On Weighted Least Squares Estimators of Parameters of a Chirp Model

Debasis Kundu*& Swagata Nandi†& Rhythm Grover ‡

Abstract

The least squares method seems to be a natural choice in estimating the parameters of a chirp model. But the least squares estimators are very sensitive to the outliers. Even in presence of very few outliers, the performance of the least squares estimators becomes quite unsatisfactory. Due to this reason, the least absolute deviation method has been proposed in the literature. But implementing the least absolute deviation method is quite challenging particularly for the multicomponent chirp model. In this paper, we propose to use the weighted least squares estimators, which seem to be more robust in presence of a few outliers. First, we consider the weighted least squares estimators of the unknown parameters of a single component chirp signal model. It is assumed that the weight function is a finite degree polynomial and the errors are independent and identically distributed random variables with mean zero and finite variance. It is observed that the weighted least squares estimators are strongly consistent and they have the same convergence rate as the least squares estimators. The weighted least squares estimators can be obtained by solving a two dimensional optimization problem. In case of the multicomponent chirp signal, we provide a sequential weighted least squares estimators and provide the consistency and asymptotic normality properties of these sequential weighted least squares estimators. To compute the sequential weighted least squares estimators one needs to solve only one two dimensional optimization problem at each stage. Extensive simulations have been performed to see the performances of the proposed estimators. Two data sets have been analyzed for illustrative purposes.

Keywords: Non-linear least squares; weighted least squares; asymptotic distribution; strong consistency; outliers.

*Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India, kundu@iitk.ac.in

†Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi - 110016, India, nandi@isid.ac.in

‡Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi - 110016, India, rhythm.grover91@gmail.com

1 Introduction

In this paper we have considered the multicomponent chirp signal model with an additive error and it can be written as follows:

$$y(t) = \sum_{k=1}^{p} \{A_k^0 \cos(\alpha_k^0 t + \beta_k^0 t^2) + B_k^0 \sin(\alpha_k^0 t + \beta_k^0 t^2)\} + X(t); \qquad t = 1, \ldots, N. \tag{1}$$

Here A_k^0s and B_k^0s are unknown real numbers and $(A_k^{0^2} + B_k^{0^2})$s are known as amplitudes, α_k^0s and β_k^0s are known as frequency and frequency rate, respectively, and $X(t)$ is an additive error, independent and identically distributed (i.i.d.) with mean zero and finite variance. The problem is to estimate the unknown parameters A_k^0, B_k^0, α_k^0 and β_k^0, $k = 1, \ldots, p$ based on a sample of size N, assuming p is known.

Retrieving the parameters of Chirp signals has received a considerable amount of attention in the signal processing literature due to its wide applications in many natural and man-made systems like audio signals, sonar, radar, etc. Unlike the sinusoidal model, a chirp signal has a frequency that changes with time. Chirp model has its roots in radar signal modeling and is used in various forms for modeling trajectories of moving objects. An extensive amount of work has been done developing different estimation procedures of the parameters of a chirp model, see for example Abatzoglou [1], Djurić and Kay [2], Farquharson, O'Shea and Ledwich [3], Gini, Montanari and Verrazzani [4], Lahiri, Kundu and Mitra [10], Grover, Kundu and Mitra [5] and see the references cited therein.

The chirp model (1) is a non-linear regression model, and the least squares method has been used in estimating the unknown parameters of this model. But the chirp model (1) does not satisfy the standard sufficient conditions of Jennrich [7] or Wu [15] so that the least squares estimators (LSEs) become consistent. Hence, although the LSEs are the most natural estimators, the consistency of the estimators is not guaranteed. Nandi and Kundu [12] first provided the formal proof of consistency and asymptotic normality properties of the LSEs of the parameters of a chirp model under the assumption that the errors are i.i.d. random variables with mean zero and finite variance. It is observed that the asymptotic variances of the frequency and frequency rate estimators are of the orders N^{-3} and N^{-5}, respectively. Therefore, the variance of the frequency rate estimator converges much faster than the variance of the frequency estimator. Due to these consistency and asymptotic normality properties the LSEs are the most preferred estimators in a perfect condition. Moreover, under the assumption of normality of the error distribution, the variances of the LSEs achieve the Cramer-Rao lower bound. Hence, they are the most efficient estimators also in case of Gaussian errors.

It is further observed that although the LSEs seem to be a natural choice, they are very sensitive to the outliers. Even if very few outliers are present, they affect the performances of the estimators quite significantly. Due to this reason Lahiri, Kundu and Mitra [9] proposed the least absolute deviation estimators (LADEs) for one component ($p = 1$) chirp model. It is observed that the LADEs are strongly consistent and they are asymptotically normally distributed. It may be mentioned that one needs a stronger set of assumptions than what are needed in case of LSEs, to establish the consistency and asymptotic normality properties of the LADEs. It is further observed that extending the results for a multicomponent model is not straightforward. Moreover, implementing the LADEs for p component chirp model involves solving a $4p$ dimensional optimization problem. Therefore, for large p, implementing the least absolute deviation (LAD) procedure is quite challenging.

In this paper, we propose to use the weighted least squares estimators (WLSEs), which seem to be quite robust in presence of few outliers. The main motivation to use the WLSEs as more robust estimators compared to LSEs is the following. In case of WLSEs we choose the weight function in such a manner that where outliers are present, less weight has been given compared to the least squares method. Hence, it is expected that this method will produce more robust estimators compared to the least squares method. The idea is very similar to the least absolute deviation criterion. We consider the WLSEs of the parameters of the multiple chirp model (1) when the weight function is a known finite degree polynomial. Therefore, the LSEs can be obtained as a special case of the proposed estimators. First, we consider the single component chirp model, i.e. when $p = 1$ in the model defined in (1). Under certain restrictions on the weight function, we have shown that the WLSEs are strongly consistent and asymptotically normally distributed. It is observed that the asymptotic variances of the WLSEs of the frequency and frequency rate are of the orders N^{-3} and N^{-5}, respectively. We further consider the WLSEs of the multiple chirp model (1). It involves solving a $2p$ dimensional non-linear optimization problem, which can be computationally challenging, if p is large. To avoid that we have proposed a sequential procedure that involves solving p two dimensional problems and we have shown that the sequential estimators have the same asymptotic properties as the WLSEs. Extensive simulations have been performed to compare the performances of the WLSEs and LADEs for one component chirp model, and they are quite comparable. Two data sets have been analyzed for illustrative purposes using the multicomponent model. Finally we have indicated how the method can be generalized for a general weight function.

The main contributions of this paper are the following. We have provided a new estimation procedure namely the weighted least squares (WLS) estimation procedure, to estimate the unknown parameters of the chirp model (1). The proposed estimation procedure is very easy to implement in practice, and it can be used quite conveniently even for large p. The proposed method has the same computational complexity as the least squares method. The proposed WLSEs are quite robust compared to the LSEs in presence of outliers. The performances of the WLSEs are very similar to the robust estimators like LADEs, although it is well known that the LADEs are more difficult to compute than the LSEs. Particularly, for large p, the implementation of the LAD procedure is a more challenging problem, and it has not been attempted so far. The asymptotic properties of the LADEs have been obtained so far only for $p = 1$ under a set of conditions which are much stronger than what are needed in case of LSEs and WLSEs. Establishing the properties of the LADEs for general p is not immediate. We have established the consistency and asymptotic normality properties of the WLSEs for general p, under the same set of conditions as what are needed in case of the LSEs. Hence, it seems we have provided a comprehensive solution to this problem.

The rest of the paper is organized as follows. In Section 2 we provide some preliminaries. One component chirp model has been considered in Section 3. In Section 4 we consider the multicomponent chirp model. Simulation results and the analysis of two data sets have been presented in Section 5 and in Section 6, respectively. The case of general weight function has been discussed in Section 7 and finally, in Section 8 we conclude the paper.

