Fundamental Frequency and its Harmonics Model: A Robust Method of Estimation

Debasis Kundu *

Abstract

In this paper we have proposed a novel robust method of estimation of the unknown parameters of a fundamental frequency and its harmonics model. Although the least squares estimators (LSEs) or the periodogram type estimators are the most efficient estimators, it is well known that they are not robust. In presence of outliers the LSEs are known to be not efficient. In presence of outliers, robust estimators like least absolute deviation estimators (LADEs) or Huber's M-estimators (HMEs) may be used. But implementation of the LADEs or HMEs are quite challenging, particularly if the number of component is large. Finding initial guesses in the higher dimensions is always a non-trivial issue. Moreover, theoretical properties of the robust estimators can be established under stronger assumptions than what are needed for the LSEs. In this paper we have proposed novel weighted least squares estimators (WLSEs) which are more robust compared to the LSEs or periodogram estimators in presence of outloers. The proposed WLSEs can be implemented very conveniently in practice. It involves in solving only one non-linear equation. We have established the theoretical properties of the proposed WLSEs. Extensive simulations suggest that in presence of outliers, the WLSEs behave better than the LSEs, periodogram estimators, LADEs and HMEs. The performance of the WLSEs depend on the weight function, and we have discussed how to choose the weight function. We have analyzed one synthetic data set to show how the proposed method can be used in practice.

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 consider the estimation of the unknown parameters of the following fundamental frequency and its harmonics model:

$$y(n) = \sum_{k=1}^{p} \{A_k^0 \cos(nk\lambda^0) + B_k^0 \sin(nk\lambda^0)\} + X(n); \qquad n = 1, \ldots, N. \tag{1}$$

Here $A_k^{0^2} + B_k^{0^2} > 0$, for $k = 1, \ldots, p$, are unknown amplitudes and λ^0 is the fundamental frequency. The number of component p is assumed to be known. The error components $X(n)$'s are assumed to be independent and identically distributed random variables with zero mean and finite variance. The problem is to estimate the unknown parameters; namely λ^0 and also A_k^0 and B_k^0, for $k = 1, \ldots, p$.

This particular model is a very important model in the statistical signal processing and time series analysis. This has been used quite successfully in analyzing musical signals. Researchers in this field are particularly interested in determining what particular characteristics of the sound produced by musical instruments permit humans to distinguish one instrument from the other. For many orchestral instruments, such as clarinet, physical modeling suggests that within short segments, the frequencies are harmonically related as in (1), see Fletcher and Rossing [1]. Kundu and Nandi [2, 3] observed that male vowel sound also can be modelled very effectively using (1).

The model was originally introduced by Walker [4]. Since then several attempts have been made in developing different efficient estimation procedures of the fundamental frequency, the amplitudes and also the number of components. See for example Nielsen et al. [5], Jensen et al. [6], Cheveigńe and Kawahara [7], Li et al. [8], Kundu and Nandi [2, 3, 9, 10, 11], Qiu et al. [12], Fu et al. [13], Chen et al. [14], Rengaswamy et al. [15]. The most efficient estimators are the least squares estimators (LSEs), which can be obtained by minimizing the residual sums of squares, i.e.

$$R(A,B,\lambda) = \sum_{n=1}^{N} \left(y(n) - \sum_{k=1}^{p} \{A_k \cos(nk\lambda) + B_k \sin(nk\lambda)\} \right)^2 . \tag{2}$$

Here, $A = (A_1, \ldots, A_p)^\top$ and $B = (B_1, \ldots, B_p)^\top$. The periodogram estimator of λ can be obtained by maximizing the periodogram function defined as follows:

$$I(\lambda) = \frac{1}{N} \sum_{j=1}^{p} \left| \sum_{n=1}^{N} y(n) e^{inj\lambda} \right|^2 . \tag{3}$$

The maximization is usually performed over the Fourier frequencies. It has been shown by Nandi and Kundu [3] that in case of LSEs, the asymptotic variances of the linear parameter estimators have the convergence rate N^{-1} and the frequency parameter estimator has the convergence rate N^{-3}. Moreover, the asymptotic variances reach the corresponding Cramer-Rao lower bound when the errors are independent and identically distributed normal random variables.

Although the LSEs are the most efficient estimators, they are quite susceptible to outliers. Even in the presence of few outliers, the performances of the LSEs deteriorate drastically. They are known to be not robust estimators. In this respect the most natural robust estimators will be the least absolute deviation estimators (LADEs), which can be obtained by minimizing

$$Q(A,B,\lambda) = \sum_{n=1}^{N} \left| y(n) - \sum_{k=1}^{p} \{A_k \cos(nk\lambda) + B_k \sin(nk\lambda)\} \right| , \tag{4}$$

see for example Kim et al. [16]. Alternatively, Huber's M-estimators (HMEs), Huber [17], also can be used as efficient robust estimators, and they can be obtained by minimizing

$$H(A,B,\lambda) = \sum_{n=1}^{N} \rho \left(y(n) - \sum_{k=1}^{p} \{A_k \cos(nk\lambda) + B_k \sin(nk\lambda)\} \right) . \tag{5}$$

Different forms of $\rho(\cdot)$ functions have been suggested in the literature. Some of the standard $\rho(\cdot)$ functions for some $c > 0$, are as follows:

$$\rho(x) = \begin{cases} \frac{1}{2}c^2 - c(x+c) & \text{if} \quad x < -c, \\ \frac{1}{2}x^2 & \text{if} \quad |x| \le c \\ \frac{1}{2}c^2 + c(x+c) & \text{if} \quad x > c, \end{cases} \tag{6}$$

or

$$\rho(x) = \begin{cases} \frac{1}{2c}x^2 & \text{if} \quad |x| < c, \\ |x| - \frac{1}{2}c^2 & \text{if} \quad |x| \geq c. \end{cases} \tag{7}$$

Here c is the tuning parameter. In both the above cases computation of the robust estimators are quite challenging. It involves multidimensional optimization problem. The dimensionality of the optimization problem increases with p. The usual optimization technique does not work as the objective function is not differentiable. Moreover, in establishing the properties of the estimators, one needs stronger assumptions on the error random variables than what are needed for the usual least squares estimators to be consistent and asymptotically normally distributed, see for example Huber [17]. Moreover, the asymptotic properties of the robust estimators of the unknown parameters of the model (1) are not available in the literature.

The main aim of this manuscript is to overcome the above issues. In this paper we have introduced a novel weighted least squares method which can be used quite effectively as an alternative to the well known robust method of estimation. The corresponding weighted least squares estimators (WLSEs) of the unknown parameters of the model (1) can be obtained by solving only one non-linear equation for any p. Since, the objective function is a differentiable function, standard optimization technique like Newton-Raphson or Gauss-Newton algorithm can be used to minimize the objective function. Further, the asymptotic properties of the WLSEs can be established under the same set of assumptions that are needed for the LSEs to be consistent and asymptotically normally distributed. Note that the consistency and the asymptotic normality properties are very strong properties of any estimator. Consistency property indicates that as the sample size icreases the estimators converge to the true parameter value. The asymptotic normality property can be used to compute the variance of the estimators for large sample sizes and also to contruct confidence intervals of the unknown parameters. It is observed that the WLSEs of the unknown parameters have the same rate of convergence as the LSEs. Moreover, LSEs can be obtained as a special case of the WLSEs. We have performed an extensive simulation experiments to see the performance of the proposed WLSEs. It is observed that in presence of outliers with the proper choice of the weight functions, the WLSEs perform better than the LSEs, Periodogram estimators, LADEs and HMEs. Another important question is how to choose the proper weight function? It is an important issue, and we have discussed how to choose the proper weight function for a given data set from a class of weight functions.

The main contributions of this paper are the following: (1) We have proposed a novel weighted least squares method to provide the robust estimators of the unknown parameters, in presence of outliers. (2) We have developed the theoretical properties of the proposed estimators under the same set of assumptions that are needed to establish the consistency and asymptotic normality properties of the least squares estimators. (3) We have some practical suggestion to choose the weight function. (4) It is observed that with the proper choice of weight function the proposed estimators behave better than the LSEs, Periodogram estimators, LADEs and HMEs if there are outliers in the observations.

Rest of the paper is organized as follows. In Section 2 we have described the WLSEs and their implementation. The asymptotic properties of the WLSEs are established in Section 3 when we have a polynomial weight function. In Section 4 the results have been extended for a general class of weight functions. In Section 5 we have provided the simulation results to show how the proposed method works in practice. Choice of weight function is discussed in Section 6. In Section 7 we have presented the analysis of a data set, and finally we conclude the paper in Section 8.

