On Robust Method of Estimation in Polynomial-Sinusoidal Regression

Debasis Kundu *

Abstract

In this paper we have considered the polynomial-sinusoidal regression model in presence of additive errors. The standard method of estimation of the unknown parameters is the least squares approach. It is well known that the least squares estimators are not robust in presence of outliers. We have proposed a weighted least squares method and it can be more robust than the least squares estimators in presence of outliers. The proposed weighted least squares method can be implemented very easily compared to some of the well known robust estimators like least absolute deviation estimators or Huber's M (Huber-M) estimators. Theoretical properties of the proposed estimators have been obtained under the same set of assumptions as what are needed for the least squares estimators. Simulations have been performed to show the effectiveness of the proposed method, and it is observed that the weighted least squares estimators behave better than the least squares estimators in presence of outliers and they behave at par with the other robust estimators. Further, the implementation of the LAD and Huber-M estimators is computationally quite challenging when the number of components is large. It is observed that the performance of the weighted least squares estimators depends on the choice of the weight function. We have proposed a method to choose the proper weight function. We have analyzed one data set for illustrative purposes.

Key Words and Phrases: sinusoidal model; least squares estimators; weighted least squares estimators; robust estimators; consistency; asymptotic normality.

AMS Subject Classifications: 62F10, 62F03, 62H12.

*Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500.

1 Introduction

In this paper we have considered the following polynomial sinusoidal regression model in presence of additive noise;

$$y(t_i) = \beta_0^0 + \beta_1^0 t_i + A^0 \cos(\theta^0 t_i) + B^0 \sin(\theta^0 t_i) + e(t_i); \quad i = 1, \ldots, n. \tag{1}$$

Here β_0^0, β_1^0, A^0, B^0 are linear parameters, $A^{0^2} + B^{0^2}$ is known as the amplitude and $0 < \theta^0 < \pi$ is the frequency. It is assumed $0 < t_1 < \ldots < t_n < \infty$ and t_i's are equidistant. Further, $e(t_i)$'s are error random variables with mean zero and finite variance $\sigma^2 > 0$. The explicit assumptions on $e(t)$'s will be mentioned later. It should be emphasized that in the model (1) and throughout the paper $\beta_0^0, \beta_1^0, A^0, B^0$ and θ^0 denote the true parameter values of the model, from which the original signal has been obtained. Some of the above parameters might be unknown, and our aim is to estimate those unknown parameters. We denote β_0, β_1, A, B and θ as the arbitrary parameters of the signal, need not be the true one, and it will be clear from the context.

Since, it involves linear (polynomial) regression term and also sinusoidal regression term the model is known as the polynomial sinusoidal regression model. The main aim of this paper is to provide robust estimators of A^0, B^0 and θ^0 in presence of outliers. Once robust estimators of A^0, B^0 and θ^0 are obtained, robust estimators of β_0^0 and β_1^0 also can be obtained by linear regression technique. Hence, we have not addressed that here explicitly. In the statistical literature sometimes they have been treated as nuisance parameters, see Li and Nie [1].

It may be mentioned that a more general form of the model (1) is also available in the literature, and it has the following form;

$$y(t_i) = \sum_{j=0}^{p} \beta_j^0 t_i^j + \sum_{k=1}^{q} \{A_k^0 \cos(\theta_k^0 t_i) + B_k^0 \sin(\theta_k^0 t_i)\} + e(t_i); \quad i = 1, \ldots, n, \tag{2}$$

where it has a p degree polynomial and q number of sinusoidal components. First we will provide the necessary methodology for model (1), and then we will show that most of the methods can be extended for the model (2). It may be mentioned that when the first component is absent, i.e. the multiple sinusoidal model has received a significant amount of attention in the signal processing literature, see for example Stoica [2], Surana et al. [3], Nandi and Kundu [4], Qian et al. [5] and the references cited therein. A special case of the model (2) has been considered by Trapero et al. [6] in the signal processing literature. See also Perron et al. [7] in this respect.

The model has several applications in different areas. When θ^0 is known and t_i's are unbounded, the model (1) has a significant application in econometrics. Abadir et al. [8] considered model (2) with $p = 1$ under a more general error structure, and provided the properties of the least squares estimators. Nandi and Kundu [9] also considered the model (2) as Abadir et al. [8] when $p = 1$, but θ^0 is assumed to be unknown. Raspaud et al. [10] used this model in the signal processing for tracking and time stretching. See also Li and Nie [1, 11], Rigdon et al. [12] in this respect for some applications in general of such partially nonlinear models in ecology, public health and in some other areas.

There exists an extensive amount of literature which deals with the design aspects of this model when t_i are bounded and θ^0s are known, see for example Lau and Studden [13] and the references cited there in. When θ^0 is known and t_i's are bounded, Eubank and Speckman [14] effectively used model (2) as an alternative to non-parametric regression technique. They have shown that the model is very effective, easy to implement in practice and it is quite competitive in terms of mean squared errors perspective, with a number of widely studied non-parametric regression techniques, see also Eubank and Speckman [15]. In this paper we will mainly consider the case when t_i's are unbounded, and without loss of generality it is assumed $t_i = i$. The frequency parameter θ^0 is assumed to be unknown also.

The aim of this paper is to develop a robust method of estimation of A^0, B^0 and θ^0 in presence of outliers. The model (1) is a non-linear regression model. In presence of additive errors, the most natural estimators are the least squares (LS) estimators. But unfortunately it is well known that the LS estimators are not robust. The performances of the LS estimators affect significantly in presence of few outliers only. The standard robust estimators are the least absolute deviation (LAD) estimators and Huber-M estimators. There are mainly two different issues about both the methods. First of all the implementation of both the methods involve solving higher dimensional non-linear optimization problem. For example, for model (1) if θ^0 is unknown both the methods involve solving five dimensional optimization problem. The problem becomes more intense for model (2). The second issue about both the methods is to develop the properties of the estimators. In developing the consistency and asymptotic normality properties of the LAD or Huber-M estimators, one needs stronger error assumptions than what are needed for the LS estimators, see for example Kim et al. [16] and Huber [17]. Moreover, if there is no outliers in the data set, the performances of the LAD and Huber-M estimators are quite poor compared to the LS estimators.

Recently Kundu and Grover [18] proposed a weighted least squares (WLS) method as a robust method of estimation in two dimensional sinusoidal frequency model. In this paper we have proposed a new WLS method to provide robust estimators of the unknown parameters of the polynomial-sinusoidal model. We have used polynomial weight function for this purpose. The basic idea of the method is quite intuitive. If it is known in advance the position where the outliers might be present, less weight is assigned to the residual squares in that portion of the data set with respect to the overall contribution of the residual sum of squares. The main advantage of the proposed method is that they can be implemented quite conveniently in practice. Further the consistency and asymptotic normality properties of the proposed estimators can be obtained under the same set of assumptions that are needed for the LS estimators. The implementation of the algorithm and the asymptotic properties of the estimators can be easily generalized for the model (2). We have performed an extensive simulations to show the effectiveness of the proposed method. It is observed that the WLS estimators behave better than the LS estimators and at par with the LAD and Huber-M estimators in terms of mean squared errors in presence of outliers and if there is no outliers, the performances of the WLS estimators are at par with the LS estimators and better than the LAD and Huber-M estimators. Further, we have provided some time comparison of the computation of the LAD and Huber-M estimators with respect to the WLS estimators, and it is observed that the computation time of the WLS estimators is significantly smaller than LAD and Huber-M estimators.

The performance of the proposed method depends very much on the choice of the weight function. As it has been mentioned before that if it is known in advance where the outliers might be present the weight function can be chosen accordingly. But in practice often it may not be known. We have proposed a method to choose the weight function also when the position of the outliers is not known in advance. We have demonstrated the method by analyzing a real data set. The results are quite satisfactory. Hence, the proposed method can be implemented in practice quite conveniently.

The rest of the paper is organized as follows. In Section 2 we have discussed about the WLS method. The asymptotic properties of the WLS estimators have been provided in Section 3. In Section 4 we have provided the simulation results. The choice of the weight function and the analysis of a real data set have been presented in Section 5. Finally we conclude the paper in Section 6. All the proofs have been presented in the Appendix.

2 Weighted Least Squares Estimators

Assumption 1: The error random variables $e(t)$'s have the following structure

$$e(t) = \sum_{k=-\infty}^{\infty} a(k)u(t-k).$$

Here $u(t)$'s are independent and identically distributed (i.i.d.) random variables with mean zero and finite variance $\sigma^2 > 0$. Further $a(k)$'s satisfy the following condition

$$\sum_{k=-\infty}^{\infty} |a(k)| < \infty.$$

Consider a polynomial of degree k as follows:

$$w(t) = c_0 + c_1 t + \ldots + c_k t^k; \quad t \geq 0. \tag{3}$$

Here, $c_0, \ldots, c_k$ are the coefficients of the polynomial and they satisfy the following conditions

$$\inf_{0 \leq t \leq 1} w(t) \geq \gamma > 0 \quad \text{and} \quad \int_0^1 w(t)dt = 1. \tag{4}$$

Let us consider the first difference series: $z(t) = y(t+1) - y(t)$, for $t = 1, \ldots, n$. Hence,

$$\begin{aligned} z(t) &= y(t+1) - y(t) \\ &= \beta_1^0 + A^0[\cos(\theta^0(t+1)) - \cos(\theta^0 t)] + B^0[\sin(\theta^0(t+1)) - \sin(\theta^0 t)] + e_d(t), \end{aligned} \tag{5}$$

where $e_d(t) = e(t+1) - e(t)$, for $t = 1, \ldots, n$. Note that $e_d(t)$ can be written as

$$\begin{aligned} e_d(t) = e(t+1) - e(t) &= \sum_{k=-\infty}^{\infty} a(k)e(t+1-k) - \sum_{k=-\infty}^{\infty} a(k)e(t-k) \\ &= \sum_{k=-\infty}^{\infty} (a(k+1) - a(k))e(t-k) = \sum_{k=-\infty}^{\infty} b(k)e(t-k), \end{aligned}$$

here $b(k) = a(k+1) - a(k)$. Since $\sum_{k=-\infty}^{\infty} |a(k)| < \infty$, implies $\sum_{k=-\infty}^{\infty} |b(k)| < \infty$. Therefore, if $e(t)$'s satisfy Assumption 1, then $e_d(t)$'s also satisfy Assumption 1.