2 Preliminaries

In order to establish the consistency and asymptotic normality of the WLSEs we need some number theoretic results and one famous number theoretic conjecture. We explicitly mention it here for easy reference.

RESULT 1: If $(\theta_1, \theta_2) \in (0,1) \times (0,1)$, and θ_2 is irrational, then the following results hold.

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \cos(\theta_1 n + \theta_2 n^2) &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \sin(\theta_1 n + \theta_2 n^2) = 0, \\
\lim_{N\to\infty} \frac{1}{N^{k+1}}\sum_{n=1}^{N} n^k \cos^2(\theta_1 n + \theta_2 n^2) &= \frac{1}{2(k+1)}, \\
\lim_{N\to\infty} \frac{1}{N^{k+1}}\sum_{n=1}^{N} n^k \sin^2(\theta_1 n + \theta_2 n^2) &= \frac{1}{2(k+1)}, \\
\lim_{N\to\infty} \frac{1}{N^{k+1}}\sum_{n=1}^{N} n^k \cos(\theta_1 n + \pi\theta_2 n^2)\sin(\pi\theta_1 n + \pi\theta_2 n^2) &= 0,
\end{aligned}$$

where $k = 0, 1, 2, \ldots$.

PROOF: The proof can be obtained from Vinogradov's [14] results. See Lahiri, Kundu and

Mitra [10] for details.

Comments: The proof of Vinogradov [14] is mathematically quite involved. But one important point it has been shown that if θ_2 is an irrational point, then Result 1 holds. Since in all practical applications, $\theta_2 = \pi\alpha$, for $\alpha \in (0,1)$, and π is irrational, hence Result 1 holds for all $\alpha \in (0,1)$.

The following famous number theoretic conjecture, see Montgomery [11], can not be established formally. But extensive numerical experiments indicate that it holds true.

Conjecture A: If $\theta_1, \theta_2, \theta'_1, \theta'_2 \in (0,\pi)$, and θ_2, θ'_2 both are irrational, then

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}N^k} \sum_{n=1}^{N} n^k \cos(\theta_1 n + \theta_2 n^2) \sin(\theta'_1 n + \theta'_2 n^2) = 0; \qquad k = 0,1,2,\ldots. \tag{2}$$

In addition if $\theta_2 \neq \theta'_2$, then

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}N^k} \sum_{n=1}^{N} n^k \cos(\theta_1 n + \theta_2 n^2) \cos(\theta'_1 n + \theta'_2 n^2) = 0; \qquad k = 0,1,2,\ldots \tag{3}$$

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}N^k} \sum_{n=1}^{N} n^k \sin(\theta_1 n + \theta_2 n^2) \sin(\theta'_1 n + \theta'_2 n^2) = 0; \qquad k = 0,1,2,\ldots \tag{4}$$

3 One component Chirp Model

In this section we consider the one component chirp model and it can be described as follows:

$$y(t) = A^0 \cos(\alpha^0 t + \beta^0 t^2) + B^0 \sin(\alpha^0 t + \beta^0 t^2) + X(t); \qquad t = 1,\ldots,N. \tag{5}$$

Here as mentioned before A^0 and B^0 are unknown real numbers, $|A^0|^2 + |B^0|^2$ is known as amplitude, and it is assumed that there exists an M, such that $0 < |A^0|, |B^0| < M$. The frequency $\alpha^0 \in (0,\pi)$ and the frequency rate $\beta^0 \in (0,\pi)$. The additive error random variables $X(t)$s are i.i.d. random variables with mean zero and finite variance $\sigma^2 > 0$, as mentioned before. The problem here is to estimate A^0, B^0, α^0 and β^0 based on a sample of size N, namely $\{y(1), \ldots, y(N)\}$. We use the following notation. $\theta = (A, B, \alpha, \beta)^\top$, $\theta^0 = (A^0, B^0, \alpha^0, \beta^0)^\top$,

$$\mu(t;\theta) = A\cos(\alpha t + \beta t^2) + B\sin(\alpha t + \beta t^2),$$

and $w(s)$ is a m-th degree polynomial, i.e.

$$w(s) = a_0 + a_1 s + a_2 s^2 + \ldots + a_m s^m, \tag{6}$$

here $a_0, a_1, \ldots, a_m$ are such that $\min_{0 \le s \le 1} w(s) > \gamma > 0$. Without loss of generality, it is assumed that $a_0 = 1$. Suppose K is such that $\sup_{0 \le s \le 1} w(s) \le K$.

Let us consider the following quantity:

$$Q(\theta) = \sum_{t=1}^{N} w\left(\frac{t}{N}\right)(y(t) - \mu(t;\theta))^2. \tag{7}$$

Suppose $\widehat{\theta}$ minimizes $Q(\theta)$, then $\widehat{\theta}$ is called the WLSE of θ^0. When $w(s) = 1$, $\widehat{\theta}$ becomes the LSE of θ^0. Although θ is a four dimensional vector, we will show that the minimization of $Q(\theta)$ can be performed by solving a two dimensional optimization problem. Before showing that we will establish the asymptotic properties of $\widehat{\theta}$.

Theorem 1: If $X(t)$s are i.i.d. random variables with mean zero and finite variance $\sigma^2 > 0$, $\alpha^0, \beta^0 \in (0, \pi)$ and $w(s)$ is the weight function as defined in (6), then $\widehat{\theta}$ is a strongly consistent estimator of θ^0.

Proof: See in Appendix A.

We need the following notations for development of the asymptotic distribution of $\widehat{\theta}$.

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{t=1}^{N} t^k w\left(\frac{t}{N}\right) &= \int_0^1 t^k w(t)dt = c_{k+1}, \\
\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{t=1}^{N} t^k w^2\left(\frac{t}{N}\right) &= \int_0^1 t^k w^2(t)dt = d_{k+1}; \quad k = 0, 1, 2, \ldots,
\end{aligned}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} d_1 & 0 & B^0 d_2 & B^0 d_3 \\ 0 & d_1 & -A^0 d_2 & -A^0 d_3 \\ B^0 d_2 & -A^0 d_2 & (A^{0^2} + B^{0^2}) d_3 & (A^{0^2} + B^{0^2}) d_4 \\ B^0 d_3 & -A^0 d_3 & (A^{0^2} + B^{0^2}) d_4 & (A^{0^2} + B^{0^2}) d_5 \end{bmatrix} \tag{8}$$

and

$$\boldsymbol{G} = \begin{bmatrix} c_1 & 0 & B^0 c_2 & B^0 c_3 \\ 0 & c_1 & -A^0 c_2 & -A^0 c_3 \\ B^0 c_2 & -A^0 c_2 & (A^{0^2} + B^{0^2}) c_3 & (A^{0^2} + B^{0^2}) c_4 \\ B^0 c_3 & -A^0 c_3 & (A^{0^2} + B^{0^2}) c_4 & (A^{0^2} + B^{0^2}) c_5 \end{bmatrix}. \tag{9}$$

Based on the same assumptions as in Theorem 1, it can be shown that with the proper normalization the distribution of $\widehat{\theta}$ is asymptotically normal having mean θ^0, and a dispersion matrix whose elements depend on α^0 and β^0. But based on the number theoretic Conjecture A, the following statement which provides a simplified form of the dispersion matrix, can be proved.

Statement 1: If (2), (3), (4) are true, then under the same assumptions as in Theorem 1, and if the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{G}$, as defined in (8) and (9), respectively are of full rank, then

$$\left(N^{1/2}(\widehat{A} - A^0), N^{1/2}(\widehat{B} - B^0), N^{3/2}(\widehat{\alpha} - \alpha^0), N^{5/2}(\widehat{\beta} - \beta^0)\right)^\top \xrightarrow{d} N_4\left(\mathbf{0}, 2\sigma^2\, \boldsymbol{G}^{-1}\boldsymbol{\Sigma}\boldsymbol{G}^{-1}\right).$$

Here $\xrightarrow{d}$ means convergence in distribution.