2 Weighted Least Squares Estimators

In this section first we define the WLSEs when the weight function is a m-degree polynomial, and later we will generalize it to any arbitrary continuous function. Suppose $w(t)$ is a m-degree polynomial defined on $[0,1]$, i.e.

$$w(t) = a_0 + a_1 t + \ldots + a_m t^m, \tag{8}$$

here the weight functions $a_0, \ldots, a_m$ are such that

$$\min_{0 \le t \le 1} w(t) > \gamma > 0. \tag{9}$$

Suppose K is such that $\max_{0 \le t \le 1} w(t) \le K$. Let $\boldsymbol{\Theta} = (A_1, \ldots, A_p, B_1, \ldots, B_p, \lambda)^\top$ as the parameter vector, and $\boldsymbol{\Theta}^0 = (A_1^0, \ldots, A_p^0, B_1^0, \ldots, B_p^0, \lambda^0)^\top$ as the true parameter. We also denote $\boldsymbol{\beta} = (A_1, \ldots, A_p, B_1, \ldots, B_p)^\top$ and $\boldsymbol{\beta}^0$ as the corresponding true parameter vector. Further, let us use

$$\mu_n(\boldsymbol{\Theta}) = \sum_{k=1}^{p} \{A_k \cos(nk\lambda) + B_k \sin(nk\lambda)\},$$

for $n = 1, \ldots, N$. Let us consider the following weighted residual sums of squares

$$Q(\boldsymbol{\Theta}) = \sum_{n=1}^{N} w\left(\frac{n}{N}\right)(y(n) - \mu_n(\boldsymbol{\Theta}))^2. \tag{10}$$

The WLSEs of $\boldsymbol{\Theta}^0$ can be obtained as

$$\widehat{\boldsymbol{\Theta}} = \text{argmin}_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}). \tag{11}$$

It may be seen that the function $Q(\boldsymbol{\Theta})$ is a smooth differentiable function of the unknown parameters, hence the minimization can be obtained by solving $Q'(\boldsymbol{\Theta}) = 0$, where

$$Q'(\boldsymbol{\Theta}) = \left[\frac{\partial Q(\boldsymbol{\Theta})}{\partial A_1}, \ldots, \frac{\partial Q(\boldsymbol{\Theta})}{\partial A_p}, \frac{\partial Q(\boldsymbol{\Theta})}{\partial B_1}, \ldots, \frac{\partial Q(\boldsymbol{\Theta})}{\partial B_p}, \frac{\partial Q(\boldsymbol{\Theta})}{\partial \lambda}\right]^\top.$$

For fixed λ, the WLSEs of $A_1, \ldots, A_p$ and $B_1, \ldots, B_p$ can be obtained by solving the following matrix equation:

$$\boldsymbol{Z}(\lambda) = \boldsymbol{X}(\lambda)\boldsymbol{\beta}, \tag{12}$$

here $\boldsymbol{Z}(\lambda)$ is a $2p \times 1$ vector, $\boldsymbol{X}(\lambda)$ is a $2p \times 2p$ matrix, $\boldsymbol{\beta}$ is a $2p \times 1$ vector, and they can be partitioned as follows:

$$\boldsymbol{Z}(\lambda) = \begin{bmatrix} \boldsymbol{Z}_1(\lambda) \\ \boldsymbol{Z}_2(\lambda) \end{bmatrix}, \quad \boldsymbol{X}(\lambda) = \begin{bmatrix} \boldsymbol{X}_{11}(\lambda) & \boldsymbol{X}_{12}(\lambda) \\ \boldsymbol{X}_{21}(\lambda) & \boldsymbol{X}_{22}(\lambda) \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}.$$

In this case $\boldsymbol{Z}_1(\lambda)$ and $\boldsymbol{Z}_2(\lambda)$ are both $p \times 1$ vectors and they are as follows.

$$\begin{aligned} \boldsymbol{Z}_1(\lambda) &= \left[\frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) y(n) \cos(n\lambda), \ldots, \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) y(n) \cos(np\lambda) \right]^\top \\ \boldsymbol{Z}_2(\lambda) &= \left[\frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) y(n) \sin(n\lambda), \ldots, \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) y(n) \sin(np\lambda) \right]^\top. \end{aligned}$$

$\boldsymbol{X}_{11}(\lambda)$, $\boldsymbol{X}_{12}(\lambda)$, $\boldsymbol{X}_{21}(\lambda)$ and $\boldsymbol{X}_{22}(\lambda)$ are all $p \times p$ matrices. The (i,j)-th elements of $\boldsymbol{X}_{11}(\lambda)$, $\boldsymbol{X}_{12}(\lambda)$, $\boldsymbol{X}_{21}(\lambda)$ and $\boldsymbol{X}_{22}(\lambda)$ are

$$\begin{array}{ll} \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \cos(ni\lambda) \cos(nj\lambda), & \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \cos(ni\lambda) \sin(nj\lambda), \\ \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \sin(ni\lambda) \cos(nj\lambda), & \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \sin(ni\lambda) \sin(nj\lambda), \end{array}$$

respectively. Further, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are both $p \times 1$ vectors, and

$$\boldsymbol{\beta}_1 = (A_1, \ldots, A_p)^\top \quad \boldsymbol{\beta}_2 = (B_1, \ldots, B_p)^\top.$$

For fixed λ, the WLSEs of $\boldsymbol{\beta}$, say $\widehat{\boldsymbol{\beta}}(\lambda)$ can be obtained as $\widehat{\boldsymbol{\beta}}(\lambda) = \boldsymbol{X}^{-1}(\lambda)\boldsymbol{Z}(\lambda)$ and the WLSE of λ can be obtained by minimizing $Q(\widehat{\boldsymbol{\beta}}(\lambda), \lambda)$ with respect to λ. Hence, it is clear that the WLSEs of the unknown parameters can be obtained by solving an one dimensional optimization problem.

Although, the WLSEs can be obtained by solving only one non-linear equation, it also involves computing the inverse of a $2p \times 2p$ symmetric matrix. If p is large, it can be a computationally challenging problem. Using the following facts as $N \to \infty$, for $i, j =$

$1, \ldots, p$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \cos^2(ni\lambda) = \frac{1}{2}; \tag{13}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{n=1}^{N} n^k w\left(\frac{n}{N}\right) \cos^2(ni\lambda) = \frac{1}{2(k+1)}; k = 1, 2, \ldots, \tag{14}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \cos(ni\lambda)\cos(nj\lambda) = 0, \quad i \neq j, \tag{15}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \sin^2(ni\lambda) = \frac{1}{2}, \tag{16}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{n=1}^{N} n^k w\left(\frac{n}{N}\right) \sin^2(ni\lambda) = \frac{1}{2(k+1)}, k = 1, 2, \ldots, \tag{17}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \sin(ni\lambda)\sin(nj\lambda) = 0; \quad i \neq j, \tag{18}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \sin(ni\lambda)\cos(nj\lambda) = 0, \tag{19}$$

see Mangulis [18], we approximate $\widehat{A}_k(\lambda)$ and $\widehat{B}(\lambda)$, for large N as

$$\widehat{A}_k(\lambda) = \frac{2a_k}{\int_0^1 w(t)dt} \quad \text{and} \quad \widehat{B}_k(\lambda) = \frac{2b_k}{\int_0^1 w(t)dt}; \quad k = 1, \ldots, p. \tag{20}$$

Here

$$a_k = \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) y(n)\cos(nk\lambda) \quad \text{and} \quad b_k = \frac{1}{N} \sum_{n=1}^{N} w\left(\frac{n}{N}\right) y(n)\sin(nk\lambda).$$

Based on $\widehat{A}_k(\lambda)$ and $\widehat{B}_k(\lambda)$, the WLSE of λ can be obtained by minimizing

$$R(\lambda) = \sum_{n=1}^{N} w\left(\frac{n}{N}\right) \left(y(n) - \sum_{k=1}^{p} \{\widehat{A}_k(\lambda)\cos(nk\lambda) + \widehat{B}_k(\lambda)\sin(nk\lambda)\}\right)^2, \tag{21}$$

with respect to λ. Since $R(\lambda)$ is a differentiable function of λ. The minimization of $R(\lambda)$ can be obtained quite conveniently. Note that

$$\begin{aligned} R'(\lambda) &= 2\sum_{n=1}^{N} w\left(\frac{n}{N}\right) \left(y(n) - \sum_{k=1}^{p} \{\widehat{A}_k(\lambda)\cos(nk\lambda) + \widehat{B}_k(\lambda)\sin(nk\lambda)\}\right) \times, \\ &\quad \left(\sum_{k=1}^{p} \{\widehat{A}'_k(\lambda)\cos(nk\lambda) + \widehat{B}'_k(\lambda)\sin(nk\lambda) - A_k nk\sin(nk\lambda) + B_k nk\cos(nk\lambda)\}\right), \end{aligned}$$

here

$$\widehat{A}_k'(\lambda) = -\frac{2}{N\int_0^1 w(t)dt}\sum_{n=1}^{N} w\left(\frac{n}{N}\right) nky(n)\sin(nk\lambda)$$

$$\widehat{B}_k'(\lambda) = \frac{2}{N\int_0^1 w(t)dt}\sum_{n=1}^{N} w\left(\frac{n}{N}\right) nky(n)\cos(nk\lambda).$$

Hence, $\widehat{\lambda}$, the WLSE of λ, can be obtained as the solution of $R'(\lambda) = 0$, and once $\widehat{\lambda}$ is obtained, the WLSEs of A_k and B_k can be obtained as $\widehat{A}_k = \widehat{A}_k(\widehat{\lambda})$ and $\widehat{B}_k = \widehat{B}_k(\widehat{\lambda})$, respectively.