2.1 One Component Sinusoidal Regression Model

In this section we will discuss about the WLS method of the model

$$y(t) = \beta_0^0 + \beta_1^0 t + A^0 \cos(\theta^0 t) + B^0 \sin(\theta^0 t) + e(t); \quad t = 1, \ldots, n+1. \tag{6}$$

Let us use the notation $\boldsymbol{\Theta}_1 = (A, B)^\top$. In this section we will work with the series $\{z(t); t = 1, \ldots, n\}$. First we will estimate β_1^0 as

$$\widetilde{\beta}_1 = \frac{1}{\sum_{i=1}^{n} w\left(\frac{i}{n}\right)} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) z(t). \tag{7}$$

We will show that $\widetilde{\beta}_1$ is a consistent estimate of β_1^0. If θ^0 is known, the WLS estimators of A^0 and B^0 can be obtained by minimizing $Q(A, B, \theta^0)$, where

$$Q(A,B,\theta^0) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)(z(t) - \widetilde{\beta}_1 - A(\cos(\theta^0(t+1)) - \cos(\theta^0 t)) - B(\sin(\theta^0(t+1)) - \sin(\theta^0 t)))^2, \tag{8}$$

with respect to A and B. If $\widetilde{A}(\theta^0)$ and $\widetilde{B}(\theta^0)$ minimize $Q(A, B, \theta^0)$ with respect to A and B, respectively, then they can be written as follows:

$$\widetilde{A}(\theta^0) = \frac{1}{a_{11}a_{22} - (a_{12})^2}[a_{22}d_1 - a_{12}d_2] \quad \text{and} \quad \widetilde{B}(\theta^0) = \frac{1}{a_{11}a_{22} - (a_{12})^2}[a_{11}d_2 - a_{12}d_1],$$

where

$$a_{11} = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\Big(\cos(\theta^0(t+1)) - \cos(\theta^0 t)\Big)^2, \quad a_{22} = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\Big(\sin(\theta^0(t+1)) - \sin(\theta^0 t)\Big)^2,$$

$$a_{12} = a_{21} = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\Big(\cos(\theta^0(t+1)) - \cos(\theta^0 t)\Big)\Big(\sin(\theta^0(t+1)) - \sin(\theta^0 t)\Big),$$

$$d_1 = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)(z(t) - \widehat{\beta}_1(\theta^0))(\cos(\theta^0(t+1)) - \cos(\theta^0 t)),$$

$$d_2 = \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)(z(t) - \widehat{\beta}_1(\theta^0))(\sin(\theta^0(t+1)) - \sin(\theta^0 t)).$$

It may be seen that as $n \to \infty$,

$$a_{11} \longrightarrow 2\sin^2\left(\frac{\theta^0}{2}\right)\int_0^1 w(t)dt, \qquad a_{22} \longrightarrow 2\sin^2\left(\frac{\theta^0}{2}\right)\int_0^1 w(t)dt, \qquad a_{12} = a_{21} \longrightarrow 0.$$

Hence, for large n

$$\widetilde{A}(\theta^0) = \frac{d_1}{2\sin^2\left(\frac{\theta^0}{2}\right)\int_0^1 w(t)dt} \quad \text{and} \quad \widetilde{B}(\theta^0) = \frac{d_2}{2\sin^2\left(\frac{\theta^0}{2}\right)\int_0^1 w(t)dt}. \tag{9}$$

When θ^0 is unknown, the WLS estimators of A^0, B^0 and θ^0 can be obtained by minimizing

$$Q(A,B,\theta)=\sum_{t=1}^{n} w\left(\frac{t}{n}\right)(z(t)-\widetilde{\beta}_1-A(\cos(\theta(t+1))-\cos(\theta t))-B(\sin(\theta(t+1))-\sin(\theta t)))^2, \tag{10}$$

with respect to A, B and θ. It is equivalent to obtain the WLS estimator of θ^0, say $\widetilde{\theta}$, first by minimizing $R(\theta)$ with respect to θ, where

$$R(\theta)=\sum_{t=1}^{n} w\left(\frac{t}{n}\right)(z(t)-\widetilde{\beta}_1-\widetilde{A}(\theta)(\cos(\theta(t+1))-\cos(\theta t))-\widetilde{B}(\theta)(\sin(\theta(t+1))-\sin(\theta t)))^2. \tag{11}$$

Here, $\widetilde{A}(\theta)$ and $\widetilde{B}(\theta)$ can be obtained from (9) by replacing θ^0 with θ. Then obtain the WLS estimator of A^0 and B^0 as $\widetilde{A}=\widetilde{A}(\widetilde{\theta})$ and $\widetilde{B}=\widetilde{B}(\widetilde{\theta})$. Hence, the WLS estimators A^0, B^0 and θ^0 can be obtained by solving a one-dimensional optimization problem.

2.2 Multicomponent Sinusoidal Regression Model

In this section we will discuss about the WLS method of the model

$$y(t)=\beta_0^0+\beta_1^0 t+\sum_{k=1}^{q}\left\{A_k^0\cos(\theta_k^0 t)+B_k^0\sin(\theta_k^0 t)\right\}+e(t); \quad t=1,\ldots,n+1. \tag{12}$$

In this case $0<\theta_1^0,\ldots,\theta_q^0<\pi$ and without loss of generality it is assumed that

$$A_1^{0^2}+B_1^{0^2}>A_2^{0^2}+B_2^{0^2}>\ldots>A_q^{0^2}+B_q^{0^2}. \tag{13}$$

Here also the sequence $\{z(t);t=1,\ldots,n\}$ is same as defined before. In this case also we estimate $\widetilde{\beta}_1$ same as in (7). In this case also $\widetilde{\beta}_1$ is a consistent estimator of β_1^0.

Let us use the notation $\mathbf{\Theta}_q=(A_1,B_1,\ldots,A_q,B_q)^\top$. If $\theta_1^0,\ldots,\theta_q^0$ are known, the WLS estimators of $A^0,\ldots,A_q^0$, $B_1^0,\ldots,B_q^0$ can be obtained by minimizing

$$\begin{aligned}Q(\beta_1,\mathbf{\Theta}_q) \;=\; &\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\Bigg(z(t)-\widetilde{\beta}_1- \\ &\sum_{k=1}^{q}A_k(\cos(\theta_k^0(t+1))-\cos(\theta_k^0 t))+B_k(\sin(\theta_k^0(t+1))-\sin(\theta_k^0 t))\Bigg)^2,\end{aligned} \tag{14}$$

with respect to $A_1, B_1, \ldots A_q, B_q$. The explicit solutions can be obtained in this case. If $\theta_1^0, \ldots, \theta_q^0$ are not known, then it involves a q-dimensional optimization problem. To avoid that we provide the following sequential procedure which involves solving q one-dimensional optimization problem. First we minimize $Q_1(A_1, B_1, \theta_1)$ with respect to A_1, B_1 and θ_1, where

$$Q_1(A_1, B_1, \theta_1) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)(z(t) - \widetilde{\beta}_1 - A_1(\cos(\theta_1(t+1)) - \cos(\theta_1 t)) - B_1(\sin(\theta_1(t+1)) - \sin(\theta_1 t)))^2. \tag{15}$$

Let $\widetilde{A}_1$, $\widetilde{B}_1$ and $\widetilde{\theta}_1$ minimize (15), we call them as the sequential WLS estimators. We have already shown that the minimization of (15) can be performed by solving a one-dimensional optimization problem. At the second stage we obtain $\{z_(t); t = 1, \ldots, n\}$, where

$$z_1(t) = z(t) - \widetilde{\beta}_1 - \widetilde{A}_1(\cos(\widetilde{\theta}_1(t+1)) - \cos(\widetilde{\theta}_1 t)) - \widetilde{B}_1(\sin(\widetilde{\theta}_1(t+1)) - \sin(\widetilde{\theta}_1 t)).$$

We minimize

$$Q_2(A_2, B_2, \theta_2) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)(z_1(t) - A_2(\cos(\theta_2(t+1)) - \cos(\theta_2 t)) - B_2(\sin(\theta_2(t+1)) - \sin(\theta_2 t)))^2, \tag{16}$$

with respect to A_2, B_2 and θ_2. The minimization of (16) can be performed by solving a one-dimensional optimization problem. We continue the process q times, and obtain the sequential estimators of $A_1, B_1, \theta_1, \ldots, A_q, B_q, \theta_q$.

3 Theoretical Properties of the WLSEs

In this section we provide the consistency and the asymptotic normality properties of the WLS estimators of the unknown parameters of the polynomial-sinusoidal regression model. We consider two cases separately.

3.1 One Component Sinusoidal Regression Model

Theorem 1: If $e(t)$'s satisfy Assumption 1, then $\widetilde{\beta}_1$ is a strongly consistent estimator of β_1.