Proof: See in Appendix B. ∎

Statement 1 is helpful in deriving the variances of the estimators for large N. Moreover, it should be mentioned that even though Conjecture A have not yet been formally proved, extensive numerical experiments support these results, see for example Vinogradov [14] and Lahiri [8] in this respect. In our extensive simulation experiments it has been observed that the elements of the dispersion matrix do not depend on α^0 and β^0. Hence, it can be concluded that the experimental results do not contradict Conjecture A.

Observe that the matrix $\mathbf{\Sigma}$ is invertible if and only if

$$(d_3 - d_2^2)(d_5 - d_3^2) \neq (d_4 - d_2 d_3)^2,$$

similarly, the matrix $\boldsymbol{G}$ is invertible if and only if

$$(c_3 - c_2^2)(c_5 - c_3^2) \neq (c_4 - c_2 c_3)^2.$$

The explicit forms of c_{k+1} and d_{k+1} are as follows

$$\begin{aligned} c_{k+1} &= \frac{1}{k+1} + \frac{a_1}{k+2} + \ldots + \frac{a_m}{k+m+1} \\ d_{k+1} &= \frac{1}{k+1} + \frac{a_1^2}{k+3} + \ldots + \frac{a_m^2}{k+2m+1} + 2 \sum_{0 \le i < j \le m} \frac{a_i a_j}{k+i+j+1}. \end{aligned}$$

When $m = 0$, then $c_j = d_j$ for $j = 1, 2, \ldots$, and $2\sigma^2 \boldsymbol{G}^{-1} \mathbf{\Sigma} \boldsymbol{G}^{-1} = 2\sigma^2 \boldsymbol{G}^{-1}$, where

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 & \frac{1}{2}B^0 & \frac{1}{3}B^0 \\ 0 & 1 & -\frac{1}{2}A^0 & -\frac{1}{3}A^0 \\ \frac{1}{2}B^0 & -\frac{1}{2}A^0 & \frac{1}{3}(A^{0^2} + B^{0^2}) & \frac{1}{4}(A^{0^2} + B^{0^2}) \\ \frac{1}{3}B^0 & -\frac{1}{3}A^0 & \frac{1}{4}(A^{0^2} + B^{0^2}) & \frac{1}{5}(A^{0^2} + B^{0^2}) \end{bmatrix} = \mathbf{\Sigma}$$

and

$$\boldsymbol{G}^{-1} = \frac{2}{A^{0^2} + B^{0^2}} \begin{bmatrix} \frac{1}{2}\left(A^{0^2} + 9B^{0^2}\right) & -4A^0B^0 & -18B^0 & 15B^0 \\ -4A^0B^0 & \frac{1}{2}\left(9A^{0^2} + B^{0^2}\right) & 18A^0 & -15A^0 \\ -18B^0 & 18A^0 & 96 & -90 \\ 15B^0 & -15A^0 & -90 & 90 \end{bmatrix},$$

which is the asymptotic variance matrix of the LSE $\widehat{\theta}$.

Now we will describe how we can obtain the WLSEs as a two dimensional optimization problem. Note that the WLSEs of $\widehat{A}$, $\widehat{B}$, $\widehat{\alpha}$, $\widehat{\beta}$ are obtained by minimizing $Q(\theta)$ as defined in (7). Since A and B are the linear parameters, therefore, for a given α and β, $\widehat{A}(\alpha,\beta)$ and $\widehat{B}(\alpha,\beta)$ minimize $Q(\theta)$, where

$$\widehat{A}(\alpha,\beta) = \frac{b_1(\alpha,\beta)a_{22}(\alpha,\beta) - b_2(\alpha,\beta)a_{12}(\alpha,\beta)}{a_{11}(\alpha,\beta)a_{22}(\alpha,\beta) - a_{21}(\alpha,\beta)a_{12}(\alpha,\beta)},$$

$$\widehat{B}(\alpha,\beta) = \frac{b_1(\alpha,\beta)a_{21}(\alpha,\beta) - b_2(\alpha,\beta)a_{11}(\alpha,\beta)}{a_{21}(\alpha,\beta)a_{12}(\alpha,\beta) - a_{11}(\alpha,\beta)a_{22}(\alpha,\beta)},$$

$$\begin{aligned} a_{11}(\alpha,\beta) &= \sum_{t=1}^{N} w\left(\frac{t}{N}\right)\cos^2(\alpha t + \beta t^2), \quad a_{22}(\alpha,\beta) = \sum_{t=1}^{N} w\left(\frac{t}{N}\right)\sin^2(\alpha t + \beta t^2), \\ a_{12}(\alpha,\beta) &= \sum_{t=1}^{N} w\left(\frac{t}{N}\right)\sin(\alpha t + \beta t^2)\cos(\alpha t + \beta t^2) = a_{21}(\alpha,\beta), \\ b_1(\alpha,\beta) &= \sum_{t=1}^{N} w\left(\frac{t}{N}\right) y(t)\cos(\alpha t + \beta t^2), \quad b_2(\alpha,\beta) = \sum_{t=1}^{N} w\left(\frac{t}{N}\right) y(t)\sin(\alpha t + \beta t^2). \end{aligned}$$

Hence $\widehat{\alpha}$ and $\widehat{\beta}$ can be obtained by minimizing $Q(\widehat{A}(\alpha,\beta),\widehat{B}(\alpha,\beta),\alpha,\beta)$ with respect to α and β. Then $\widehat{A}$ and $\widehat{B}$ can be obtained as $\widehat{A} = \widehat{A}(\widehat{\alpha},\widehat{\beta})$ and $\widehat{B} = \widehat{B}(\widehat{\alpha},\widehat{\beta})$, respectively.

4 Multicomponent Chirp Model

In this section we consider the multicomponent chirp model as it has been defined in (1). First, we consider the WLSEs and discuss its properties, and then we consider the sequential WLSEs. It is assumed that p is known and without loss of generality

$$0 < A_p^{0^2} + B_p^{0^2} < \ldots < A_1^{0^2} + B_1^{0^2} < M. \tag{10}$$

We use the following notations for $k = 1, \ldots, p$: $\theta_k = (A_k, B_k, \alpha_k, \beta_k)^\top$ and $\theta_k^0 = (A_k^0, B_k^0, \alpha_k^0, \beta_k^0)^\top$. The WLSEs of $\theta_1^0, \ldots, \theta_p^0$ can be obtained by minimizing

$$Q_p(\theta_1, \ldots, \theta_p) = \sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left(y(t) - \sum_{k=1}^{p} \mu(t;\theta_k)\right)^2, \tag{11}$$

with respect to the unknown parameters. Let us denote them as $\widehat{\theta}_1 \ldots, \widehat{\theta}_p$, respectively. It can be shown as before that WLSEs of $\theta_1, \ldots, \theta_p$ can be obtained by solving a $2p$ dimensional optimization problem. Now we provide the asymptotic properties of the WLSEs of $\theta_1, \ldots, \theta_p$ without proofs, because these can be proved along the similar lines as the one component model and using Conjecture A.

Theorem 2: If $X(t)$s are same as defined in Theorem 1, $w(s)$ is the weight function as defined in (6), $\alpha_k^0, \beta_k^0 \in (0, \pi)$, then $\widehat{\theta}_k$ is a strongly consistent estimator of θ_k^0 for $k = 1, \ldots, p$.

Along the same line as the one component model, it can be shown that with the proper normalization, the asymptotic distribution of $\widehat{\theta}_k$ is normally distributed with the mean vector θ_k^0 and a dispersion matrix whose elements depend on the frequency and frequency rate parameters also. But based on Conjecture A, similar to Statement 1, we can make the following statement which provides a simplified form of the dispersion matrix. Statement 2 can be proved along the same line as Statement 1.