3 Theoretical Properties

In this section we provide the theoretical properties of the WLSEs of the unknown parameters of the model (1), as proposed in the previous section.

Theorem 1: If $X(n)$'s are independent and identically distributed random variables with mean zero and finite variance, and the weight function $w(t)$ has the form (8), then $\widehat{\boldsymbol{\Theta}}$, the WLSE of $\boldsymbol{\Theta}^0$, is a strongly consistent estimator, i.e. $\widehat{\boldsymbol{\Theta}}$ converges to $\boldsymbol{\Theta}^0$ almost surely.

Proof: See in the Appendix A. ∎

We need the following notations for further development. For $k = 0, 1, \ldots,$

$$u_k = \int_0^1 t^k w(t)dt \quad \text{and} \quad v_k = \int_0^1 t^k w^2(t)dt.$$

The $2p+1$ diagonal matrix

$$\boldsymbol{D} = \text{diag}\{N^{1/2}, \ldots, N^{1/2}, N^{3/2}\}.$$

Theorem 2: Under the same assumptions as in Theorem 1,

$$\left((\widehat{A}_1 - A_1^0), (\widehat{B}_1 - B_1^0), \ldots, (\widehat{A}_p - A_p^0), (\widehat{B}_p - B_p^0), (\widehat{\lambda} - \lambda^0\right) \xrightarrow{d} N_{2p+1}(\mathbf{0}, 2\sigma^2\boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1}).$$

Here

$$\mathbf{\Sigma} = \begin{bmatrix} v_0 & 0 & 0 & \dots & 0 & 0 & B_1^0 v_1 \\ 0 & v_0 & 0 & \dots & 0 & 0 & -A_1^0 v_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & v_0 & 0 & B_p^0 v_1 \\ 0 & 0 & 0 & \dots & 0 & v_0 & -A_p^0 v_1 \\ B_1^0 v_1 & -A_1^0 v_1 & B_2^1 v_1 & \dots & B_p^0 v_1 & -A_p^0 v_1 & c v_0 \end{bmatrix},$$

$$\mathbf{\Gamma} = \begin{bmatrix} u_0 & 0 & 0 & \dots & 0 & 0 & B_1^0 u_1 \\ 0 & u_0 & 0 & \dots & 0 & 0 & -A_1^0 u_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & u_0 & 0 & B_p^0 u_1 \\ 0 & 0 & 0 & \dots & 0 & u_0 & -A_p^0 u_1 \\ B_1^0 u_1 & -A_1^0 u_1 & B_2^1 u_1 & \dots & B_p^0 u_1 & -A_p^0 u_1 & c u_2 \end{bmatrix}$$

and $c = \sum_{j=1}^{p} (A_j^{0^2} + B_j^{0^2})$. The notation $\xrightarrow{d}$ means convergence in distribution, $N_{2p+1}(\mathbf{0}, \boldsymbol{A})$ denotes a $(2p+1)$-variate normal distribution with the mean vector $\mathbf{0}$, and the dispersion matrix $\boldsymbol{A}$.

Proof: See in the Appendix A. ∎

Note that when $w(t) = 1$, for $0 \le t \le 1$, then the WLSEs become the LSEs, and in this case $u_k = v_k = \frac{1}{k+1}$, for $k = 0, 1, \ldots$. Hence, $\mathbf{\Sigma} = \mathbf{\Gamma}$.

4 General Weight Function

So far we have assumed that the weight function $w(t)$ is of the form (8) and it satisfies (9). In this section we will show that the results are true for a general continuous function also, when it satisfies a similar condition like (9). In this section also we will be using the same notation as before, and it should not create any confusion. The following assumptions are made on the general weight function $w(t)$.

Assumption 1: Suppose $w(t)$ is a non-negative continuous function defined on $[0, 1]$, and it satisfies $\min_{0 \le t \le 1} w(t) > \gamma > 0$ and $\max_{0 \le t \le 1} w(t) \le K < \infty$.

Theorem 3: If $X(n)$'s are independent and identically distributed random variables with mean zero and finite variance, and the weight function $w(t)$ satisfies Assumption 1, then $\widehat{\boldsymbol{\Theta}}$, the WLSE of $\boldsymbol{\Theta}^0$, is a strongly consistent estimator.

Proof: See Appendix B.

Theorem 4: Under the same assumptions as in Theorem 3,

$$\left((\widehat{A}_1 - A_1^0), (\widehat{B}_1 - B_1^0), \ldots, (\widehat{A}_p - A_p^0), (\widehat{B}_p - B_p^0), (\widehat{\lambda} - \lambda^0\right) \xrightarrow{d} N_{2p+1}(\mathbf{0}, 2\sigma^2 \boldsymbol{\Gamma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Gamma}^{-1}).$$

Here $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are same as defined in Theorem 2.

Proof: See in the Appendix B.

5 Simulation Results

In this section we present some simulation results to show how the proposed method compares with the existing methods. The main aim of this section is to see how the proposed WLSEs behave compared with the LSEs, Periodogram estimators, LADEs and HMEs in terms of biases and mean squared errors in presence of outliers. We have considered the following model:

Model: $p = 2$, $A_1^0 = B_1^0 = 3$ and $A_2^0 = B_2^0 = 1$, $\lambda_0 = 0.75\,\pi = 2.356194$

We have assumed that $e(n)$'s are independent and identically distributed normal random variables with mean 0 and variance $\sigma^2 = 1$. It is assumed that 10%, 20%, 40% and 50% outliers are present in the sample and in those cases $e(n)$'s are independent and identically distributed normal random variables with mean zero and variance $\sigma^2_{out} = 25$. Moreover, it is assumed that the outliers are present in the middle of the data set. We have considered the LSEs, Periodogram estimators, LADEs, HMEs and two different WLSEs. The main reason to choose the LSEs and the Periodogram estimators is that they are the most popular estimators which are being used in practice. Moreover, the LSEs are the most efficient estimators also when the errors are Gaussian. The reason to choose the LADEs and HMEs is that they are the two most popular robust estimators which are available in the literature.

The weight functions of the two WLSEs are as follows for $0 \le t \le 1$;

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

Here $w_1(t)$ is a polynomial of degree two and $w_2(t)$ is a continuous function in $[0, 1]$, and it satisfies Assumption 1. Since the outliers are known to be present in the middle of the data set, the above two weight functions have been chosen so that they put less weights to the residual sums of squares in the middle portion of the data set. We have varied the sample sizes from 100 to 1000.

We have generated samples from the above model with the given error structure. We have calculated the LADEs, HMEs, LSEs, Periodogram estimators and two different WLSEs. In each case we have used the maximum of the periodogram function at the Fourier frequncies as the initial guess of the frequency parameter. In case of LADEs and HMEs we need to provide the initial guesses of the A_k's and B_k's also. Based on the Fourier frequency where the maximum of the periodogram function occurs, the corresponding LSEs of A_k's and B_k's have been used as the initial guesses to compute the LADEs and HMEs. In this case the computation of LADEs and HMEs involve solving a five dimensional optimization problem, and we have used the Nelder-Mead algorithm for that purpose. The computation of the LSEs, Periodogram estimators and WLSEs involve solving only one dimensional optimization problem.

We have obtained the average estimates of the two amplitudes, the fundamental frequency and the associated mean squared errors (MSEs) based on 1000 replications. Since in all the cases the average estimates are very close to the corresponding true values (biases are negligible), we have not reported those. We have only reported the MSEs in the form -10log(MSE) in each case. The results are presented in Figures 1,2 and 3.