Proof: See in the Appendix A. ∎

Theorem 2: If $e(t)$'s satisfy Assumption 1 and $0 < \theta^0 < \pi$, there exists M such that $|A^0| \le M$ and $|B^0| \le M$, then $\widetilde{A}$, $\widetilde{B}$ and $\widetilde{\theta}$, the WLS estimators of A, B and θ, respectively, are strongly consistent.

Proof: See in the Appendix A. ∎

We use the following notations for further development. For $p = 0, 1, 2, \ldots$,

$$u_p = \int_0^1 t^p w(t) dt \quad \text{and} \quad v_p = \int_0^1 t^p w^2(t) dt.$$

The 3×3 diagonal matrix $\boldsymbol{D}$ is

$$\boldsymbol{D} = \text{diag}\{n^{1/2}, n^{1/2}, n^{3/2}\},$$

the matrices

$$\boldsymbol{\Sigma}(\boldsymbol{\Gamma}^0) = 2(1 - \cos(\theta^0)\left[\begin{array}{ccc} v_0 & 0 & B^0 v_1 \\ 0 & v_0 & -A^0 v_1 \\ B^0 v_1 & -A^0 v_1 & (A^{0^2} + B^{0^2}) v_2 \end{array}\right] \tag{17}$$

and

$$\boldsymbol{\Psi}(\boldsymbol{\Gamma}^0) = 2(1 - \cos(\theta^0)\left[\begin{array}{ccc} u_0 & 0 & B^0 u_1 \\ 0 & u_0 & -A^0 u_1 \\ B^0 u_1 & -A^0 u_1 & (A^{0^2} + B^{0^2}) u_2 \end{array}\right] \tag{18}$$

Theorem 3: Under the same assumptions as in Theorem 3, if the matrices as defined in (17) and (18) are of full rank, then

$$\left((\widetilde{A} - A^0), (\widetilde{B} - B^0), (\widetilde{\theta} - \theta^0)\right) \boldsymbol{D} \xrightarrow{d} N_3(\mathbf{0}, (2\sigma^2 \xi(\theta^0))(\boldsymbol{\Psi}(\boldsymbol{\Gamma}^0))^{-1} \boldsymbol{\Sigma}(\boldsymbol{\Gamma}^0)(\boldsymbol{\Psi}(\boldsymbol{\Gamma}^0))^{-1}).$$

Here, $\xrightarrow{d}$ means convergence in distribution, $N_3(\mathbf{0}, \boldsymbol{A})$ indicates a 3-variate normal distribution with the mean vector $\mathbf{0}$ and the dispersion matrix $\boldsymbol{A}$ and

$$\xi(\theta^0) = \left| \sum_{k=-\infty}^{\infty} \beta(k) e^{i\theta^0 k} \right|^2.$$

Proof: See in the Appendix A. ∎

3.2 Multicomponent Sinusoidal Model

Theorem 4: If $e(t)$'s satisfy Assumption 1 and $0 < \theta_1^0, \ldots, \theta_q^0 < \pi$, there exists M such that $0 < |A_1^0|, \ldots, |A_q^0| \leq M$, $0 < |B_1^0|, \ldots, |B_q^0| \leq M$, and the amplitudes satisfy (13), then $\widetilde{A}_k$, $\widetilde{B}_k$ and $\widetilde{\theta}_k$, the WLS estimators of A_k, B_k and θ_k, for $k = 1, \ldots, q$, respectively, are strongly consistent.

Proof: See in the Appendix B. ∎

To provide the asymptotic distribution of the WLS estimators, we need to define the following matrices for $j = 1, \ldots, q$.

$$\boldsymbol{\Sigma}_j(\boldsymbol{\Gamma}_j^0) = 2(1-\cos(\theta_j^0) \begin{bmatrix} v_0 & 0 & B_j^0 v_1 \\ 0 & v_0 & -A_j^0 v_1 \\ B_j^0 v_1 & -A_j^0 v_1 & (A_j^{0^2} + B_j^{0^2}) v_2 \end{bmatrix} \tag{19}$$

and

$$\boldsymbol{\Psi}(\boldsymbol{\Gamma}^0) = 2(1-\cos(\theta_j^0) \begin{bmatrix} u_0 & 0 & B^0 u_1 \\ 0 & u_0 & -A^0 u_1 \\ B_j^0 u_1 & -A_j^0 u_1 & (A_j^{0^2} + B_j^{0^2}) u_2 \end{bmatrix} \tag{20}$$

Theorem 5: Under the same assumptions as in Theorem 4, if the matrices as defined in (19) and (20) are of full rank, then for $j = 1, \ldots, q$,

$$\Big((\widetilde{A}_j - A_j^0), (\widetilde{B}_j - B_j^0), (\widetilde{\theta}_j - \theta_j^0)\Big)\, \boldsymbol{D} \xrightarrow{d} N_3(\mathbf{0}, (2\sigma^2 \xi(\theta_j^0))(\boldsymbol{\Psi}_j(\boldsymbol{\Gamma}_j^0))^{-1} \boldsymbol{\Sigma}_j(\boldsymbol{\Gamma}_j^0)(\boldsymbol{\Psi}_j(\boldsymbol{\Gamma}_j^0))^{-1}).$$

Here,

$$\xi(\theta_j^0) = \left| \sum_{k=-\infty}^{\infty} \beta(k) e^{i\theta_j^0 k} \right|^2.$$

Moreover, for $1 \leq j \neq k \leq q$, $(\widetilde{A}_j, \widetilde{B}_j, \widetilde{\theta}_j)$ is asymptotically independent of $(\widetilde{A}_j, \widetilde{B}_j, \widetilde{\theta}_j)$.

Proof: See in the Appendix A. ∎

4 Simulation Results

In this section we present the results of some simulation experiments to show how the proposed methods work in practice for different sample sizes, for different error variances and for different weight functions in presence of outliers. The main aim of this section is to see how the proposed WLS estimators (WLSE), the LS estimators (LSE) and the other robust estimators like LAD estimators (LADE) and Huber-M estimators perform in terms of average estimates and mean squared errors (MSEs) for this polynomial-sinusoidal regression model in presence of outliers. We consider the following models:

Model 1: $p = q = 1$, $\beta_0 = 1.0$, $\beta_1 = 2.0$, $A = B = 5.0$ and $\omega = 0.75\pi = 2.35619$.

Model 2: $p = 1$, $q = 2$, $\beta_0 = 1.0$, $\beta_1 = 2.0$, $A_1 = B_1 = 10.0$, $A_2 = B_2 = 5.0$,

. $\omega_1 = 0.75\pi = 2.35619$, $\omega_2 = 0.15\pi = 0.47124$.

We have considered the following error sequences: $e(t) = u(t) + 0.5u(t-1)$, where $u(t)$'s are i.i.d. normal random variables with $E(u(t)) = 0$ and $V(u(t)) = \sigma_u^2 = 1$.

It is assumed that 10%, 20%, 30% and 40% outliers are present in the sample and in those cases $V(u(t)) = \sigma_{out}^2 = 100$. Moreover, it is also assumed that the outliers are present in the middle portion of the data sequence. We have considered the following two weight functions to see the effect of weight functions.

$$w_1(t) = \frac{1}{12} + 11\left(t - \frac{1}{2}\right)^2 \quad \text{and} \quad w_2(t) = \begin{cases} 2 - 4t & \text{if} \quad 0 \le t \le \frac{1}{4} \\ 1 & \text{if} \quad \frac{1}{4} \le t \le \frac{3}{4} \\ 4t - 2 & \text{if} \quad \frac{3}{4} \le t \le 1. \end{cases}$$

In both the cases it has put less weight in the middle portion of the data sequence. Different sample sizes namely $n = 25$, 50, 75 and 100 have been considered. In each case we have computed the average estimates and the mean squared errors based on 10000 replications. The results are compared with the LAD and Huber-M estimators. We have reported the results in Tables 1 to 6. In Tables 1 to 4 we have reported the results for Model 1, and in

Tables 5 and 6 we have reported the results for Model 2 parameters. In all the cases we have reported the asymptotic theoretical lower variances also (CRLB). Due to space restriction for Model 2 parameters we have reported the results for $n = 25$ and 100, and when the percentage of outliers are 20% and 40% only. The other results are available with the author, and they can be obtained on request. Note that to compute the LAD and Huber-M estimators, for Model 1, one needs to solve a five dimensional optimization problem, and for Model 2, it is a eight dimensional optimization problem. In case of the WLSEs for Models 1 and 2, one needs to solve one and two, respectively, 1-dimensional optimization problem only.

Some of the points are quite clear from these simulation results. First of all in each case as the sample size increases the mean squared errors (MSEs) decrease. It verifies the consistency properties of the estimators. As the percentage of the outliers increases, the MSEs increase, as expected. The performances of the WLSEs are very similar to the LSEs when the percentage of outliers is small and they are very similar with the robust estimators when the percentage of the outliers increases. Between the two WLSEs sometimes one performs better than the other and vice versa. Same phenomena have been observed between the two robust estimators also. From these simulation results it can be concluded that the WLSEs perform quite satisfactorily in presence of outliers and the choice of the weight function is an important issue. It will be discussed in details in the next section.