Statement 2: If (2), (3), (4) are true, then under the same assumptions as in Theorem 2,

$$\left((\widehat{\theta}_1-\theta_1^0)^\top \boldsymbol{D} \ldots (\widehat{\theta}_p-\theta_p^0)^\top \boldsymbol{D}\right)^\top \xrightarrow{d} N_{4p}\left(\mathbf{0}, 2\sigma^2 \boldsymbol{H}\right),$$

where $\mathbf{D} = \text{diag}\{N^{\frac{1}{2}}, N^{\frac{1}{2}}, N^{\frac{3}{2}}, N^{\frac{5}{2}}\}$ and $\boldsymbol{H}$ is a $4p \times 4p$ block diagonal matrix with the k-th block as $\boldsymbol{G}_k^{-1}\boldsymbol{\Sigma}_k\boldsymbol{G}_k^{-1}$. Here $\boldsymbol{\Sigma}_k$ and $\boldsymbol{G}_k$ can be obtained from $\boldsymbol{\Sigma}$ and $\boldsymbol{G}$, by replacing A^0, B^0, α^0, β^0 with A_k^0, B_k^0, α_k^0, β_k^0, respectively for $k = 1, \ldots, p$.

The procedure to obtain the sequential WLSEs is outlined below. Consider

$$Q_1(\theta_1) = \sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left(y(t) - \mu(t;\theta_1)\right)^2. \tag{12}$$

Obtain the sequential WLSEs of θ_1 by minimizing (12) with respect to θ_1, and let us denote it by $\widetilde{\theta}_1 = (\widetilde{A}_1, \widetilde{B}_1, \widetilde{\alpha}_1, \widetilde{\beta}_1)^\top$. Note that the minimization of $Q_1(\theta_1)$ can be obtained as a two dimensional optimization problem. At the second stage, obtain the new data $\{y_1(t); t = 1, \ldots, N\}$, where $y_1(t) = y(t) - \mu(t; \widetilde{\theta}_1)$, for $t = 1, \ldots, N$. Consider,

$$Q_2(\theta_2) = \sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left(y_1(t) - \mu(t;\theta_2)\right)^2. \tag{13}$$

Obtain the sequential WLSEs of θ_2 by minimizing (13) with respect to θ_2. We denote it by $\widetilde{\theta}_2 = (\widetilde{A}_2, \widetilde{B}_2, \widetilde{\alpha}_2, \widetilde{\beta}_2)^\top$. At the third stage, obtain the new data $\{y_2(t); t = 1, \ldots, N\}$, where $y_2(t) = y_1(t) - \mu(t; \widetilde{\theta}_2)$, for $t = 1, \ldots, N$ and continue the process till the p-th stage. Therefore, sequentially we can obtain $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_p$. It is clear that the computational complexity of the sequential WLSEs is same as the sequential LSEs. Both of them require solving p two dimensional optimization problem. Hence, the method which is being used to compute the LSEs can be used to compute the WLSEs also.

Now we will show that $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_p$ are strongly consistent estimators of $\theta_1^0, \ldots, \theta_p^0$, respectively. In the following theorems, $\boldsymbol{\Sigma}_k$ and $\mathbf{G}_k$, $k = 1, \ldots, p$ are same as defined in Statement 2.

Theorem 3: Under the same assumption as in Theorem 2, $\widetilde{\theta}_1$ is a strongly consistent estimator of θ_1^0.

Proof: See in Appendix C.

Regarding the asymptotic equivalence between the WLSEs and sequential WLSEs, we can prove the following statement based on Conjecture A.

Statement 3: If (2), (3), (4) are true, then under the same assumption as in Theorem 2

$$\left(N^{1/2}(\widetilde{A}_1 - A_1^0), N^{1/2}(\widetilde{B}_1 - B_1^0), N^{3/2}(\widetilde{\alpha}_1 - \alpha_1^0), N^{5/2}(\widetilde{\beta}_1 - \beta_1^0)\right)^\top \xrightarrow{d} N_4\left(\mathbf{0}, 2\sigma^2\ \boldsymbol{G}_1^{-1}\boldsymbol{\Sigma}_1\boldsymbol{G}_1^{-1}\right).$$

Proof: See in Appendix D.

Theorem 4: Under the same assumption as in Theorem 2, $\widetilde{\theta}_2$ is a strongly consistent estimator of θ_2^0.

Proof: See in Appendix D.

Similar to Statement 3, we can prove the following Statement 4, based on Conjecture A. Hence, it is avoided.

Statement 4: If (2), (3), (4) are true, then under the same assumption as in Theorem 2

$$\left(N^{1/2}(\widetilde{A}_2 - A_2^0), N^{1/2}(\widetilde{B}_2 - B_2^0), N^{3/2}(\widetilde{\alpha}_2 - \alpha_2^0), N^{5/2}(\widetilde{\beta}_2 - \beta_2^0)\right)^\top \xrightarrow{d} N_4\left(\mathbf{0}, 2\sigma^2\ \boldsymbol{G}_2^{-1}\boldsymbol{\Sigma}_2\boldsymbol{G}_2^{-1}\right).$$

Now if the process is repeated beyond p steps and $\widehat{A}_{p+1}$ and $\widehat{B}_{p+1}$ are the estimates of A and B at step $p+1$, respectively, then we have the following result.

Theorem 5: Under the same assumption as in Theorem 2, $\widehat{A}_{p+1} \xrightarrow{a.s} 0$ and $\widehat{B}_{p+1} \xrightarrow{a.s} 0$.

Proof: See in Appendix D.

5 Simulation Results

In this section, we provide a demonstration of the performance of the proposed WLSEs through numerical simulations. We consider a simple one-component model:

$$y(t) = 4\cos(0.38\pi t + 0.01\pi t^2) + 3\sin(0.38\pi t + 0.01\pi t^2) + X(t). \tag{14}$$

Here, $X(t)$s are i.i.d. random variables simulated from Gaussian distribution with mean 0 and variance σ^2. For different sample sizes varying from 100 to 500 and for error standard deviations $\sigma = 0.5, 1, 2, 5$ and 10, we generate data from the above model. To each of these data sets, we add a few outliers to assess the robustness of the proposed estimators. These outliers are added to 10 percent of the middle section of the data. The outliers are generated from normal distribution with mean 0 and standard deviation 50.

Since the outliers are added in the middle, we choose the following weight function for the computation of WLSEs:

$$w\left(\frac{t}{N}\right) = \frac{1}{4} - \frac{t}{N} + \frac{t^2}{N^2}. \tag{15}$$

It is important to note that the choice of weights is vital for the computation of optimal WLSEs. This weight function puts less weight in the middle of the data. We compute the WLSEs for different sample sizes and different error variances. For a comparative study, we also compute the usual LSEs and the robust LADEs. It may be mentioned here that there are other robust estimators available in the literature, see for example Rousseeuw and Leroy [13]. Among the different robust estimators, other than the LADEs, Huber's M-estimators are also quite popular. The behavior of Huber's estimators are very similar to the LADEs. Unfortunately, in case of chirp models, the properties of Huber's M-estimators are not yet established. Moreover, computationally it is more challenging than the LADEs, depending on the choice of the influence function. Hence, it has not been attempted here. For each set of parameters, the experiment

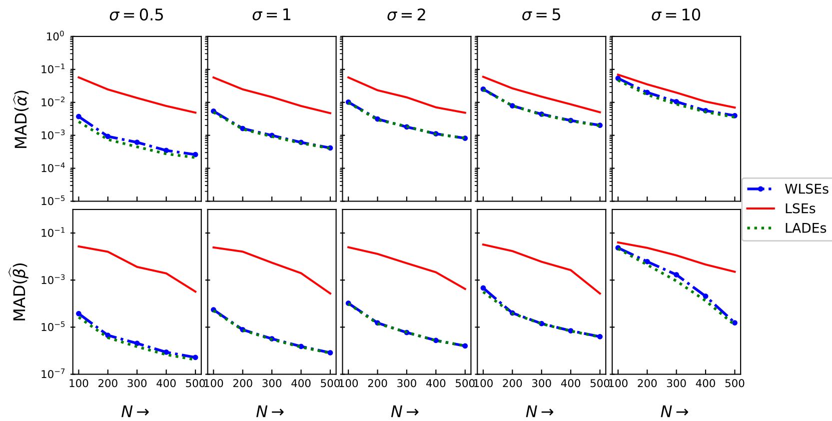

Figure 1: In each sub-plot, the graphed line represents the MADs of the WLSEs, LSEs, and LADEs of the simulated one component model.