[Figures 1,2 and 3 should be placed after this]

Some of the points are quite clear from these simulation experiments. First of all for the all the cases the MSEs decrease as the sample size increases, and as the proportion of outliers decrease. The performances of the WLSE-1 in terms of the lower MSEs are the best among these six estimators for all the parameters and the performances of the Periodogram estimators are the worst. The performances of the LADEs and HMEs are very similar in all the cases. Another point which has been observed but not reported here that the performances of LADEs and HMEs depend very much on the choice of the initial guesses. It is a quite challenging problem. Hence, the implementation of the LADEs and HMEs in practice are quite difficult. Where as the performances of the LSEs, Periodogram estimators or the WLSEs do not depend much on the choice of the initial guess of the frequency parameter. Hence, the implementation of the LSEs, Periodogram estimators and WLSEs is quite simple in practice.

6 Choice of the Weight Function

In the simulation experiments we have seen that if we know where the outliers are present, we can choose the weight function accordingly and the performances of the WLSEs are quite satisfactory. With the proper choice of the weight function, the WLSEs behave better than the LSEs, Periodogram estimators and other robust estimators like LADEs and HMEs also. But, it has also been observed that the performances of the WLSEs depend very much about the weight function itself. If the weight function is not proper, the corresponding WLSEs perform quite poorly compared to the LSEs also. Hence, the natural question is how to choose the proper weight function in practice. We propose the following method which can be used in practice quite conveniently, when nothing is known about the location of the outliers.

The method is based on choosing a proper data dependent weight function from a class of weight functions. One does not have to choose the weight function before. The method works as the following. First we choose a class of weight functions may be 5 or 6, which has different forms, i.e. one may put less weight at the beginning, one put less weight at the end, one may put less weight in the middle etc. Further, in case of time varying outliers a mixture of different types of weight functions may be chosen. It may be remembered that the LSEs also can be obtained as WLSEs when the weight function $w(t) = 1$, for all $0 \leq t \leq 1$. We will keep this weight function also as a member of this class. Now compute the WLSEs based on all the weight functions of this specific class, and choose that particular weight function which has the smallest weighted residual sums of squares. The details will be explained with the analysis of a synthetic data set in the next section.

7 Synthetic Data Analysis

In this section we have performed the analysis of a synthetic data set to show how the proposed methods work in practice and how to choose a proper weight function when outliers are present in the data set, but the exact locations are not known. We have considered the following model:

$$p = 4, A_1^0 = B_1^0 = 4, \quad A_2^0 = B_2^0 = 3, \quad A_3^0 = B_3^0 = 2, \quad A_4^0 = B_4^0 = 1, \quad \lambda^0 = 0.75\pi, \quad N = 100.$$

It is assumed that $X(n)$'s are independent and identically distributed random variables and it has the following structure;

$$X(n) = 0.8N(0,1) + 0.2N(0,25),$$

i.e. there are 20% outliers present in the data, but the locations are not known. The original and the observed signals are presented in Figure 4. The periodogram function of the data is presented in Figure 5.

[Figures 4 and 5 should be placed after this]

Since, the locations of the outliers are not known we have used the following six weight functions which put different weights at different parts as they have different shapes.

$$w_1(t) = 0.3 + 4.8(t - 0.25)^2, \quad w_2(t) = 1.0, \quad w_3(t) = \frac{1}{12} + 11(t - 0.5)^2$$

$$w_4(t) = \begin{cases} \frac{19}{15}(2 - 4t) & \text{if} \quad 0 \leq t \leq 0.25 \\ 0.1 & \text{if} \quad 0.25 \leq t \leq 0.75 \\ \frac{19}{15}(4t - 2) & \text{if} \quad 0.75 \leq t \leq 1.00 \end{cases}$$

$$w_5(t) = \frac{4}{3} - 4(t - 0.5)^2, \quad w_6(t) = \frac{12}{19} + \frac{48}{19}(t - 0.75)^2.$$

We have plotted the weight functions in Figure 6. It is very clear they take various shapes, and putting different weights across the time.

[Figure 6 should be placed after this.]

Estimates of the amplitudes and the frequency and the associated weighted residual sums of squares (WRSS) based on different weight functions are presented in Table 1.

[Table 1 should be placed after this.]

Based on the WRSS, we suggest to use the weight function (v) for this data set. We have plotted the predicted signal based on the estimated parameters obtained using weight function (v) in Figure 7. They match quite well with the original signal.

[Figure 7 should be placed after this]

Finally, it may be mentioned that in this case it is very difficult to implement the LADEs and HMEs, as both of them involved solving a nine dimensional optimization problem, and they require very efficient initial guesses at the beginning to start any iterative process. Where as in this case it is quite simple to implement the proposed WLSEs to analyze the data when outliers are present.

8 Conclusions

In this paper we have proposed a novel robust method to estimate the unknown parameters of a fundamental frequency and its harmonics model in presence of outliers. The proposed method is very easy to implement in practice, as it involves solving only one dimensional optimization problem. The performances of the proposed estimators depend very much on the choice of the weight function. If the location of the outliers are known, then it is easy to choose the weight function, but if it is unknown, we have proposed a method to choose a proper weight function. Simulation results and the data analysis indicate that the proposed method works quite well in practice, and its implementation is also quite simple.

Acknowledgements:

The author would like to thank the Associate Editor and three unknown reviewers for their constructive comments which have helped to improve the earlier version of the paper significantly.

Appendix A

To prove Theorem 1, we need the following lemmas.

Lemma 1: Let $\{X(n)\}$ be a sequence of independent and identically distributed random variables with mean 0 and finite variance, then for $k = 0, 1, \ldots,$

$$\lim_{N\to\infty} \sup_{\theta} \left| \frac{1}{N^{k+1}} \sum_{n=1}^{N} X(n) n^k \cos(n\theta) \right| = 0 \quad a.s.$$

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

Proof: For $k = 0$, the result is available in Kundu and Mitra [19]. For general k, the result follows from the fact $\dfrac{n}{N} \leq 1$, for $1 \leq n \leq N$. ∎

Lemma 2: Let $\{X(n)\}$ be a sequence of independent and identically distributed random variables with mean 0 and finite variance, and the weight function $w(t)$ has the form (8), then

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

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

Proof: Using Lemma 1, Lemma 2 can be easily obtained. ∎

Lemma 3: Let $\boldsymbol{\Theta}$ be the same as before, and let us define the set

$$\begin{aligned} S_{\delta,M} \;&=\; \{\boldsymbol{\Theta} : |\lambda - \lambda^0| > \delta \;\text{ or }\; |A_k^0 - A_k| > \delta \;\text{ or }\; |B_k^0 - B_k| > \delta, \text{for any } k = 1, \ldots, p \\ &\qquad \text{and } |A_k| \leq M, |B_k| \leq M, \text{ for all } k = 1, \ldots, p\}. \end{aligned}$$

If for any $\delta > 0$ and for some $M < \infty$,

$$\liminf_{N\to\infty} \inf_{\boldsymbol{\Theta} \in S_{\delta,M}} \frac{1}{N} \{Q(\boldsymbol{\Theta}) - Q(\boldsymbol{\Theta}^0)\} > 0, \tag{22}$$

where $Q(\boldsymbol{\Theta})$ is same as defined in (10), then $\widehat{\Theta}$ is a strongly consistent estimator of $\boldsymbol{\Theta}^0$.