Now we provide the time comparison of the computational time of the LAD and Huber-M estimators with respect to the WLSEs. We provide the ratio of the computational time of the LAD and Huber-M estimators to the WLSEs. It is observed that the ratio does not depend on the sample size and the proportion of the outliers, it mainly depends on the order of the model, i.e. mainly on p and q. Further, the computational time of the LSEs and the WLSEs are almost same, hence they have not reported here. We have presented the results for the LAD and Huber-M estimators with respect to the WLSEs, when the weight function

PoO	Par	LSE	W1	W2	LADE	Huber-M	CRLB
10	A	5.47106	4.72271	5.15166	5.01454	4.88476	
		(2.2247E-01)	(3.6696E-02)	(2.3083E-02)	(2.0781E-02)	(2.3188E-02)	(1.9945E-02)
	B	4.52999	4.95479	4.75931	5.24694	4.97341	
		(2.2069E-01)	(1.9989E-02)	(5.7677E-02)	(6.1141E-02)	(7.0823E-02)	(1.9945E-02)
	ω	2.35344	2.35639	2.35480	2.35696	2.35639	
		(6.0439E-06)	(4.8106E-08)	(1.6871E-07)	(4.6754E-07)	(3.7112E-08)	(3.1234E-08)
20	A	5.96563	4.81615	5.45158	5.24557	5.10516	
		(9.3399E-01)	(3.9679E-02)	(3.9399E-02)	(6.0535E-02)	(4.1027E-02)	(3.8059E-02)
	B	4.85884	5.01194	4.94909	5.25329	5.05927	
		(2.0053E-01)	(8.5352E-02)	(5.6498E-02)	(6.4110E-02)	(6.5396E-02)	(3.8059E-02)
	ω	2.35295	2.35582	2.35414	2.35552	2.35552	
		(1.1959E-05)	(3.4966E-07)	(4.4261E-07)	(6.9633E-07)	(6.4477E-07)	(5.9601E-08)
30	A	4.71065	4.24914	4.59465	5.09259	5.14178	
		(8.4068E-01)	(6.6324E-02)	(6.6441E-02)	(7.5485E-02)	(7.0147E-02)	(5.6175E-02)
	B	4.83561	5.05888	5.06221	5.26905	5.16892	
		(8.6835E-01)	(6.4347E-02)	(6.8374E-02)	(7.2445E-02)	(7.8572E-02)	(5.6175E-02)
	ω	2.34447	2.35577	2.35176	2.35639	2.35551	
		(1.3728E-04)	(9.6458E-08)	(1.9767E-07)	(9.0232E-08)	(7.5317E-07)	(8.7971E-08)
40	A	5.65696	4.84269	5.24000	5.07498	5.55352	
		(4.3242E-01)	(7.4852E-02)	(7.7340E-02)	(8.6189E-02)	(8.0604E-02)	(7.4291E-02)
	B	4.32323	4.88243	4.76677	5.26283	5.19927	
		(4.5769E-01)	(7.6904E-02)	(7.7380E-02)	(8.9024E-02)	(8.9837E-02)	(7.4291E-02)
	ω	2.34705	2.35420	2.35161	2.35652	2.35237	
		(8.1419E-05)	(3.3065E-07)	(2.0900E-06)	(1.0267E-07)	(1.4888E-06)	(1.1634E-07)

Table 1: Average estimates, MSEs of the unknown parameters based on LSEs, LADEs, Huber-M estimators WLSEs for Model 1, when sample size $n = 25$.

PoO	Par	LSE	W1	W2	LADE	Huber-M	CRLB
10	A	3.95955	5.12643	4.42492	5.03074	5.00858	
		(1.0835E-01)	(1.8118E-02)	(1.3110E-02)	(1.1205E-02)	(1.5081E-02)	(9.9725E-03)
	B	5.20274	4.99915	5.07054	5.00868	4.82230	
		(2.1361E-01)	(1.9837E-02)	(1.5950E-02)	(1.1848E-02)	(1.6663E-02)	(9.9725E-03)
	ω	2.35581	2.35658	2.35537	2.35714	2.35475	
		(3.4162E-06)	(2.8447E-08)	(9.1110E-07)	(8.9013E-07)	(2.7312E-08)	(1.5617E-08)
20	A	5.25519	4.90373	5.15610	5.03353	5.04792	
		(6.5461E-02)	(2.2480E-02)	(2.4419E-02)	(3.1051E-02)	(2.2887E-02)	(1.9030E-02)
	B	4.48491	4.92250	4.72543	5.12113	5.08936	
		(4.6543E-02)	(3.0116E-02)	(3.5481E-02)	(3.4602E-02)	(3.0138E-023)	(1.9030E-02)
	ω	2.35639	2.35533	2.35652	2.35533	2.35715	
		(5.4817E-06)	(1.4345E-07)	(1.0607E-07)	(1.1459E-07)	(1.0027E-07)	(2.9801E-08)
30	A	4.62571	5.01334	4.84186	4.73935	4.98255	
		(4.3987E-01)	(3.8203E-02)	(3.4879E-02)	(3.7166E-02)	(3.5101E-02)	(2.8088E-02)
	B	5.93699	5.07028	5.45544	4.98759	4.92527	
		(6.7903E-01)	(4.0542E-02)	(3.0802E-02)	(3.6574E-02)	(3.6561E-02)	(2.8088E-02)
	ω	2.35796	2.35582	2.35582	2.35770	2.35533	
		(7.2514E-05)	(5.4153E-08)	(6.3839E-08)	(5.4461E-08)	(3.7403E-07)	(4.3986E-08)
40	A	5.67753	4.71064	4.66062	5.02145	4.86217	
		(2.5857E-01)	(4.4117E-02)	(4.7518E-02)	(4.1237E-02)	(4.8979E-02)	(3.7146E-02)
	B	4.32981	4.52517	5.08132	5.08132	5.08344	
		(4.4858E-01)	(4.2556E-02)	(5.6087E-02)	(6.6030E-02)	(7.0221E-02)	(3.7246E-02)
	ω	2.35456	2.35399	2.35820	2.35533	2.35775	
		(3.1689E-05)	(1.5236E-07)	(5.1751E-07)	(1.1933E-07)	(1.7882E-07)	(5.8171E-08)

Table 2: Average estimates, MSEs of the unknown parameters based on LSEs, LADEs, Huber-M estimators WLSEs for Model 1, when sample size $n = 50$.

PoO	Par	LSE	W1	W2	LADE	Huber-M	CRLB
10	A	5.01167	4.88565	4.96540	5.03539	5.01453	
		(1.4523E-02)	(8.3125E-03)	(9.1944E-03)	(9.2514E-03)	(8.0827E-03)	(6.6483E-03)
	B	4.80165	5.02731	4.85945	5.24511	5.07204	
		(3.9428E-02)	(8.4389E-04)	(9.9570E-03)	(9.9937E-03)	(8.1778E-03)	(6.6483E-03)
	ω	2.35582	2.35699	2.35639	2.35696	2.35653	
		(4.5207E-07)	(1.2961E-08)	(2.2071E-08)	(5.0383E-08)	(1.6717E-08)	(1.0411E-08)
20	A	5.03691	5.01489	5.01445	4.99248	5.00631	
		(2.3654E-02)	(2.2511E-03)	(2.0227E-03)	(5.5443E-03)	(3.2394E-03)	(1.2686E-03)
	B	5.25644	5.03542	5.10876	5.21226	5.08970	
		(3.5815E-02)	(2.2658E-03)	(2.1902E-03)	(4.5298E-03)	(6.1160E-03)	(1.2686E-03)
	ω	2.35718	2.35695	2.35699	2.35696	2.35657	
		(1.0481E-06)	(1.9137E-09)	(5.9594E-09)	(4.4251E-08)	(1.9038E-09)	(1.9867E-08)
30	A	4.86533	4.99095	4.99879	5.03682	4.81558	
		(1.8092E-02)	(3.2587E-03)	(5.1924E-03)	(4.3460E-03)	(3.4211E-03)	(1.8725E-03)
	B	4.32669	4.79688	4.54000	5.25040	5.16928	
		(4.5337E-02)	(4.1350E-03)	(5.1103E-03)	(6.2992E-03)	(4.8796E-03)	(1.8725E-03)
	ω	2.35559	2.35639	2.35576	2.35696	2.35713	
		(2.4119E-07)	(5.5573E-08)	(4.2669E-08)	(4.8863E-08)	(8.8329E-08)	(2.9324E-08)
40	A	4.99543	4.98781	5.00381	4.99430	4.98783	
		(4.2973E-02)	(4.4042E-03)	(5.5350E-03)	(6.9500E-03)	(5.3413E-03)	(2.4652E-03)
	B	4.78735	4.95765	4.82996	5.21511	5.02110	
		(4.5333E-02)	(4.7810E-03)	(6.9056E-03)	(4.6165E-03)	(4.4820E-03)	(2.4652E-03)
	ω	2.35557	2.35657	2.35581	2.35696	2.35652	
		(3.4201E-07)	(5.1216E-08)	(8.4210E-08)	(7.3085E-08)	(7.8347E-08)	(3.9654E-8)

Table 3: Average estimates, MSEs of the unknown parameters based on LSEs, LADEs, Huber-M estimators WLSEs for Model 1, when sample size $n = 75$.