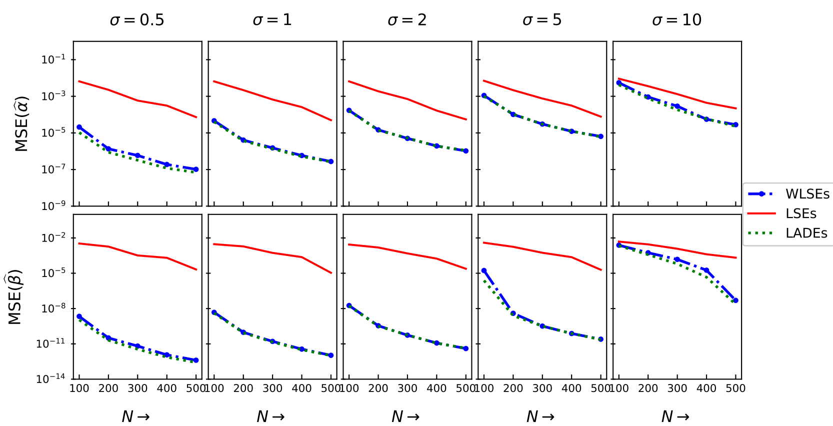

Figure 2: In each sub-plot, the graphed line represents the MSEs of the WLSEs, LSEs, and LADEs of the simulated one component model.

is replicated 1000 times and the mean absolute deviations (MADs) as well as the mean square errors (MSEs) of WLSEs, LSEs, and LADEs are reported here. Computing both the LSE and WLSEs, one needs to solve two-dimensional optimization problem, where as in case of LADEs it is a four dimensional optimization problem. We have mainly used Nelder-Mead algorithm to solve both the two and four dimensional optimization problems. We have tried with other set of parameters values also, but the overall pattern of the behavior of the different estimators remains same. Some noteworthy observations are listed below.

- As the sample size increases, the MSEs and MADs of the proposed estimators decrease, validating their consistency property. This is also observed for the LSEs and LADEs.

- As the standard deviation increases, the MSEs and MADs curves move upward along the y-axis, thereby implying the estimators perform better with increasing signal to noise ratio.

- WLSEs perform much better than LSEs in terms of both MADs and MSEs for all values of N and σ.

- WLSEs are observed to be robust to the outliers.

- WLSEs perform at par with the LADEs for all values of N and for all σ. WLSEs and LADEs are more efficient than the LSEs in presence of outliers.

From the simulation results, we can conclude that the WLSEs are robust than the LSEs in presence of outliers. As a matter of fact, their performance is as good as the LADEs which are known to be robust. Another important point to note here is that although the LADEs perform the best, computing them for a one component model involves solving a 4 dimensional optimisation problem as there is no closed form solution for the linear parameters in this case. For the multiple component model, the problem becomes more complex because unlike the least squares and the weighted least squares methods, this problem cannot be broken down into a series of two-dimensional optimisation problems and will require a $4p$ dimensional search. Also initial guesses are required not only for the frequency and frequency rate parameters but also for the linear parameters. Since the number of components required for analysis of real data can be very large, we use the sequential LSEs and sequential WLSEs for the parameter estimation in the next section and compare their performances.

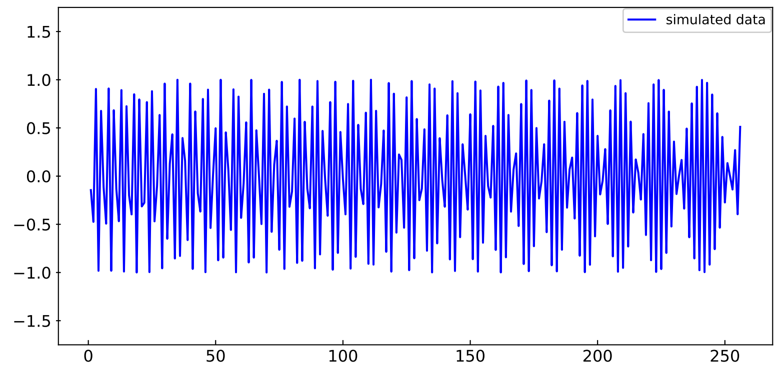

Figure 3: Simulated data.

6 Data Analysis

6.1 Simulated Data Analysis

We illustrate the performance of sequential WLSEs in comparison with the sequential LSEs through a simulated example in this section. We consider a two-component chirp model given by:

$$y(t) = 10\cos(0.80\pi t + 0.001\pi t^2) + 10\sin(0.80\pi t + 0.001\pi t^2)$$
$$+8\cos(0.48\pi t + 0.002\pi t^2) + 8\sin(0.48\pi t + 0.002\pi t^2) + X(t)$$

for $t = 1, \ldots, 256$. Here $e(t)$s are independent and identically normally distributed with mean 0 and standard deviation 5. Figure 3 shows the simulated data from the above-defined model equation. To the simulated data, we add 10 outliers in the middle of the data set. These outliers are generated from normal distribution as well. However, in this case they have standard deviation 20. To this contaminated data, we fit a two-component chirp model using the sequential LSEs first. Figure 4 shows the estimated signal using the LSEs along with the simulated data. Figure 4 reveals that using the sequential LSEs to fit the model to the data with outliers leads to unsatisfactory results. Since the LSEs are sensitive to perturbations in the data, it is expected

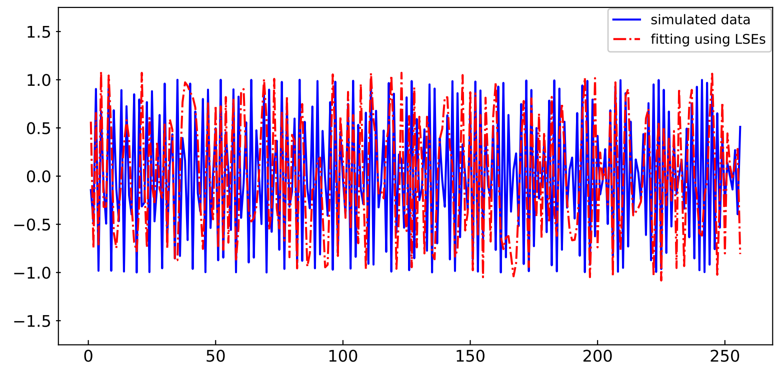

Figure 4: The estimated signal using sequential LSEs along with the simulated data.

that sequential WLSEs may yield better results than the sequential LSEs. To choose an appropriate weight function, we first plot the residuals obtained by fitting the model using sequential LSEs. Figure 5 shows the residual plot obtained after fitting the two-component chirp model to the simulated data using the sequential LSEs. It can be seen that there are large residuals in the middle indicating the presence of outliers. We set the weight function to the following:

