), "On the Least Trimmed Squares Estimators for JS Circular Regression Model ", Kuwait J. Sci., Vol.48 (3), July.2021, pp(1-13).
4. Al-Yasiri, Tahani Mahdi Abbas, (2007), “Comparing the Estimations of the Hippocampal Basal with other Estimations to Estimate the Approximate Reliability Function of the Whipple Distribution”, Ph.D thesis, College of Administration and Economics, University of Baghdad.
5. Arthur Pewsey, Markus Neuhäuser, and Graeme D. Ruxton, (2013), "Circular Statistics in R", Oxford and New York: Oxford University Press. xiv+183 p.; ill.; index. ISBN: 978-0-19-967113-7.
6. Ehab Mohamed Almetwally, Hisham Mohamed Almongy, (2018), "Comparison Between M-Estimation, S-Estimation, and MM Estimation Methods of Robust Estimation with Application and Simulation", Journal of Mathematical Archive-9(11), 2018, 55-63 available online through www.ijma.info ISSN 2229 – 5046.
7. Farcomeni, A.; Greco, L. (2013), “Robust Methods for Data Reduction”, Boca Raton, FL: Chapman & Hall/ CRC Press, ISBN 978-1-4665-9062-5.
8. Ferreira, Sergio L. C.; Caires, Adriana O.; da S. Borges, Thaise; M.D.S. Lima Ariana, O.B. Silva, Laiana and N.L. dos Santos Walter ,(2017), "Robustness Evaluation in Analytical Methods Optimized Using Experimental Designs", Microchemical Journal 131 (2017) 163-169, http://dx.doi.org/10.1016/j.microc.2016.12.004 0026-265X/© 2016 Elsevier B.V. All rights reserved.
9. Hekimoglu, S., and Erenoglu, R. Z. (2013) “A New GM-Estimate with High Breakdown Point” Acta Geod Geophys, No. 48, PP 419-437.
10. Ibrahim, S., Rambli, A., Hussin, A.G., Mohamed, I., (2013). “Outlier Detection in a Circular Regression Model Using COVRATIO Statistic”; Commun. Stat.-Simul. Comput. 42 (10), 2272-2280.
11. Lukman, A. F., Osowole, O. I. & Ayinde, K.; (2015); “Two Stage Robust Ridge Method in a Linear Regression Model”, Journal of Modern Applied Statistical Methods, NO. 14(2), PP 53-67.

12. Rusiecki m Andrzej, (2009), “Robust Learning Algorithm with LTS Error Function'', Encyclopedia of Artificial Intelligence, DOI: 10.4018/978-1-59904-849-9.ch204.

13. Sarma, Y., Jammalamadaka, S., (1993); “Circular Regression”; In: Matusita, K. (Ed.), Statistical Theory and Data Analysis. VSP, Utrecht, pp. 109128109–109128128.

14. Shokrya Alshqaq a, Ali Abuzaid b Abdullah Ahmadini a, “Robust Estimators for Circular Regression Models”, Journal of King Saud University-Science 33 (2021) 101576.

15. Susanti, Yuliana §, Hasih Pratiwi, Sri Sulistijowati H., Twenty Liana, (2014), "M Estimation, S Estimation, and MM Estimation in Robust Regression", International Journal of Pure and Applied Mathematics Volume 91 No. 3 2014, 349-360 ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version) URL: http://www.ijpam.eu doi: http://dx.doi.org/10.12732/ijpam.v91i3.7.

مقارنة مقدرات المربعات الصغرى الدائرية الحصينة والدائرية S الحصينة لأنموذج الانحدار الدائري باستخدام المحاكاة

الباحث/ هدى هديب عباس
جامعة بغداد- المعهد العالي للدراسات المحاسبية والمالية
Hoda.Hadib201a@coadec.uobaghdad.edu.iq

أ.م. سهيل نجم عبود
جامعة بغداد-كلية الادارة والاقتصاد – قسم الاحصاء، بغداد، العراق
suhnaj2005@coadec.uobaghdad.edu.iq

Received:1/12/2022 Accepted:13/12/2022 Published: December / 2022

مستخلص البحث

في هذا البحث تم استعمال اسلوب المحاكاة مونت- كارولوا لغرض مقارنة مقدر S الدائري الحصين مع طريقة المربعات الصغرى الدائرية (Circular Least squares method) في حالة عدم تلوث البيانات وفي حالة وجود تلوث في البيانات من خلال اتجاهين الاول تلوث بنقاط الانعطاف العالية التي يمثل التلوث في المتغير المستقل الدائري والثاني التلوث في المتغير العمودي الذي يمثل المتغير المعتمد الدائري باستعمال ثلاثة معايير للمقارنة هي وسيط الخطأ المعياري (Median SE) ووسيط متوسط مربعات الخطأ (Median MSE) ووسيط متوسط الجتا للبواقي الدائرية (Median A(k)) . وتم التوصل الى ان طريقة المربعات الصغرى افضل من طرائق طريقة S الحصينة الدائرية في حالة عدم احتواء البيانات على قيم ملوثة كونها سجلت اقل معيار وسيط متوسط مربعات خطأ (Median MSE) واقل وسيط خطأ معياري (Median SE) واكبر قيمة لمعيار وسيط متوسط جتا البواقي الدائرية A(K) ولكافة احجام العينات الافتراضية (n=20, 50, 100) . وفي حالة التلوث في البيانات العمودية اتضح عدم افضلية طريقة المربعات الصغرى الدائرية عند كافة نسب التلوث ولكافة احجام العينات. وانه كلما ازادت نسبة التلوث في البيانات العمودية زادت افضلية طرائق التقدير الحصينة بحيث يقل معيار وسيط متوسط مربعات خطأ (Median MSE) ومعيار وسيط خطأ معياري (Median SE) وتزداد قيمة لمعيار وسيط متوسط جتا البواقي الدائرية A(K) ولكافة احجام العينات الافتراضية. وفي حالة التلوث في نقاط الرفع العالية عدم افضلية طريقة المربعات الصغرى الدائرية بنسبة كبيرة عند كافة نسب التلوث ولكافة احجام العينات. وانه كلما ازادت نسبة التلوث في نقاط الرفع زادت افضلية طرائق التقدير الحصينة بحيث يقل معيار وسيط متوسط مربعات خطأ (Median MSE) ومعيار وسيط خطأ معياري (Median SE) وتزداد قيمة لمعيار وسيط متوسط جتا البواقي الدائرية A(K) ولكافة احجام العينات الافتراضية.

نوع البحث: ورقة بحثية

المصطلحات الرئيسة للبحث: الحصانة ، الانحدار الدائري ، مقدر S ، مقدر المربعات الصغرى ، الاحصاء الدائري، البيانات الدائرية، المشاهدات الشاذة.

*البحث مستل من رسالة ماجستير