PoO	Par	LSE	W1	W2	LADE	Huber-M	CRLB
10	A	5.14171	4.97862	5.03054	5.00202	4.99428	
		(9.0062E-03)	(6.1685E-03)	(9.1906E-03)	(7.3287E-03)	(6.1231E-03)	(4.9863E-03)
	B	5.72720	4.77987	5.45326	4.84940	5.02588	
		(9.2960E-03)	(5.8365E-03)	(7.0488E-03)	(8.2741E-03)	(6.7869E-03)	(4.9863E-03)
	ω	2.35715	2.35557	2.35658	2.35559	2.35582	
		(1.0364E-07)	(9.0070E-09)	(8.8495E-09)	(9.4756E-09)	(1.5207E-08)	(7.8085E-09)
20	B	4.93013	5.01670	4.71639	5.01158	4.99722	
		(4.9216E-02)	(1.1171E-02)	(1.0161E-02)	(1.3735E-02)	(1.2675E-02)	(9.5148E-03)
	B	4.43501	5.18614	4.75216	5.24827	4.85224	
		(3.18790E-02)	(1.4459E-02)	(1.1465E-02)	(1.1577E-02)	(1.1923E-02)	(9.5148E-03)
	ω	2.35314	2.35638	2.35481	2.35638	2.35563	
		(1.0041E-06)	(1.7652E-09)	(1.61068E-09)	(2.1046E-09)	(2.0574E-09)	(1.4900E-09)
30	A	5.49313	4.97302	5.04481	4.99391	5.00717	
		(6.4291E-02)	(3.5207E-02)	(1.9664E-02)	(3.9751E-02)	(5.7516E-02)	(1.4043E-02)
	B	5.71853	4.77940	5.54518	4.84933	5.07878	
		(5.1639E-02)	(4.8799E-02)	(2.9650E-02)	(2.2800E-02)	(6.3097E-02)	(1.4043E-02)
	ω	2.35714	2.35557	2.35701	2.35559	2.35577	
		(1.5212E-07)	(4.3398E-08)	(2.5242E-08)	(2.4641E-08)	(4.8318E-08)	(2.1192E-08)
40	A	4.73653	4.99248	4.98115	5.03827	4.98726	
		(6.9471E-02)	(3.5724E-02)	(2.5623E-02)	(2.4299E-04)	(2.7592E-04)	(1.8573E-02)
	B	5.17368	5.23577	4.94483	5.25465	4.71399	
		(5.0155E-02)	(2.5767E-02)	(2.0182E-02)	(2.4680E-02)	(2.1853E-02)	(1.8573E-02)
	ω	2.35396	2.35657	2.35457	2.35581	2.35552	
		(2.1597E-07)	(3.0050E-08)	(4.2528E-08)	(4.3212E-08)	(5.9343E-08)	(2.9085E-08)

Table 4: Average estimates, MSEs of the unknown parameters based on LSEs, LADEs, Huber-M estimators WLSEs for Model 1, when sample size $n = 100$.

PoO	Par	LSE	W1	W2	LADE	Huber-M	CRLB
20	A_1	10.13401	10.34824	10.95772	10.00410	10.01362	
		(7.7880E-02)	(4.2209E-02)	(5.1735E-02)	(6.1338E-02)	(5.9510E-02)	(3.8059E-02)
	B_1	10.65339	9.68928	9.63136	9.99074	10.10536	
		(4.2849E-01)	(5.6732E-02)	(5.3625E-02)	(9.5370E-02)	(6.1036E-02)	(3.8059E-02)
	ω_1	2.35892	2.35320	2.35312	2.35581	2.35555	
		(7.8432E-06)	(8.2622E-08)	(3.0675E-08)	(4.1569E-08)	(5.6185E-08)	(1.4900E-08)
	A_2	5.93336	5.30719	5.42681	5.00679	4.83355	
		(8.7188E-01)	(9.4348E-02)	(4.8263E-02)	(4.6563E-02)	(6.7697E-02)	(2.8765E-02)
	B_2	3.23776	3.58041	4.16570	4.99880	3.89633	
		(3.1043-01)	(7.0160E-02)	(6.9575E-02)	(5.0180E-02)	(7.2171E-02)	(2.8765E-02)
	ω_2	0.46201	0.45843	0.46527	0.47266	0.46558	
		(1.4280E-04)	(1.6267E-07)	(1.5069E-07)	(2.0792E-07)	(3.1485E-07)	(6.2354E-08)
40	A_1	9.61297	9.85146	9.59093	9.94676	11.1738	
		(1.5011E-00)	(2.1998E-01)	(1.6708E-01)	(2.8855E-01)	(2.3760-01)	(7.5356E-02)
	B_1	9.90004	9.18471	10.56675	10.14708	9.90220	
		(1.8098E-00)	(2.6345E-01)	(2.2156E-01)	(2.1588E-01)	(3.5742E-01)	(7.5356E-02)
	ω_1	2.35771	2.35395	2.35714	2.35639	2.35099	
		(1.5608E-06)	(5.4766E-08)	(9.0668E-08)	(5.9367E-08)	(2.5720E-08)	(2.9651E-08)
	A_2	5.37362	4.92582	3.91055	5.01451	4.12929	
		(1.3937E-00)	(9.4625E-02)	(9.1862E-02)	(1.0445E-01)	(9.5858E-02)	(5.7242E-02)
	B_2	3.69551	3.14475	2.57799	4.99543	4.02975	
		(1.7011E-00)	(9.4420E-02)	(8.8663E-02)	(2.3174E-01)	(9.4059E-02)	(5.7242E-02)
	ω_2	0.52003	0.47601	0.50343	0.47288	0.47689	
		(2.3795E-03)	(2.3196E-06)	(1.0400E-06)	(2.5883E-06)	(3.2491E-06)	(1.2408E-07)

Table 5: Average estimates, MSEs of the unknown parameters based on LSEs, LADEs, Huber-M estimators WLSEs for Model 2, when sample size $n = 25$.

PoO	Par	LSE	W1	W2	LADE	Huber-M	CRLB
20	A_1	10.73565	9.93063	9.96949	9.99449	10.00027	
		(5.40876E-02)	(1.8754E-02)	(1.1443E-02)	(3.3247E-02)	(5.0221E-02)	(9.5148E-03)
	B_1	10.65249	9.84759	10.02020	9.93136	9.97467	
		(4.2528E-02)	(2.3437E-02)	(1.1036E-02)	(4.6797E-02)	(5.8203E-02)	(9.5148E-03)
	ω_1	2.35653	2.35576	2.35639	2.35581	2.35581	
		(1.5139E-07)	(1.0054E-08)	(5.3709E-08)	(1.3098E-08)	(1.1510E-08)	(3.7251E-09)
	A_2	5.31684	4.99579	5.01345	4.98735	4.99725	
		(1.0065E-01)	(1.6815E-02)	(1.9062E-02)	(1.6767E-02)	(1.6070E-02)	(7.9125E-03)
	B_2	5.63272	4.83475	4.99428	4.98303	4.98297	
		(4.0033E-01)	(2.7403E-02)	(3.1795E-02)	(2.7985E-02)	(2.8404E-02)	(7.9125E-03)
	ω_2	0.46970	0.47111	0.47035	0.47094	0.47087	
		(2.2063E-06)	(2.6236E-08)	(6.8193E-08)	(6.3230E-08)	(5.3443E-08)	(1.5589E-08)
40	A_1	9.96230	10.11183	10.01066	9.99738	9.93436	
		(1.4237E-01)	(8.2514E-02)	(7.1255E-02)	(8.5320E-02)	(6.2844E-02)	(1.8839E-02)
	B_1	9.75040	10.21666	9.98802	9.96949	9.73995	
		(9.2135E-02)	(4.6854E-02)	(7.3989E-02)	(9.2146E-02)	(6.7881E-02)	(1.8839E-02)
	ω_1	2.35462	2.35638	2.35552	2.35563	2.35563	
		(2.5124E-06)	(1.7195E-08)	(6.3379E-08)	(1.4826E-08)	(1.2299E-08)	(7.4128E-09)
	A_2	4.92425	4.98772	5.00392	4.99877	4.95589	
		(5.7233E-01)	(8.4761E-02)	(7.8590E-02)	(8.4267E-02)	(7.9360E-02)	(1.4311E-02)
	B_2	5.06412	5.05626	4.94287	4.97964	4.85422	
		(4.1558E-01)	(6.1890E-02)	(6.3032E-02)	(7.1473E-02)	(8.1208E-02)	(1.4311E-02)
	ω_2	0.46825	0.47145	0.46967	0.47065	0.47065	
		(8.7547E-06)	(9.1515E-08)	(8.5169E-08)	(9.8791E-08)	(1.8705E-07)	(3.1025E-08)

Table 6: Average estimates, MSEs of the unknown parameters based on LSEs, LADEs, Huber-M estimators WLSEs for Model 2, when sample size $n = 100$.

Model	LADE	Huber-M
1	4.7453	4.8326
2	7.2217	7.6598

Table 7: Ratio of the computational time of the LAD and Huber-M estimators with respect to the WLSEs.

is $w_1(t)$, in Table 7. In this case we have computed the ratio of the time taken to compute LAD estimators and Huber-M estimators to the WLSEs for 10000 replications The results are presented when $n = 100$ and PoO is 20. It is clear that as the number of parameters increases, the ratio increases.

5 Weight Function

In the previous section it has been observed that the performances of the WLSEs depend on the weight function. In this section we discuss about the choice of the weight function in practice and also we illustrate the proposed method by analyzing a real data set.

5.1 How to Choose the Weight Function?

The main idea about the WLSEs is to put less weight on the residual where the outliers are present. Note that it is the same idea as the Huber-M estimators. In case of Huber-M estimators also less weight is assigned where the outliers are present. Now if it is known apriori where the outliers are present, then one can choose the weight function accordingly. For example if it is known that the outliers are present at the beginning of the data sequence, then one can choose a weight function which puts less weight at the beginning, i.e. a non-decreasing weight function. Similarly if the outliers are present towards the end of the data sequence, then one can choose the weight function accordingly may be a non-increasing weight function. But in practice, one may not know apriori where the outliers are present. Hence, the position of the outliers are unknown. In this case we suggest to choose a class of weight functions, where each weight function puts different weights at different places. For example some of the weight functions put less weight at the beginning, and some put less weight in the middle etc. The main idea is to choose different shapes of the weight functions in this class. The possible choices are provided in the data analysis Section 5.2. Now compute the weighted residual sums of squares in each case, and finally choose that weight function which provides the lowest weights residual sums of squares. Since the LS estimator is also a special case of the WLS estimator when $w(t) = 1$, we suggest to keep this weight function also in the class. For any practical use, we suggest to keep 6-8 weight functions in the class, the details are explained using a real data set in the next section.