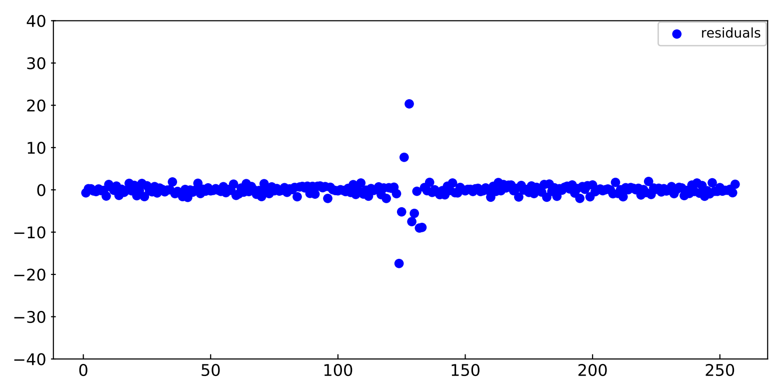

Figure 5: The residual plot of the simulated data and the fitted LSEs signal.

$$w\left(\frac{t}{N}\right) = \frac{1}{4} - \frac{t}{N} + \frac{t^2}{N^2}, \tag{16}$$

a convex function so that lesser weights are assigned to the observations in the middle, given the residuals are large in the middle section in Figure 5. Using this weight function, we compute

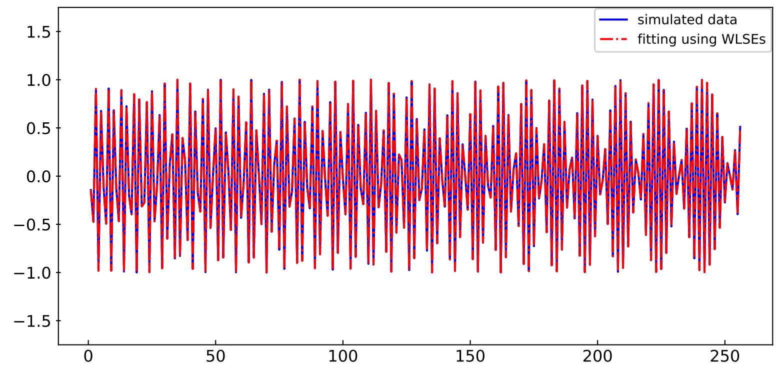

Figure 6: The estimated signal using sequential WLSEs along with the simulated data.

the sequential WLSEs and the estimated signal using these estimators along with the simulated data is shown in Figure 6.

We observe in Figure 6 that the estimated signal using the sequential WLSEs fits the data reasonably well. On comparing this with Figure 4, it can be concluded that in the presence of outliers, the sequential WLSEs provide a satisfactory fit.

6.2 Real Data Analysis

In this section, the applicability of proposed sequential WLSEs is demonstrated through analysis of an observed EEG signal. Only a part of the observed signal has been used and Figure 7 shows the original signal with 64 data points.

To this data, we add outliers to first five observations. To test the sensitivity of sequential LSEs to these outliers, we first fit the model using the sequential least squares method. Figure 8 shows the estimated chirp signal along with the original EEG signal. Next, we use the sequential WLSEs to fit the model to the EEG data. For their computation, the following weight function is used:

$$w\left(\frac{t}{N}\right) = a_0 + a_1\frac{t}{N}, \tag{17}$$

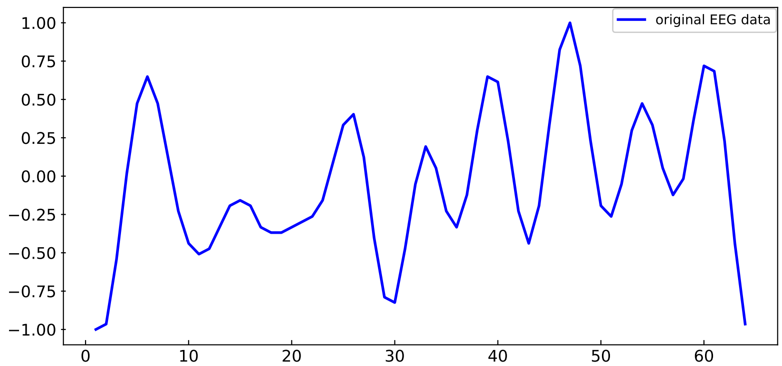

Figure 7: A segment of EEG data.

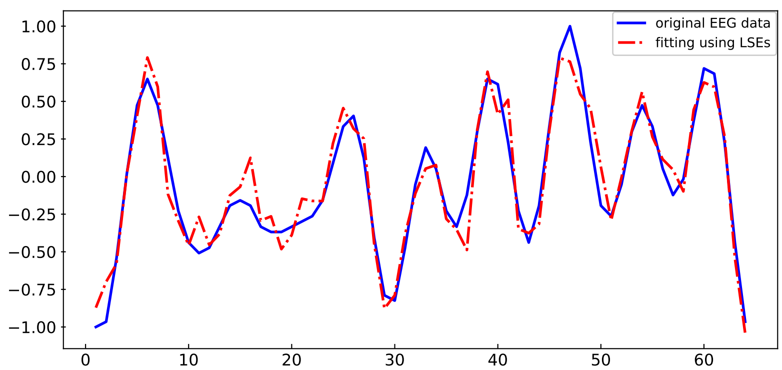

Figure 8: The estimated signal using sequential LSEs along with the EEG data.

with $a_0 = a_1 = \frac{1}{20}$. Note that we choose an increasing weight function as the simulated outliers are added at the beginning of the data. In Figure 9, we plot the estimated signal along with the original signal. It can be seen that there is an improvement in the fitting when the sequential WLSEs are used instead of the sequential LSEs. This improvement is quantified in terms of the residual sum of squares also. The residual sum of squares obtained by using the sequential LSEs for the fitting is 1.1177, whereas that obtained using the sequential WLSEs is 0.8961. Clearly, the sequential WLSEs outperform the sequential LSEs in terms of the fitting.

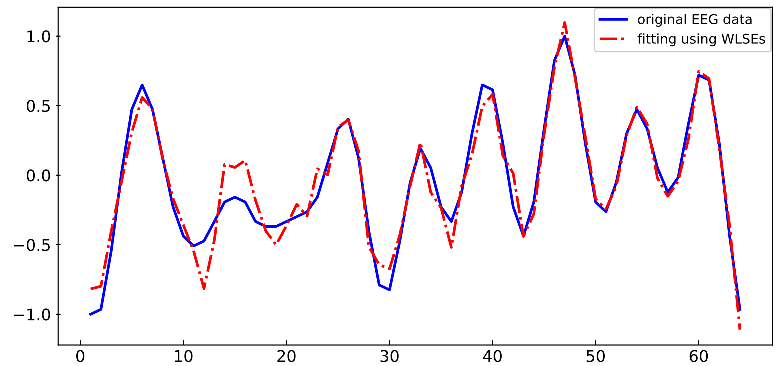

Figure 9: The estimated signal using sequential WLSEs along with the EEG data.

7 General Weight Function

So far we have assumed that the weight function $w(s)$ to be a polynomial weight function. In this section we will outline how the method can be extended for a general class of weight functions and for the single component chirp model (5). The result can be easily extended for the multiple chirp model (1) along the same line as before. In this section we will be using some of the same notations for convenience, which have also been used before in a similar context. It should not create any problem. We need the following assumption and result for further development.

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

Result 2: Suppose $(\theta_1,\theta_2) \in (0,\pi)\times(0,\pi)$, and it satisfies Result 1, $w(s)$ satisfies Assumption 1, then

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

Proof: See in Appendix E.

Theorem 6: If $X(t)$, $A^0, B^0, \alpha^0, \beta^0$ satisfy the same assumptions as in Theorem 1, and $w(s)$

satisfies Assumption 1, then $\widehat{\theta}$ is a strongly consistent estimator of θ^0.