PROOF: It mainly follows by contradiction, along the same line as the proof of Lemma 1 of Wu [20]. Hence, the details are avoided. ∎

LEMMA 4: For any given $\delta > 0$ and for some $M < \infty$,

$$\liminf_{N\to\infty} \inf_{\boldsymbol{\Theta}\in S_{\delta,M}} \frac{1}{N}\{Q(\boldsymbol{\Theta}) - Q(\boldsymbol{\Theta}^0)\} > 0.$$

PROOF: Consider

$$\begin{aligned}
\frac{1}{N}Q(\boldsymbol{\Theta}) &= \frac{1}{N}\sum_{n=1}^{N} w\left(\frac{n}{N}\right)(y(n) - \mu_n(\boldsymbol{\Theta}))^2 \\
&= \frac{1}{N}\sum_{n=1}^{N} w\left(\frac{n}{N}\right)(\mu_n(\boldsymbol{\Theta}^0) - \mu_n(\boldsymbol{\Theta}) + X(n))^2 \\
&= \frac{1}{N}\sum_{n=1}^{N} w\left(\frac{n}{N}\right)(\mu_n(\boldsymbol{\Theta}^0) - \mu_n(\boldsymbol{\Theta}))^2 + \frac{1}{N}\sum_{n=1}^{N} w\left(\frac{n}{N}\right)X^2(n) + \\
&\quad \frac{2}{N}\sum_{n=1}^{N} w\left(\frac{n}{N}\right)(\mu_n(\boldsymbol{\Theta}^0) - \mu_n(\boldsymbol{\Theta}))X(n)
\end{aligned}$$

Hence, using Lemma 2, and the condition of the weight function $w(t)$, we obtain

$$\liminf_{N\to\infty} \inf_{\boldsymbol{\Theta}\in S_{\delta,M}} \frac{1}{N}\{Q(\boldsymbol{\Theta}) - Q(\boldsymbol{\Theta}^0)\} \geq \liminf_{N\to\infty} \inf_{\boldsymbol{\Theta}\in S_{\delta,M}} \frac{\gamma}{N}\sum_{n=1}^{N}(\mu_n(\boldsymbol{\Theta}^0) - \mu_n(\boldsymbol{\Theta}))^2.$$

Consider the following sets for $k = 1, \ldots, p$:

$$\begin{aligned}
\Gamma_{1k} &= \{\boldsymbol{\Theta} : |A_j| \leq M, |B_j| \leq M, |A_k - A_k^0| > \delta\} \\
\Gamma_{2k} &= \{\boldsymbol{\Theta} : |A_j| \leq M, |B_j| \leq M, |B_k - B_k^0| > \delta\} \\
\Gamma_0 &= \{\boldsymbol{\Theta} : |A_j| \leq M, |B_j| \leq M, |\lambda - \lambda^0| > \delta\},
\end{aligned}$$

and $\Gamma_1 = \cup_{k=1}^{p}\Gamma_{1k}$, $\Gamma_2 = \cup_{k=1}^{p}\Gamma_{2k}$. Since

$$\{\boldsymbol{\Theta} : \boldsymbol{\Theta} \in S_{\delta,M}\} \subset \Gamma_1 \cup \Gamma_2 \cup \Gamma_0,$$

$$\liminf_{N\to\infty} \inf_{\mathbf{\Theta}\in S_{\delta,M}} \frac{\gamma}{N}\sum_{n=1}^{N}(\mu_n(\mathbf{\Theta}^0)-\mu_n(\mathbf{\Theta}))^2 \geq \liminf_{N\to\infty} \inf_{\mathbf{\Theta}\in\Gamma_1\cup\Gamma_2\cup\Gamma_0} \frac{\gamma}{N}\sum_{n=1}^{N}(\mu_n(\mathbf{\Theta}^0)-\mu_n(\mathbf{\Theta}))^2.$$

Observe that

$$\liminf_{N\to\infty} \inf_{\mathbf{\Theta}\in\Gamma_{1k}} \frac{\gamma}{N}\sum_{n=1}^{N}(\mu_n(\mathbf{\Theta}^0)-\mu_n(\mathbf{\Theta}))^2 = \liminf_{N\to\infty} \frac{\gamma|A_k-A_k^0|^2}{N}\sum_{n=1}^{N}\cos^2(\lambda^0 n) = \frac{\gamma|A_k-A_k^0|^2}{2} > 0.$$

Similarly, it can be shown for other sets also, hence the result follows. ∎

Proof of Theorem 1: Using Lemma 3 and Lemma 4, it immediately follows. ∎

Proof of Theorem 2: To prove this result, let us consider the following $2p+1$ vector $Q'(\mathbf{\Theta})$, where

$$Q'(\mathbf{\Theta}) = \left(\frac{\partial Q(\mathbf{\Theta})}{\partial A_1}, \frac{\partial Q(\mathbf{\Theta})}{\partial B_1}, \ldots, \frac{\partial Q(\mathbf{\Theta})}{\partial A_p}, \frac{\partial Q(\mathbf{\Theta})}{\partial B_p}, \frac{\partial Q(\mathbf{\Theta})}{\partial \lambda}\right),$$

$Q''(\mathbf{\Theta})$ is a $(2p+1)\times(2p+1)$ matrix contains the double derivative of $Q(\mathbf{\Theta})$. Now using the Taylor series expansion

$$Q'(\widehat{\mathbf{\Theta}}) - Q'(\mathbf{\Theta}^0) = (\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^0)Q''(\overline{\mathbf{\Theta}}), \tag{23}$$

here $\overline{\mathbf{\Theta}}$ is a point on the line joining $\widehat{\mathbf{\Theta}}$ and $\mathbf{\Theta}^0$. Since $Q'(\widehat{\mathbf{\Theta}}) = \mathbf{0}$, hence (23) can be written as

$$-Q'(\mathbf{\Theta}^0)\boldsymbol{D}^{-1}[\boldsymbol{D}^{-1}Q''(\overline{\mathbf{\Theta}})\boldsymbol{D}^{-1}]^{-1} = (\widehat{\mathbf{\Theta}} - \Theta^0)\boldsymbol{D}.$$

Now using Central limit theorem and (13) to (19), it follows that

$$Q'(\mathbf{\Theta}^0)\boldsymbol{D}^{-1} \xrightarrow{d} N_{2p+1}(\mathbf{0}, 2\sigma^2\mathbf{\Sigma}),$$

and using (13) to (19), it can be shown that

$$\lim_{N\to\infty} \boldsymbol{D}^{-1}Q''(\overline{\mathbf{\Theta}})\boldsymbol{D}^{-1} = \lim_{N\to\infty} \boldsymbol{D}^{-1}Q''(\mathbf{\Theta}^0)\boldsymbol{D}^{-1} = \mathbf{\Gamma},$$

hence, the result follows. ∎

Appendix B

To prove Theorem 3, we need the following Lemma.

Lemma 5: Let $\{X(n)\}$ be a sequence of independent and identically distributed random variables with mean 0 and finite variance, and the weight function $w(t)$ satisfies Assumption 1, then

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

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

Proof: Let $Z(n) = X(n)I_{[|X(n)|\le\sqrt{n}]}$, where $I_{[|X(n)|\le\sqrt{n}]} = 1$, if $|X(n)| \le \sqrt{n}$, and 0, otherwise. Thus

$$\begin{aligned}
\sum_{n=1}^{\infty}P(X(n)\neq Z(n)) &= \sum_{n=1}^{\infty}P(|X(n)|>\sqrt{n}) \\
&= \sum_{n=1}^{\infty}\big(1-F(\sqrt{n}\big) \quad \text{(here } F(\cdot) \text{ is the distribution function of } |X(n)|) \\
&\le 1+\int_0^{\infty}(1-F(\sqrt{x})dx \\
&= 1+2\int_0^{\infty}y(1-F(y))dy \\
&= 1+2\int_0^{\infty}\int_y^{\infty}dF(z)\ dy \\
&= 1+2\int_0^{\infty}\int_0^{z}ydy\ dF(z) \\
&= 1+\int_0^{\infty}z^2dF(z)<\infty.
\end{aligned}$$

Hence, $\{X(n)\}$ and $\{Z(n)\}$ are equivalent sequences. Thus it is enough to show that

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

Let $U(n) = Z(n) - E(Z(n))$. Note that if $G(\cdot)$ denotes the distribution function of $X(1)$, then

$$\sup_{\theta}\left|\frac{1}{N}\sum_{n=1}^{N}E[Z(n)w\left(\frac{n}{N}\right)\cos(n\theta)]\right| \le \frac{1}{N}\sum_{n=1}^{N}|E[Z(n)]| = \frac{1}{N}\sum_{n=1}^{N}\left|\int_{|x|\le\sqrt{n}} xdG(x)\right| \longrightarrow 0.$$

Hence, the result is proved if we can show that

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

For any fixed θ and $\epsilon > 0$, and for $0 < h < \frac{1}{4\sqrt{N}}$, we have

$$\begin{aligned} P\left\{\left|\frac{1}{N}\sum_{n=1}^{N}U(n)w\left(\frac{n}{N}\right)\cos(n\theta)\right| \ge \epsilon\right\} &\le e^{-hN\epsilon}E\,\exp\left\{h\left|\sum_{n=1}^{N}U(n)w\left(\frac{n}{N}\right)\cos(n\theta)\right|\right\} \\ &\le e^{-hN\epsilon}\prod_{n=1}^{N}E\,\exp\left|hU(n)w\left(\frac{n}{N}\right)\cos(n\theta)\right|. \end{aligned}$$