5.2 Analysis of a Data Set

In this section we present the analysis of a real data set to show how the proposed method works in practice. We have considered the airlines passenger data which gives information of monthly passenger totals of a US airline from 1949 to 1960. It is available in https://www.kaggle.com/datasets/chirag19/air-passengers. Four outliers have been added in the data set at random. The logarithm of the data set (in presence of outliers) and the corresponding periodogram function have been plotted in Figure 1. From Figure 1 it seems the model (2) may be used to analyze this data set with p =1 and q = 5. The details will be discussed later.

Since it is not known in advance where the outliers are present we have used the following six weight functions which put different weights at different parts of the data sequences. The different weight functions for $0 \leq t \leq 1$ are as follows:

$$w_1(t) = 1, \quad w_2(t) = 1+10.0(t-0.75)^2, \quad w_3(t) = 1-3.0(t-0.5)^2, \quad w_4(t) = 1+0.01(t-0.5)^2$$

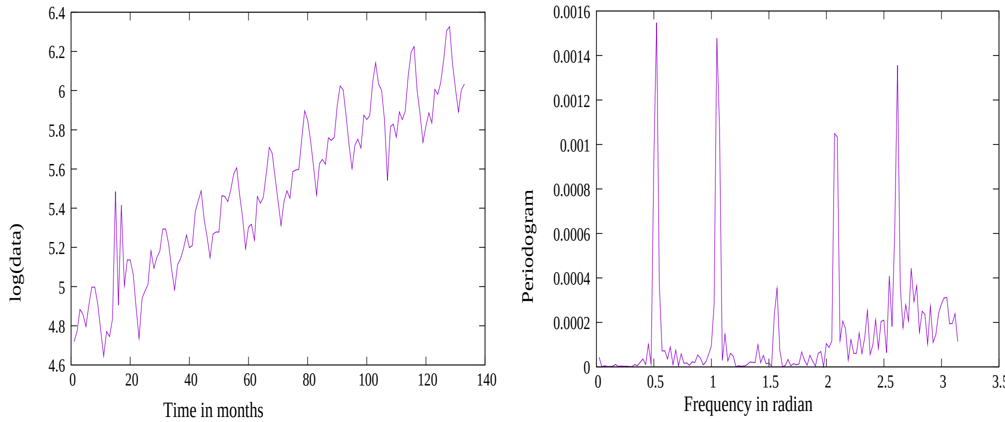

Figure 1: The logarithm of the airlines passenger data set when the outliers are present and the corresponding periodogram function: (a) log(data) (b) periodogram.

$$w_5(t) = 1 + 0.01(t - 0.25)^2 + 10.0(t - 0.75)^2, \quad \text{and} \quad w_6(t) = \begin{cases} 1 & \text{if} \quad 0 \le t \le \frac{1}{4} \\ 2 & \text{if} \quad \frac{1}{4} \le t \le \frac{3}{4} \\ 1 & \text{if} \quad \frac{3}{4} \le t \le 1. \end{cases}$$

Note that the weight function $w(t) = 1$ is a member of this class. Among the different weight functions all the five weight functions $w_2(t)$ to $w_6(t)$ provide better fit in terms of lower residual sums of squares than $w_1(t)$. Among the five weight functions, $w_6(t)$ provide the best fit. The original (with outliers) and the predicted plots are provided in Figure 2. For comparison purposes we have provided the predicted plots based on LAD and also Huber-M estimators also in Figures 3. All the predicted curves match very well. Note that the implementation of the proposed WLS method is quite simple in this case, it involves differencing and solving five one dimensional optimization problem, where as both LAD and Huber-M estimators involve solving seventeen dimensional optimization problem. The computational times of both LAD and Huber-M estimators are significantly more than the the WLS estimators.

Another natural question is whether the above choice of p and q is proper or not. For that we have used Akike Information Criterion (AIC) and Bayes Information Criterion (BIC), see

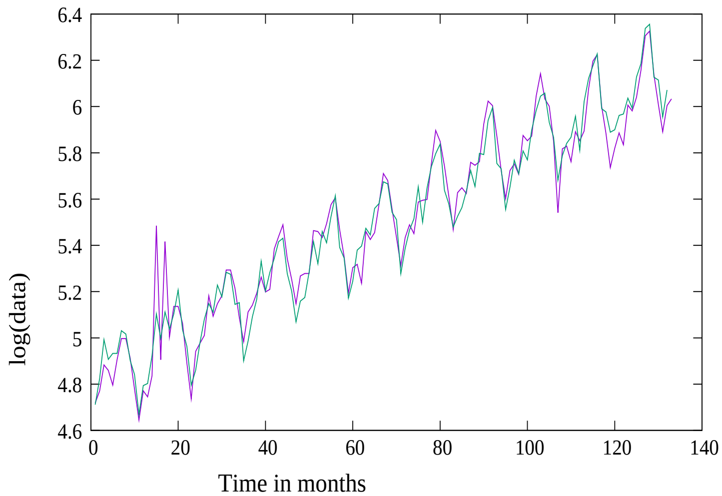

Figure 2: The log(data) and the log(predicted) values based on the WLSEs are provided. Here the 'purple' line indicates the log(data) and the 'green' line indicates log(predicted) values.

for example Nandi and Kundu [4], for this model. For the given p and q, AIC and BIC will take the following forms:

$$\mathrm{AIC}(p,q) = n\ln(R_{p,q}) + 2(p+q) \tag{21}$$

$$\mathrm{BIC}(p,q) = n\ln(R_{p,q}) + 2(p+q)\ln n. \tag{22}$$

Here $R_{p,q}$ denotes the residual sums of squares of the fitted model with p and q numbers polynomial and sinusoidal regression parameters, respectively. We will choose that particular p and q which will minimize AIC and BIC values. We have taken different choices of p and q and reported the results in Table 8. Both AIC and BIC suggest the choice of $p = 1$ and $q = 5$. Hence, it is justified to choose those particular values of p and q.

6 Conclusion

In this paper we have considered a robust estimation procedure of the parameters of a polynomial sinusoidal regression model in presence of outliers. We have used a new weighted

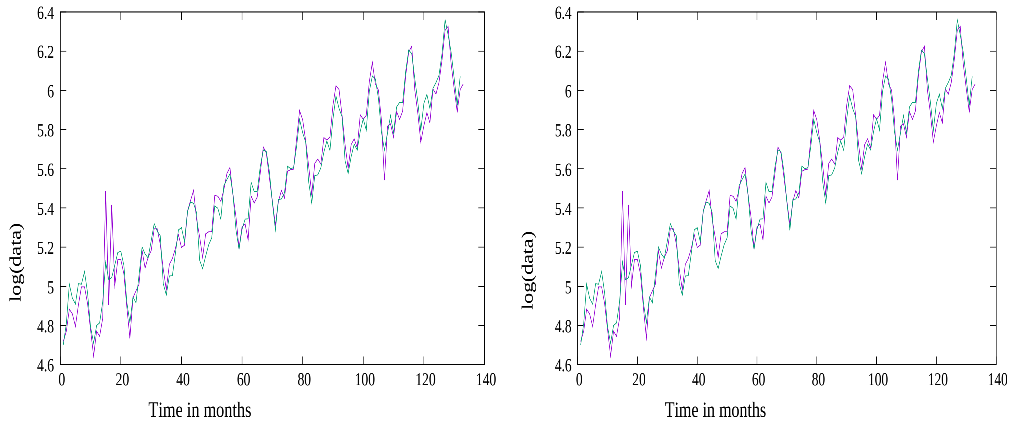

Figure 3: The log(data) and the log(predicted) values based on: (a) LAD (b) Huber-M, estimators. In both the caes, the 'purple' lines indicate the log(data) and the 'green' lines indicate log(predicted) values.

(p,q)	(1,3)	(1,4)	(1,5)	(1,6)	(2,3)	(2,4)	(2,5)	(2,6)
AIC	169.98	168.22	165.03	167.29	169.87	171.37	173.76	174.23
BIC	305.45	302.21	297.31	300.21	301.54	303.87	306.54	307.87

Table 8: AIC and BIC values for different choices of p and q.

least squares method to estimate the unknown parameters instead of the standard least absolute estimators or Huber's M-estimators. There are two advantages of the proposed method: (a) it can be implemented very easily even for any number of sinusoidal components and (b) the theoretical properties of the proposed estimators can be established under the same set of assumptions as it is needed in case of the least squares estimators. Extensive simulations suggest that the proposed method works at par with the LAD estimators and Huber-M estimators in-terms of the MSEs in presence of outliers. We have suggested how to choose the weight function in practice and we have illustrated the method using a real data set. The results are quite satisfactory. Note that although we have proposed the method for linear polynomial only, but it can be easily extended for a general p-degree polynomial also.

Acknowledgements:

The author would like to thank two unknown reviewers for their constructive suggestions which has helped to improve the manuscript significantly.

Conflict of Interest Statement

The author does not have any conflict of interest in writing this manuscript.

Appendix A:

To prove Theorem 1, we need the following lemmas.