Proof: The proof of Theorem 6 mainly depends on Lemma 1. It can be easily seen that the proof of Lemma 1 goes through if $w(s)$ satisfies Assumption 1. ∎

We need the following notations. Similarly as before, for $k = 0, 1, 2, \ldots$,

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{t=1}^{N} t^k w\left(\frac{t}{N}\right) = \int_0^1 t^k w(t)dt = c_{k+1} \tag{18}$$

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \sum_{t=1}^{N} t^k w^2\left(\frac{t}{N}\right) = \int_0^1 t^k w^2(t)dt = d_{k+1}. \tag{19}$$

These c_ks and d_ks are defined for a weight function that satisfies Assumption 1.

Regarding the asymptotic distribution, the following statement can be proved along the same line as the proof of Statement 1, using Conjecture A. Hence, it is avoided.

Statement 5: If (2), (3), (4) are true, $X(t)$, $A^0, B^0, \alpha^0, \beta^0$ satisfy the same assumptions as in Theorem 1, $w(s)$ satisfies Assumption 1, $\mathbf{\Sigma}$ and $\boldsymbol{G}$ are same as in Theorem 1, where c_k and d_k are as above, and the matrix $\mathbf{\Sigma}$ and $\boldsymbol{G}$ are of full rank, then

$$\left(N^{1/2}(\widehat{A} - A^0), N^{1/2}(\widehat{B} - B^0), N^{3/2}(\widehat{\alpha} - \alpha^0), N^{5/2}(\widehat{\beta} - \beta^0)\right)^\top \xrightarrow{d} N_4\left(\mathbf{0}, 2\sigma^2\, \boldsymbol{G}^{-1}\mathbf{\Sigma}\boldsymbol{G}^{-1}\right).$$

The results of the WLSEs and sequential WLSEs for multiple chirp model follow exactly the same way as before and therefore, they are omitted.

8 Concluding Remarks

In this paper, we have considered the WLSEs for multiple chirp model. The proposed estimators are robust in presence of a few outliers and they outperform the LSEs in terms of lower mean absolute deviations and mean squared errors. The asymptotic properties of the proposed estimators have been established and the proposed WLSEs have the same convergence rates as the LSEs. Extensive simulations have emphasized that the performance of WLSEs is as good as the LADEs. But the strong consistency and asymptotic distribution results are not available for LADEs in case of multiple chirp model and are open problems. Also computing the LADEs for p component signal involves solving a $4p$ dimensional optimisation problem as there is no closed form solution for the linear parameters in this case. Another important point is that WLSEs can be estimated component-wise sequentially and the sequential WLSEs have the same asymptotic distribution as the WLSEs. Therefore, the proposed weighted least squares is a method which is as good as the LAD estimation procedure computationally and are strongly consistent and asymptotically normal. Also the method is relatively much easier to implement. The analysis of a real data set have been performed by artificially including outliers and the performances are quite satisfactory. One important point to be noted here is that the performance of the proposed estimation method depends on the choice of the weight function. Further studies are needed in that direction.

Data Availability Statement

This manuscript has no associated data.

Acknowledgements

The authors would like to thank the unknown reviewers for their constructive comments which have helped to improve the manuscript significantly.

Appendix A

We need the following lemmas to prove Theorem 1.

LEMMA 1: If $\{X(t)\}$s are i.i.d. random variables with mean 0 and variance σ^2, and $w(s)$ are same as in Theorem 1, and $0 < \alpha, \beta < \pi$, then as $N \to \infty$,

$$\sup_{\alpha,\beta} \left| \frac{1}{N} \sum_{t=1}^{N} X(t) w\left(\frac{t}{N}\right) \cos(\alpha t) \cos(\beta t^2) \right| \stackrel{a.s.}{\to} 0.$$

PROOF OF LEMMA 1: Consider the following random variable

$$Z(t) = \begin{cases} X(t) & \text{if} \quad |X(t)| \le t^{\frac{3}{4}} \\ 0 & o.w. \end{cases}$$

Then

$$\begin{aligned} \sum_{t=1}^{\infty} P[X(t) \neq Z(t)] &= \sum_{t=1}^{\infty} P[|X(t)| > t^{\frac{3}{4}}] = \sum_{t=1}^{\infty} \sum_{2^{t-1} \le s < 2^t} P[|X(t)| > s^{\frac{3}{4}}] \\ &\le \sum_{t=1}^{\infty} \sum_{2^{t-1} \le s < 2^t} P\left[|X(1)| > 2^{(t-1)\frac{3}{4}}\right] \le \sum_{t=1}^{\infty} 2^t P\left[|X(1)| > 2^{(t-1)\frac{3}{4}}\right] \\ &\le \sum_{t=1}^{\infty} 2^t \frac{E|X(1)|^2}{2^{(t-1)\frac{3}{2}}} \le C \sum_{t=1}^{\infty} 2^{-\frac{t}{2}} < \infty. \end{aligned}$$

Therefore, $\{X(t)\}$ and $\{Z(t)\}$ are equivalent sequences. So

$$\sup_{\alpha,\beta} \left| \frac{1}{N} \sum_{t=1}^{N} X(t) w\left(\frac{t}{N}\right) \cos(\alpha t) \cos(\beta t^2) \right| \stackrel{a.s.}{\to} 0 \Leftrightarrow$$

$$\sup_{\alpha,\beta} \left| \frac{1}{N} \sum_{t=1}^{N} Z(t) w\left(\frac{t}{N}\right) \cos(\alpha t) \cos(\beta t^2) \right| \stackrel{a.s.}{\to} 0.$$

Let $U(t) = Z(t) - E(Z(t))$. Therefore,

$$\sup_{\alpha,\beta}\left|\frac{1}{N}\sum_{t=1}^{N} E(Z(t))w\left(\frac{t}{N}\right)\cos(\alpha t)\cos(\beta t^2)\right| \leq \frac{K}{N}\sum_{t=1}^{N}|E(Z(t))| = \frac{K}{N}\sum_{t=1}^{N}\left|\int_{|x|<t^{\frac{3}{4}}} x dF(x)\right| \to 0,$$

where $\sup_{0\leq s\leq 1} w(s) \leq K$, same as defined in Section 3 and $F(x)$ is the distribution of $X(1)$. Therefore, it is enough to prove that

$$\sup_{\alpha,\beta}\left|\frac{1}{N}\sum_{t=1}^{N} U(t)w\left(\frac{t}{N}\right)\cos(\alpha t)\cos(\beta t^2)\right| \stackrel{a.s.}{\to} 0.$$

For any fixed α, β and $\epsilon > 0$, and $0 \leq h \leq \frac{1}{4KN^{\frac{3}{4}}}$, we have

$$P\left[\left|\frac{1}{N}\sum_{t=1}^{N} U(t)w\left(\frac{t}{N}\right)\cos(\alpha t)\cos(\beta t^2)\right| \geq \epsilon\right] \leq 2e^{-hN\epsilon}\prod_{t=1}^{N} E(e^{hU(t)w\left(\frac{t}{N}\right)\cos(\alpha t)\cos(\beta t^2)})$$
$$\leq 2e^{-hN\epsilon}\prod_{t=1}^{N}(1 + h^2\sigma^2) \leq 2^{-hN\epsilon+Nh^2\sigma^2}.$$

The first inequality follows from Markov inequality. From the definition of $Z(t)$, $V(U(t)) = V(Z(t)) \leq V(X(t)) = \sigma^2$. Since $|hU(t)w\left(\frac{t}{N}\right)\cos(\alpha t)\cos(\beta t^2)| \leq \frac{1}{2}$, and for $|x| \leq \frac{1}{2}$, $e^x \leq 1 + x + x^2$, the second inequality holds true.