Note that $hU(n)w\left(\frac{n}{N}\right)\cos(n\theta) \le \frac{1}{4}$, for all $n = 1,\ldots,N$, and on using the fact that $e^{|x|} \le 2e^{x}$ and $e^{x} \le (1+x+x^2)$, then for $|x| \le \frac{1}{4}$

$$\begin{aligned} e^{-hN\epsilon}\prod_{n=1}^{N}E\,\exp\left|hU(n)w\left(\frac{n}{N}\right)\cos(n\theta)\right| &\le 2e^{-hN\epsilon}\prod_{n=1}^{N}E\,\exp\left(hU(n)w\left(\frac{n}{N}\right)\cos(n\theta)\right) \\ &\le 2e^{-hN\epsilon}\prod_{n=1}^{N}(1+h^2\sigma^2) \le 2\exp\,\left(-hN\epsilon + Nh^2\sigma^2\right). \end{aligned}$$

Choose $h = \frac{1}{4\sqrt{N}}$, then for large N,

$$P\left\{\left|\frac{1}{N}\sum_{n=1}^{N}U(n)w\left(\frac{n}{N}\right)\cos(n\theta)\right| \ge \epsilon\right\} \le 2\,\exp\,\left(-\frac{\sqrt{N}\epsilon}{4} + \frac{\sigma^2}{16}\right) \le C\exp\,\left(-\frac{\sqrt{N}\epsilon}{4}\right).$$

For some constant $C > 0$. Let $k = N^2$, and choose $\theta_1,\ldots\theta_k$, such that for each $\theta \in [0,\pi]$, there exists a θ_j, such that $|\theta_j - \theta| \le \frac{\pi}{N^2}$. Hence

$$\left|\frac{1}{N}\sum_{n=1}^{N}U(n)w\left(\frac{n}{N}\right)(\cos(n\theta)-\cos(n\theta_j))\right| \le \frac{1}{N}\sum_{n=1}^{N}\sqrt{n}n\frac{\pi}{N^2} \le \frac{\pi}{\sqrt{N}} \longrightarrow 0.$$

Therefore, for large N, we have

$$P\left\{\sup_{\theta}\left|\frac{1}{N}\sum_{n=1}^{N}U(n)w\left(\frac{n}{N}\right)\cos(n\theta)\right|\geq 2\epsilon\right\} \leq \left\{\max_{j\leq N^2}\left|\frac{1}{N}\sum_{n=1}^{N}U(n)w\left(\frac{n}{N}\right)\cos(n\theta_j)\right|\geq \epsilon\right\} \leq$$
$$\leq 2CN^2\text{exp}\left(-\frac{\sqrt{N}\epsilon}{4}\right).$$

Since, $\sum_{N=1}^{\infty}N^2\text{exp}\left(-\frac{\sqrt{N}\epsilon}{4}\right) < \infty$, hence, by Borel-Cantelli lemma

$$\sup_{\theta}\left|\sum_{n=1}^{N}U(n)w\left(\frac{n}{N}\right)\cos(n\theta)\right| \longrightarrow 0, \quad a.s.$$

Similarly, it can be shown when the $\cos(n\theta)$ is replaced by $\sin(n\theta)$. ∎

LEMMA 6: For any given $\delta > 0$ and for some $M < \infty$, if the weight function $w(t)$ satisfies Assumption 1, then

$$\liminf_{N\to\infty}\inf_{\boldsymbol{\Theta}\in S_{\delta,M}}\frac{1}{N}\{Q(\boldsymbol{\Theta}) - Q(\boldsymbol{\Theta}^0)\} > 0.$$

PROOF: Note that based on Lemma 5, and following the proof of Lemma 4, it can be obtained. ∎

PROOF OF THEOREM 3: Since we have a similar version of Lemma 3, where the weight function $w(t)$ satisfies Assumption 1, and using Lemma 5, Theorem 3 follows. ∎

We need the following Lemma to prove Theorem 4.

LEMMA 7: Suppose $0 < \theta < \pi$ and $w(t)$ satisfies Assumption 1, then

$$\lim_{N^{k+1}\to\infty}\frac{1}{N}\sum_{n=1}^{N}n^k w\left(\frac{n}{N}\right)\sin^2(\theta t) = \lim_{N^{k+1}\to\infty}\frac{1}{N}\sum_{n=1}^{N}n^k w\left(\frac{n}{N}\right)\cos^2(\theta t)$$
$$= \frac{1}{2}\int_0^1 t^k w(t)dt, \quad \text{for } k = 0, 1, \ldots.$$

PROOF OF LEMMA 7: We will show the result for $k = 0$, for general k, it follows along the same line.

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}w\left(\frac{n}{N}\right)\cos^2(\theta t) = \frac{1}{2}\int_0^1 w(t)dt.$$

For $\epsilon > 0$, there exists a polynomial $p_\epsilon(t)$, such that $|w(t) - p_\epsilon(t)| \leq \epsilon$, for all $t \in [0, 1]$. Hence

$$\int_0^1 w(t)dt - \epsilon \leq \int_0^1 p_\epsilon(t)dt \leq \int_0^1 w(t)dt + \epsilon. \tag{24}$$

Further,

$$\frac{1}{N}\sum_{n=1}^{N} p_\epsilon\left(\frac{n}{N}\right)\cos^2(n\theta) - \frac{\epsilon}{N}\sum_{n=1}^{N}\cos^2(n\theta) \leq \frac{1}{N}\sum_{n=1}^{N} w\left(\frac{n}{N}\right)\cos^2(n\theta) \leq$$
$$\frac{1}{N}\sum_{n=1}^{N} p_\epsilon\left(\frac{n}{N}\right)\cos^2(n\theta) + \frac{\epsilon}{N}\sum_{n=1}^{N}\cos^2(n\theta). \tag{25}$$

Suppose

$$p_\epsilon(t) = a_0 + a_1 t + \ldots + a_k t^k \quad \Rightarrow \quad \int_0^1 p_\epsilon(t)dt = a_0 + \frac{a_1}{2} + \ldots + \frac{a_k}{k+1}.$$

Now using (14),

$$\begin{aligned} \frac{1}{N}\sum_{n=1}^{N} p_\epsilon\left(\frac{n}{N}\right)\cos^2(t\theta) \quad &= \quad \frac{1}{N}\sum_{n=1}^{N}\left\{a_0 + \frac{a_1 n}{N} + \ldots + \frac{a_k n^k}{N^k}\right\}\cos^2(n\theta) \\ &\longrightarrow \quad \frac{1}{2}\left[a_0 + \frac{a_1}{2} + \ldots + \frac{a_k}{k+1}\right] = \frac{1}{2}\int_0^1 p_\epsilon(t)dt. \end{aligned}$$

Taking $N \to \infty$ in (25), we obtain

$$\frac{1}{2}\int_0^1 p_\epsilon(t)dt - \frac{\epsilon}{2} \leq \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} w\left(\frac{n}{N}\right)\cos^2(n\theta) \leq \frac{1}{2}\int_0^1 p_\epsilon(t)dt + \frac{\epsilon}{2},$$

and using (24), it follows

$$\frac{1}{2}\int_0^1 w(t)dt - \frac{3\epsilon}{2} \leq \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} w\left(\frac{n}{N}\right)\cos^2(n\theta) \leq \frac{1}{2}\int_0^1 w(t)dt + \frac{3\epsilon}{2}.$$

Since ϵ is arbitrary, the result follows. ∎

Theorem 4: Following the same line as the proof of Theorem 2, it can be proved. ∎

Data Availability Statement

The manuscript has no associated data.

Conflict of Interest Statement:

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

References

[1] Fletcher, N.H. and Rossing, T.D. (1991), *The Physics of Musical Instruments*, New York, Springer-Verlag.

[2] Kundu, D. and Nandi, S. (2004), "A Note on Estimating the Fundamental Frequency of a Periodic Function", *Signal Processing*, vol. 84, 653–661.

[3] Nandi, S. and Kundu, D. (2003), "Estimating the Fundamental Frequency of a Periodic Function", *Statistical Methods and Applications*, vol. 12, 341 – 360.

[4] Walker, A.M. (1971), "On the estimation of a harmonic component in a time series with stationary residuals", *Biometrika*, vol. 58, 21 – 36.

[5] Nielsen, J.K., Jensen, T.L., Jensen, J.R., Christensen, M.G., Jensen, S.H. (2017), "Fast fundamental frequency estimation: Making a statistically efficient estimator computationally efficient", *Signal Processing*, vol. 135, 188 – 197.

[6] Jensen, J.R., Glentis, G.O., Christensen, M.G., Jakobsson, A. and Jensen, S.H. (2013), "Fast LCMV-Based Methods for Fundamental Frequency Estimation,", *IEEE Transactions on Signal Processing*, vol. 61, 3159–3172.