Lemma 0: For $0 < \lambda \neq \theta < \pi$ and for $w(t)$ satisfies (4)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) \cos^2(t\lambda) = \frac{1}{2}; \tag{23}$$

$$\lim_{n\to\infty} \frac{1}{n^{k+1}} \sum_{t=1}^{n} t^k w\left(\frac{t}{n}\right) \cos^2(t\lambda) = \frac{1}{2(k+1)}; \qquad k = 1, 2, \ldots, \tag{24}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) \cos(t\lambda)\cos(t\theta) = 0, \qquad i \neq j, \tag{25}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) \sin^2(t\lambda) = \frac{1}{2}, \tag{26}$$

$$\lim_{n\to\infty} \frac{1}{n^{k+1}} \sum_{t=1}^{n} t^k w\left(\frac{t}{n}\right) \sin^2(t\lambda) = \frac{1}{2(k+1)}, \qquad k = 1, 2, \ldots, \tag{27}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) \sin(t\lambda)\sin(t\theta) = 0; \qquad i \neq j, \tag{28}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} w\left(\frac{t}{n}\right) \sin(t\lambda)\cos(t\theta) = 0, \tag{29}$$

Proof: See Mangulis [19]. ∎

Proof of Theorem 1: Observe that

$$\widetilde{\beta}_1 = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} w\left(\frac{i}{n}\right)} \sum_{t=1}^{n} \frac{1}{n} w\left(\frac{t}{n}\right) z(t).$$

Note that

$$\frac{1}{n}\sum_{i=1}^{n} w\left(\frac{i}{n}\right) \longrightarrow \int_0^1 w(t)dt$$

and for $k = 0, 1, \ldots,$

$$\frac{1}{n^{k+1}} \sum_{t=1}^{n} t^k e_d(t) \longrightarrow 0, \quad \text{a.s.}$$

due to strong law of large numbers of stochastic process, see Fuller[20]). Therefore, by using Lemma 0, the result follows. ∎

To prove Theorem 3, we need the following lemmas.

Lemma 1: If $e(t)$'s satisfy Assumption 1, then for m = 0,1,2, ...,

$$\lim_{n\to\infty} \sup_{\theta} \left| \frac{1}{n^{m+1}} \sum_{t=1}^{n} e(t) t^m \cos(t\theta) \right| = 0 \quad \text{a.s..}$$

The same result holds when $\cos(t\theta)$ is replaced by $\sin(t\theta)$.

Proof: For $m = 0, 1, 2$, see Lemma 1 and its associated corollaries of Kundu [21]. For general m, the results can be obtained along the same line. ∎

Lemma 2: If $e(t)$'s satisfy Assumption 1, and $w(t)$ satisfies (3), then

$$\lim_{n\to\infty} \sup_{\theta} \left| \frac{1}{n} \sum_{t=1}^{n} e(t) w\left(\frac{t}{n}\right) \cos(t\theta) \right| = 0 \quad \text{a.s..}$$

The same result holds when $\cos(t\theta)$ is replaced by $\sin(t\theta)$.

Proof: Since $w(t)$ is a polynomial then using Lemma 1, the result immediately follows. ∎

Lemma 3: Let $S_{\delta,M} = \{\mathbf{\Gamma}; \mathbf{\Gamma} = (A, B, \theta), |\mathbf{\Gamma} - \mathbf{\Gamma}^0| \geq 3\delta, |A| \leq M, |B| \leq M\}$. If for any $\delta > 0$, and for some $M < \infty$,

$$\lim_{n\to\infty} \inf_{\mathbf{\Gamma}\in S_{\delta,M}} \frac{1}{n}\left[Q(\mathbf{\Gamma}) - Q(\mathbf{\Gamma}^0)\right] > 0, \qquad a.s.,$$

then $\widetilde{A}$, $\widetilde{B}$ and $\widetilde{\lambda}$ are strongly consistent estimators of A^0, B^0 and λ^0, respectively. Here $Q(\mathbf{\Gamma}) = Q(A, B, \theta)$ is same as defined in (8).

Proof: The proof mainly follows by contradiction argument. The details are avoided. ∎

Proof of Theorem 2: Observe that $Q(\mathbf{\Gamma})$ can be written as follows:

$$Q(\mathbf{\Gamma}) = \sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left[z(t) - \widetilde{\beta}_1 - C_1(\mathbf{\Gamma})\sin(\theta t) - C_2(\mathbf{\Gamma})\cos(\theta t)\right]^2,$$

where $C_1(\mathbf{\Gamma}) = -A\sin(\theta) - 2B\sin^2(\theta/2)$ and $C_2(\mathbf{\Gamma}) = B\sin(\theta) - 2A\sin^2(\theta/2)$. Therefore,

$$\frac{1}{n}\left[Q(\mathbf{\Gamma}) - Q(\mathbf{\Gamma}^0)\right] = f_n(\mathbf{\Gamma}) + g_n(\mathbf{\Gamma}) + o_n(1),$$

here

$$\begin{aligned}
f_n(\mathbf{\Gamma}) &= \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left[C_1(\mathbf{\Gamma}^0)\sin(\theta^0 t) - C_1(\mathbf{\Gamma})\sin(\theta t) + C_2(\mathbf{\Gamma}^0)\cos(\theta^0 t) - C_2(\mathbf{\Gamma})\cos(\theta t)\right]^2,\\
g_n(\mathbf{\Gamma}) &= \frac{2}{n}\sum_{t=1}^{n} e_d(t) w\left(\frac{t}{n}\right)\left[C_1(\mathbf{\Gamma}^0)\sin(\theta^0 t) - C_1(\mathbf{\Gamma})\sin(\theta t) + C_2(\mathbf{\Gamma}^0)\cos(\theta^0 t) - C_2(\mathbf{\Gamma})\cos(\theta t)\right]^2,
\end{aligned}$$

and $o_n(1)$ converges to 0 almost surely as $n \to \infty$. Since $e_d(t)$ satisfies Assumption 1, and A, B are bounded in $S_{\delta,M}$, it follows from Lemma 2 that

$$\lim_{n\to\infty} \sup_{\mathbf{\Gamma}\in S_{\delta,M}} |g_n(\mathbf{\Gamma})| = 0 \quad a.s. \tag{30}$$

Now we would like to show that

$$\lim_{n\to\infty} \inf_{\mathbf{\Gamma}\in S_{\delta,M}} f_n(\mathbf{\Gamma}) > 0.$$

We consider three different sets:

$$\begin{aligned}
S_{\delta,M,1} &= \{\mathbf{\Gamma} : \mathbf{\Gamma} = (A, B, \theta), |A - A^0| \geq \delta, |A| \leq M, |B| \leq M\}, \\
S_{\delta,M,2} &= \{\mathbf{\Gamma} : \mathbf{\Gamma} = (A, B, \theta), |B - B^0| \geq \delta, |A| \leq M, |B| \leq M\}, \\
S_{\delta,M,3} &= \{\mathbf{\Gamma} : \mathbf{\Gamma} = (A, B, \theta), |\omega - \omega^0| \geq \delta, |A| \leq M, |B| \leq M\}.
\end{aligned}$$

Now let us consider

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{\mathbf{\Gamma}\in S_{\delta,M,1}} \frac{1}{n}\left[Q(\mathbf{\Gamma}) - Q(\mathbf{\Gamma}^0)\right] = \liminf_{n\to\infty} \inf_{\mathbf{\Gamma}\in S_{\delta,M,1}} f_n(\mathbf{\Gamma}) \\
= &\liminf_{n\to\infty} \inf_{|A-A^0|>\delta} \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left[C_1(\mathbf{\Gamma}^0)\sin(\theta^0 t) - C_1(\mathbf{\Gamma})\sin(\theta t) + C_2(\mathbf{\Gamma}^0)\cos(\theta^0 t) - C_2(\mathbf{\Gamma})\cos(\theta t)\right]^2 \\
= &\liminf_{n\to\infty} \inf_{|A-A^0|>\delta} \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left[-\left\{A^0\sin(\theta^0) + 2B^0\sin^2(\theta/2)\right\}\sin(\theta^0 t)\right. \\
&-\left\{A\sin(\theta t) + 2B\sin^2(\theta/2)\right\}\sin(\theta t) + \left\{B^0\sin(\theta^0) - 2A^0\sin^2(\theta^0/2)\right\}\cos(\theta^0 t) \\
&\left.-\left\{B^0\sin(\theta) - 2A\sin^2(\theta^0/2)\right\}\sin(\theta t)\right]^2 \\
= &\lim_{n\to\infty} \inf_{|A-A^0|>\delta} \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left[(A-A^0)^2\left\{\sin(\theta^0)\sin(\theta^0 t) + 2\sin^2(\theta^0/2)\cos(\theta^0 t)\right\}\right]^2 \\
\geq &\ \delta^2 \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} w\left(\frac{t}{n}\right)\left[\left\{\sin^2(\theta^0)\sin^2(\theta^0 t) + 4\sin^4(\theta^0/2)\cos^2(\theta^0 t)\right\}\right] \\
= &\ \frac{\delta^2}{2}\left[\sin^2(\theta^0) + 4\sin^4(\theta^0/2)\right] > 0, \quad \forall \quad \theta^0 \in (0, \pi).
\end{aligned}$$

Similarly, it can be shown that

$$\liminf_{n\to\infty} \inf_{\mathbf{\Gamma}\in S_{\delta,M,2}} f_n(\mathbf{\Gamma}) > 0 \quad \text{and} \quad \liminf_{n\to\infty} \inf_{\mathbf{\Gamma}\in S_{\delta,M,3}} f_n(\mathbf{\Gamma}) > 0.$$

Proof of Theorem 3:

We use $Q'(\mathbf{\Gamma})$ and $Q''(\mathbf{\Gamma})$ to denote the 1×3 vector of first derivatives and 3×3 matrix of second derivatives of $Q(\mathbf{\Gamma})$ as defined in (8). Now expanding $Q'(\widetilde{\mathbf{\Gamma}})$ around the true parameter value $\mathbf{\Gamma}^0$ by Taylor series, we obtain

$$Q'(\widetilde{\mathbf{\Gamma}}) - Q'(\mathbf{\Gamma}^0) = (\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma}^0)Q''(\overline{\mathbf{\Gamma}}), \tag{31}$$

where $\overline{\mathbf{\Gamma}}$ is a point on the line joining the points $\widetilde{\mathbf{\Gamma}}$ and $\mathbf{\Gamma}^0$. Since, $Q'(\mathbf{\Gamma}^0) = \mathbf{0}$, the above equation (31) can be written as

$$(\widetilde{\mathbf{\Gamma}} - \mathbf{\Gamma}^0)\boldsymbol{D} = -[Q'(\mathbf{\Gamma}^0)\boldsymbol{D}^{-1}][\boldsymbol{D}^{-1}Q''(\overline{\mathbf{\Gamma}})\boldsymbol{D}^{-1}]^{-1}. \tag{32}$$

Using the Central Limit Theorem of linear processes, see for example Fuller [20] and Lemma 0, it can be shown that

$$Q'(\mathbf{\Gamma}^0)\boldsymbol{D}^{-1} \xrightarrow{d} N_3(\mathbf{0}, 2\sigma^2\xi(\theta^0)\mathbf{\Sigma}(\mathbf{\Gamma}^0))$$

and

$$\boldsymbol{D}^{-1}Q''(\overline{\mathbf{\Gamma}})\boldsymbol{D}^{-1} \longrightarrow \mathbf{\Psi}(\mathbf{\Gamma}^0) \quad a.s.,$$

hence the result follows. ∎

Appendix B

To prove Theorem 4, we need the following Lemmas, similar to Lemma 3.

Lemma 4: Let $S^1_{\delta,M} = \{\mathbf{\Gamma}_1; \mathbf{\Gamma}_1 = (A_1, B_1, \theta_1), |\mathbf{\Gamma}_1 - \mathbf{\Gamma}^0_1| \geq 3\delta, |A_1| \leq M, |B_1| \leq M\}$. If for any $\delta > 0$, and for some $M < \infty$,

$$\lim_{n\to\infty} \inf_{\mathbf{\Gamma}_1 \in S^1_{\delta,M}} \frac{1}{n}\left[Q_1(\mathbf{\Gamma}_1) - Q_1(\mathbf{\Gamma}^0_1)\right] > 0, \qquad a.s.,$$

then $\widetilde{A}_1$, $\widetilde{B}_1$ and $\widetilde{\lambda}_1$ are strongly consistent estimators of A^0_1, B^0_1 and λ^0_1, respectively. Here $Q_1(\mathbf{\Gamma}_1) = Q_1(A_1, B_1, \theta_1)$ is same as defined in (15).

Proof of Lemma 4: This proof also can be obtained by contradiction. ∎

Lemma 5: Let $S^2_{\delta,M} = \{\mathbf{\Gamma}_2; \mathbf{\Gamma}_2 = (A_2, B_2, \theta_2), |\mathbf{\Gamma}_2 - \mathbf{\Gamma}^0_2| \geq 3\delta, |A_2| \leq M, |B_2| \leq M\}$. If for any $\delta > 0$, and for some $M < \infty$,

$$\lim_{n\to\infty} \inf_{\mathbf{\Gamma}_2 \in S^2_{\delta,M}} \frac{1}{n}\left[Q_2(\mathbf{\Gamma}_2) - Q_2(\mathbf{\Gamma}^0_2)\right] > 0, \qquad a.s.,$$

then $\widetilde{A}_2$, $\widetilde{B}_2$ and $\widetilde{\lambda}_2$ are strongly consistent estimators of A_2^0, B_2^0 and λ_2^0, respectively. Here $Q_2(\mathbf{\Gamma}_2) = Q_2(A_2, B_2, \theta_2)$ is same as defined in (16).

Proof of Lemma 5: This proof also can be obtained by contradiction. ∎

Proof of Theorem 4:

Similar to Theorem 2, it can be shown that

$$\lim_{n\to\infty} \inf_{\mathbf{\Gamma}_1 \in S^1_{\delta,M}} \frac{1}{n}\left[Q_1(\mathbf{\Gamma}_1) - Q_1(\mathbf{\Gamma}_1^0)\right] > 0, \qquad a.s..$$

Hence, $\widetilde{A}_1$, $\widetilde{B}_1$ and $\widetilde{\theta}_1$ are strongly consistent estimators. If $\widetilde{\mathbf{\Gamma}}_1$ minimizes $Q_1(\mathbf{\Gamma}_1)$, then expanding $Q_1(\widetilde{\mathbf{\Gamma}}_1)$ around the true value of $\mathbf{\Gamma}_1^0$ by multivariate Taylor series, and following similar procedure as in the proof of Theorem 3, it can be shown that

$$\widetilde{A}_1 = A_1^0 + o_n(1), \qquad \widetilde{B}_1 = B_1^0 + o_n(1), \qquad \widetilde{\theta}_1^0 = \theta_1^0 + o_n(1/n), \tag{33}$$

here $o_n(1/n)$ indicates a term such that $no_n(1/n)$ converges to zero almost surely as $n \to \infty$. Using (33) in the expression of $z_1(t)$ in $Q_2(A_2, B_2, \theta_2)$ of equation (16) and Lemma 5, it can be shown along the same line as above that $\widetilde{A}_2$, $\widetilde{B}_2$ and $\widetilde{\theta}_2$ are strongly consistent estimators. Proceedings along the same line the result can be obtained for general q. ∎

References

[1] Li, R. and Nie, L. (2007), "A new estimation procedure for a partially nonlinear model via a mixed-effect approach", *The Canadian Journal of Statistics*, vol. 35, no. 3, 399 – 411.

[2] Stoica, P. (1993), "List of references on spectral line analysis", *Signal Processing*, vol. 31, 329–340.

[3] Surana, K., Mitra, S., Mitra, A. and Stoica, P. (2016), "Estimating the order of sinusoidal models using the adaptively penalized likelihood approach: Large sample consistency properties", *Signal Processing*, vol. 128, 204 – 211.

[4] Nandi, S. and Kundu, D. (2020), *Statistical Signal Processing-Frequency Estimation*, Second edition, Springer, Singapore, ISBN: 978-981-15-6279-2; 978-981-15-6280-8.

[5] Qian, X., Liang, W., Chen, Z., Zhang, Z., Zhang, X., Chen, Z. and Lin, Y. (2023), "High resolution adaptive estimator of sinusoidal based on covariance matrix reconstruction", *Signal Processing*, vol. 209, 109034.

[6] Trapero, J.R., Sira-Ramirez, H. and Batlle, V.F. (2007), "An algebraic frequency estimator for a biased and noisy sinusoidal signal", *Signal Processing*, vol. 87, 1188 – 1201.

[7] Perron, P., Shintani, M. and Yabu, T. (2020), "Trigonometric Trend Regressions of Unknown Frequencies with Stationary or Integrated Noise", https://www.bu.edu/econ/files/2020/05/FourierSup20200131.pdf.

[8] Abadir, K.M., Distaso, W. and Giraitis, L. (2010), "An I(d) model with trend and cycles", *Journal of Econometrics*, vol. 163, 186 – 199.

[9] Nandi, S. and Kundu, D. (2013), "Estimation of parameters of partially sinusoidal frequency model", *Statistics*, vol. 47, no. 1, 45 – 60.

[10] Raspaud, M., Marchand, S., Girin, L. (2005), "A Generalized Polynomial And Sinusoidal Model For Partial Tracking And Time Stretching", *Proceedings of the Digital Audio Effects (DAFx05) Conference*, Sep 2005, Spain. pp.24–29.

[11] Li, R. and Nie, L. (2008), "Efficient statistical inference procedures for partially nonlinear models and their applications", *Biometrics*, vol. 64, 904 – 911.

[12] Rigdon, S.E., Turabelidze, G. and Jahanpour, E. (2014), "Trigonometric Regression for Analysis of Public Health Surveillance Data", *Journal of Applied Mathematics*, Article ID 673293, pp 13.

[13] Lau, T-S and Studden, W.J. (1985), "Optimal design for trigonometric and polynomial regression using canonical moments", *Annals of Statistics*, vol. 13, 383 – 394.

[14] Eubank, R.L. and Speckman, P. (1990), "Curve fitting by polynomial-trigonometric regression", *Biometrika*, vol. 77, 1–9.

[15] Eubank, R. L. and Speckman, P. (1991), "Convergence rates for trigonometric and polynomial-trigonometric regression estimators", *Statistics and Probability Letters*, vol. 11, 119–124.

[16] Kim, T.S., Kim, H.K. and Choi, S.H. (2000), "Asymptotic properties of LAD estimators of a nonlinear time series regression model", *Journal of the Korean Statistical Society*, vol. 29, 187 – 199.

[17] Huber, P. (1981), *Robust Statistics*, John Wiley, New York.

[18] Kundu, D. and Grover, R. (2023), "Robust estimators of two dimensional sinusoidal model parameters", *Signal Processing*, vol. 212, Article No. 109167.

[19] Mangulis, V. (1965), *Handbook of series for scientists and engineers*, Academic Press, New York.

[20] Fuller, W. A. (1976), *Introduction to Statistical Time Series*, John Wiley and Sons, New York.

[21] Kundu, D. (1997), "Asymptotic Properties of the Least Squares Estimators of Sinusoidal Signals", *Statistics*, vol. 30, 221 – 238.