[7] Cheveigńe, A de and Kawahara, H. (2002), "YIN, a fundamental frequency estimator for speech and music", *Journal of Acoustic Society of America*, vol. 111, 1917 – 1930.

[8] Li, H., Stoica, P. and Li, J. (2000), "Computationally efficient parameter estimation for harmonic sinusoidal signals", *Signal Processing*, vol. 80, 1937 – 1944.

[9] Kundu, D. and Nandi, S. (2005), "Estimating the number of components of the fundamental frequency model", *Journal of the Japan Statistical Society*, vol. 35, 41–59.

[10] Nandi, S. and Kundu, D. (2019), "Estimating the fundamental frequency using modified Newton-Raphson algorithm", *Statistics*, vol. 53, 440 – 458.

[11] 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.

[12] Qiu, L., Yang, H. and Koh, S-N (1995), "Fundamental frequency determination based on instantaneous frequency estimation", *Signal Processing*, vol. 44, 233 – 241.

[13] Fu, L., Zhang, M., Liu, Z., Li, H. (2019), "Robust Frequency Estimation of Multi-sinusoidal Signals Using Orthogonal Matching Pursuit with Weak Derivatives Criterion", *Circuits Syst Signal Process*, vol. 38, 1194–1205.

[14] Chen, Y., Yang, X.L., Huang, L-T, So, H.C. (2018), "Robust Frequency Estimation in Symmetric α–Stable Noise, *Circuits Syst Signal Process*, vol. 37, 4637–4650.

[15] Rengaswamy, P., Rao, K.S. and Dasgupta, P. (2021), "SongF0: A Spectrum-Based Fundamental Frequency Estimation for Monophonic Songs", *Circuits Syst Signal Process*, vol. 40, 772–797.

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

[19] Kundu, D. and Mitra, A. (1996), "A Note on the Consistency of the Undamped Exponential Signals Model", *Statistics*, vol. 28, 25–33.

[20] Wu, C.F.J. (1981), "Asymptotic theory of the nonlinear least squares estimation", *Annals of Statistics*, vol. 9, 501 – 513.

Weight	$\widehat{A_1^2}+\widehat{B_1^2}$	$\widehat{A_2^2}+\widehat{B_2^2}$	$\widehat{A_3^2}+\widehat{B_3^2}$	$\widehat{A_4^2}+\widehat{B_4^2}$	$\widehat{\lambda}$	WRSS
(i)	28.6888	22.0877	9.7080	0.7357	2.3588	1352.844
(ii)	28.9277	19.0103	6.7963	2.6742	2.3594	1235.015
(iii)	31.3035	18.9592	4.8104	0.5013	2.3616	1202.227
(iv)	23.4238	6.5302	0.3317	0.3750	2.4366	2136.165
(v)	28.3764	18.8499	7.5327	4.2211	2.3571	1186.725
(vi)	29.9024	17.5617	4.9168	2.4056	2.3611	1218.111

Table 1: WLSEs, normalized residual sums of squares and squared error distance of the fitted and the originals based on different weight functions

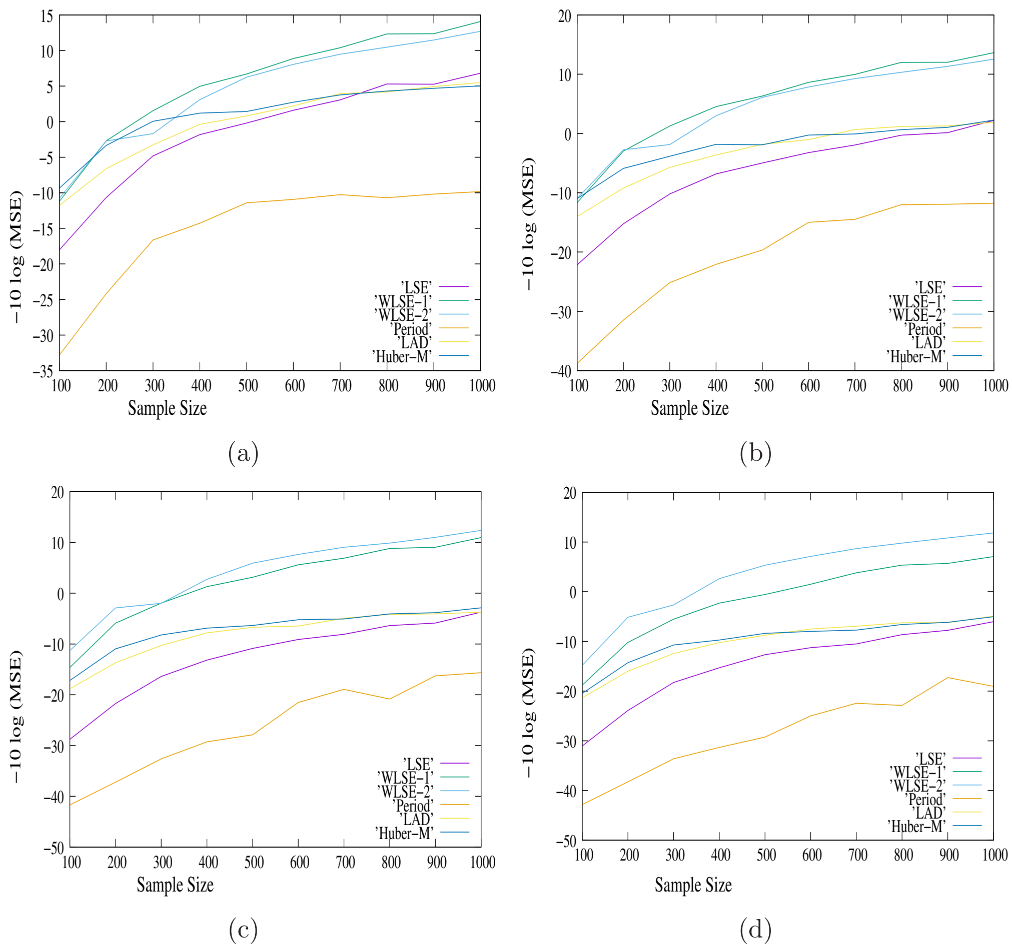


Figure 1: -10log(MSE) of the first amplitude estimators in presence of outliers: (a) 10% outliers, (b) 20% outliers, (c) 40% outliers, (d) 50% outliers.

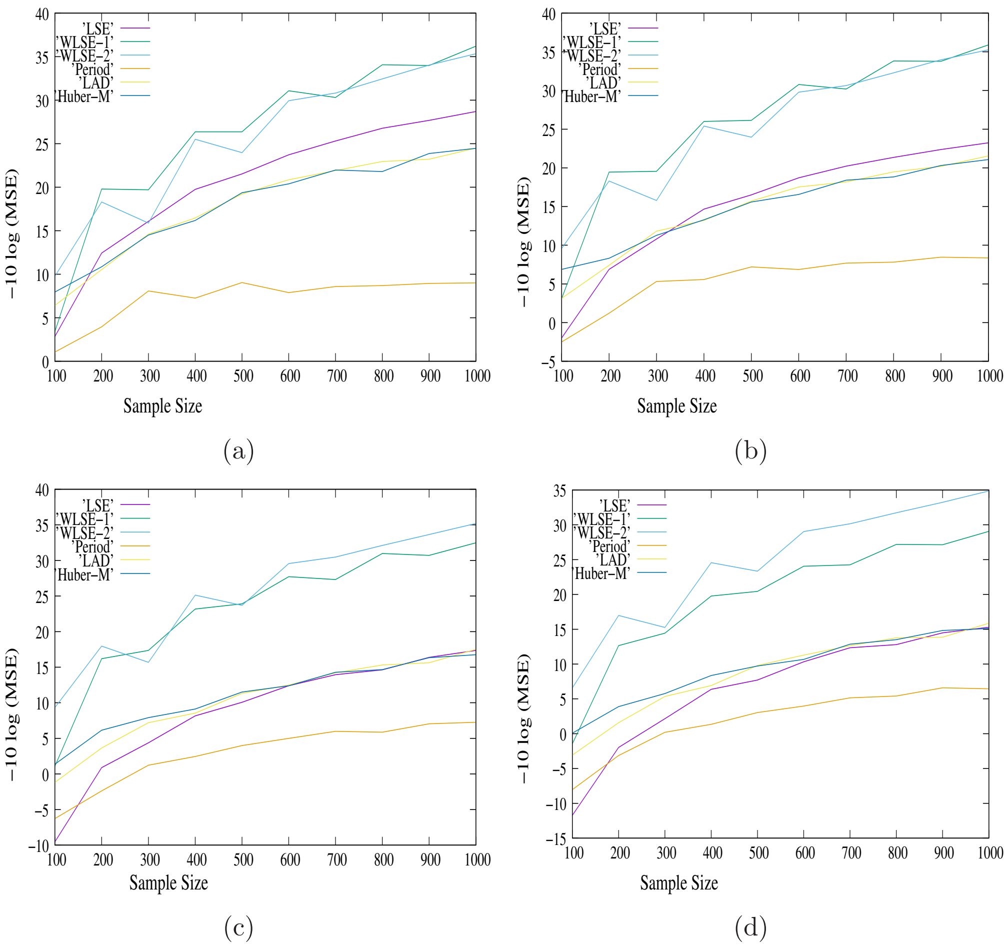


Figure 2: -10log(MSE) of the second amplitude estimators in presence of outliers: (a) 10% outliers, (b) 20% outliers, (c) 40% outliers, (d) 50% outliers.

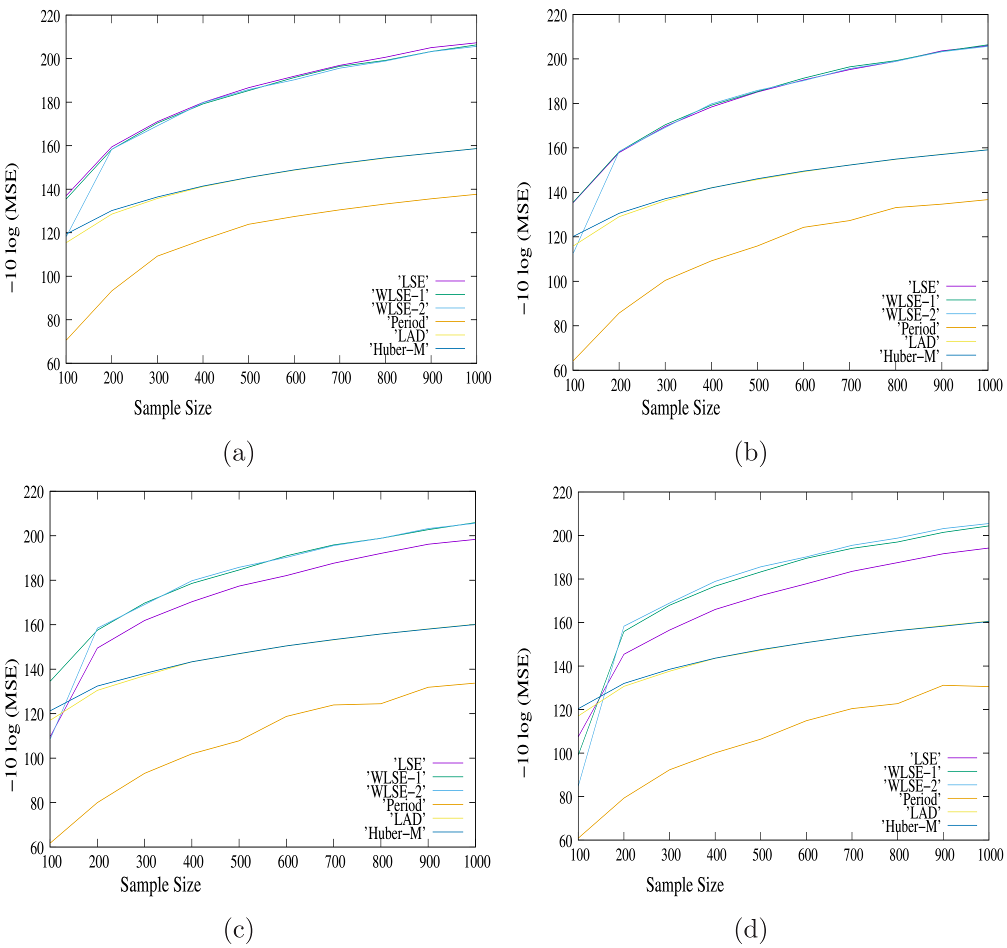
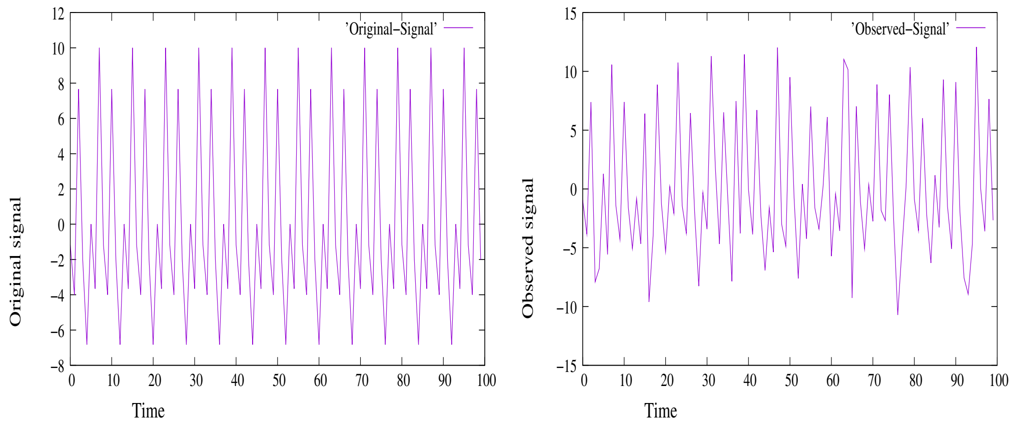


Figure 3: -10log(MSE) of the frequency estimators in presence of outliers: (a) 10% outliers, (b) 20% outliers, (c) 40% outliers, (d) 50% outliers.



(a) (b)

Figure 4: (a) Original Signal and (b) Observed Signal.

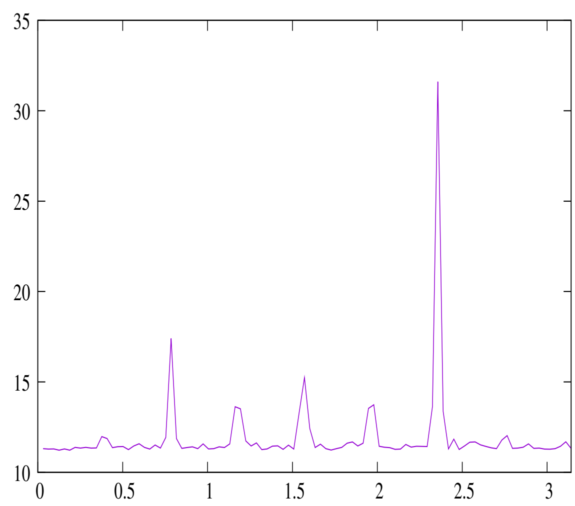


Figure 5: Periodogram function of the observed signal.

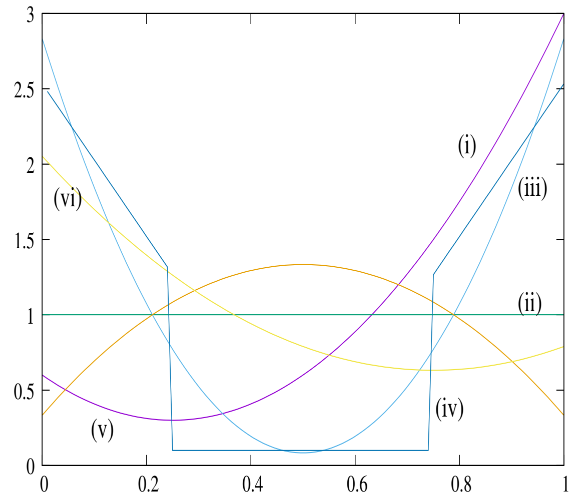


Figure 6: Different weight functions.

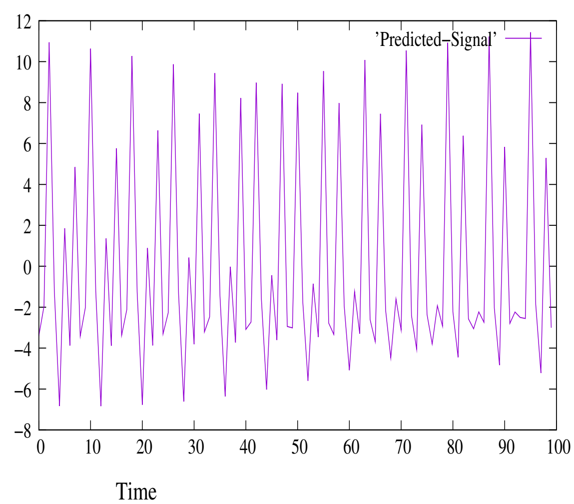


Figure 7: Predicted Signal based on the weight function (v).