Choose $h = \dfrac{1}{4KN^{\frac{3}{4}}}$, therefore for large N,

$$\left[\left|\frac{1}{N}\sum_{t=1}^{N} U(t)w\left(\frac{t}{N}\right)\cos(\alpha t)\cos(\beta t^2)\right| \geq \epsilon\right] \leq 2e^{-\frac{N^{\frac{1}{4}}\epsilon}{4}+\frac{\sigma^2}{16N^{\frac{1}{2}}}} \leq 4e^{-\frac{N^{\frac{1}{4}}\epsilon}{4}}.$$

Let $J = N^6$, and choose J points $(\alpha_1, \beta_1), \ldots, (\alpha_J, \beta_J)$, such that for any point (α, β) in $[0, \pi] \times [0, \pi]$, we have a point (α_k, β_k) satisfying

$$|\alpha_k - \alpha| \leq \frac{\pi}{N^3} \quad \text{and} \quad |\beta_k - \beta| \leq \frac{\pi}{N^3}.$$

Now by Taylor series expansion can be used to estimate

$$|\cos(\beta t^2) - \cos(\beta_k t^2)| \leq t^2|\beta - \beta_k| \quad \text{and} \quad |\cos(\alpha t) - \cos(\alpha_k t)| \leq |t||\alpha - \alpha_k|$$

therefore,

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{t=1}^{N} U(t) w\left(\frac{t}{N}\right)\left\{\cos(\alpha t)\cos(\beta t^2) - \cos(\alpha_k t)\cos(\beta_k t^2)\right\}\right| \\
&\leq \left|\frac{1}{N}\sum_{t=1}^{N} U(t) w\left(\frac{t}{N}\right)\cos(\alpha t)\left\{\cos(\beta t^2) - \cos(\beta_k t^2)\right\}\right| \\
&+ \left|\frac{1}{N}\sum_{t=1}^{N} U(t) w\left(\frac{t}{N}\right)\cos(\beta_k t^2)\left\{\cos(\alpha t) - \cos(\alpha_k t)\right\}\right| \\
&\leq C\left[\frac{1}{N}\sum_{t=1}^{N} t^{\frac{3}{4}} t^2 \frac{\pi}{N^3} + \frac{1}{N}\sum_{t=1}^{N} t^{\frac{3}{4}} t \frac{\pi}{N^3}\right] \leq C\left[\frac{\pi}{N^{\frac{1}{4}}} + \frac{\pi}{N^{\frac{5}{4}}}\right] \to 0.
\end{aligned}$$

Therefore, for large N

$$\begin{aligned}
&P\left[\sup_{a,b}\left|\frac{1}{N}\sum_{t=1}^{N} U(t) w\left(\frac{t}{N}\right)\cos(\alpha t)\cos(\beta t^2)\right| \geq 2\epsilon\right] \\
&\leq P\left[\max_{k \leq N^6}\left|\frac{1}{N}\sum_{t=1}^{N} U(t) w\left(\frac{t}{N}\right)\cos(\alpha_k t)\cos(\beta_k t^2)\right| \geq 2\epsilon\right] \leq 4N^6 e^{-\frac{N^{\frac{1}{4}}\epsilon}{4}}.
\end{aligned}$$

Since $\sum_{N=1}^{\infty} N^6 e^{-\frac{N^{\frac{1}{4}}\epsilon}{4}} < \infty$, by using Borel-Cantelli lemma the result follows. ∎

Lemma 2: Let us denote

$$S_c = \{\theta : \theta = (A, B, \alpha, \beta)^\top, |\theta - \theta^0| \geq 4c\}.$$

If there exists a $c > 0$,

$$\underline{\lim} \inf_{\theta \in S_c} \frac{1}{N}[Q(\theta) - Q(\theta^0)] > 0 \quad a.s. \tag{20}$$

then $\widehat{\theta}$ is a strongly consistent estimator of θ^0.

Proof of Lemma 2: It follows by simple argument by contradiction, exactly similar to a lemma by Wu [15]. ∎

Proof of Theorem 1: Consider

$$\begin{aligned}\frac{1}{N}[Q(\theta)-Q(\theta^0)] &= \frac{1}{N}\left[\sum_{t=1}^{N} w\left(\frac{t}{N}\right)(y(t)-\mu(t;\theta))^2 - \sum_{t=1}^{N} w\left(\frac{t}{N}\right)X^2(t)\right] \\ &= \frac{1}{N}\left[\sum_{t=1}^{N} w\left(\frac{t}{N}\right)(\mu(t;\theta^0)-\mu(t;\theta))^2\right] \\ &\quad +\frac{2}{N}\left[\sum_{t=1}^{N} w\left(\frac{t}{N}\right)X(t)(\mu(t;\theta^0)-\mu(t;\theta))\right] \\ &\quad = f_1(\theta)+f_2(\theta).\end{aligned}$$

Here

$$f_1(\theta) = \frac{1}{N}\sum_{t=1}^{N} w\Big(\frac{t}{N}\Big)\Big(\mu(t;\theta^0)-\mu(t;\theta)\Big)^2$$
$$f_2(\theta) = \frac{2}{N}\sum_{t=1}^{N} w\left(\frac{t}{N}\right)X(t)\Big(\mu(t;\theta^0)-\mu(t;\theta)\Big).$$

Consider

$$S_{c,1} = \{\theta : \theta = (A,B,\alpha,\beta)^\top, |A-A^0|\ge c\}$$
$$S_{c,2} = \{\theta : \theta = (A,B,\alpha,\beta)^\top, |B-B^0|\ge c\}$$
$$S_{c,3} = \{\theta : \theta = (A,B,\alpha,\beta)^\top, |\alpha-\alpha^0|\ge c\}$$
$$S_{c,4} = \{\theta : \theta = (A,B,\alpha,\beta)^\top, |\beta-\beta^0|\ge c\}.$$

Then $S_c \subset S_{c,1}\cup S_{c,2}\cup S_{c,3}\cup S_{c,4} = S$. Therefore,

$$\underline{\lim}\inf_{\theta\in S_c} f_1(\theta) \ge \underline{\lim}\inf_{\theta\in S} f_1(\theta) = \underline{\lim}\inf_{\theta\in\cup_j S_{c,j}} f_1(\theta).$$

Now

$$\begin{aligned}\underline{\lim}\inf_{\theta\in S_{c,1}} f_1(\theta) &= \underline{\lim}\inf_{|A-A^0|\geq c}(A-A^0)^2\frac{1}{N}\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\cos^2(\alpha^0 t+\beta^0 t^2)\\ &\geq \gamma\ \underline{\lim}\inf_{|A-A^0|\geq c}(A-A^0)^2\frac{1}{N}\sum_{t=1}^{N}\cos^2(\alpha^0 t+\beta^0 t^2) > 0. \qquad \text{(using Result 1).}\end{aligned}$$

Similarly, it can be shown for $S_{c,2}, S_{c,3}$ and $S_{c,4}$ also. Therefore,

$$\underline{\lim}\inf_{\theta\in S_c} f_1(\theta) > 0.$$

Since from Lemma 1, it follows that

$$\lim\sup_{\theta} |f_2(\theta)| = 0,$$

hence, we have

$$\underline{\lim}\inf_{\theta\in S_c}\frac{1}{N}[Q(\theta)-Q(\theta^0)] > 0 \qquad a.s.$$

Using Lemma 2, the result follows. ∎

Appendix B

In this Appendix we provide the Proof of Statement 1 based on Conjecture A.

Since

$$Q(\theta) = \sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left(y(t)-\mu(t;\theta)\right)^2,$$

therefore,

$$Q'(\theta^0) = \begin{bmatrix} \frac{\partial Q(\theta)}{\partial A} \\ \frac{\partial Q(\theta)}{\partial B} \\ \frac{\partial Q(\theta)}{\partial \alpha} \\ \frac{\partial Q(\theta)}{\partial \beta} \end{bmatrix}_{\theta=\theta^0} = -2 \begin{bmatrix} \sum_{t=1}^{N} w\left(\frac{t}{N}\right) X(t)\cos(\alpha^0 t + \beta^0 t^2) \\ \sum_{t=1}^{N} w\left(\frac{t}{N}\right) X(t)\sin(\alpha^0 t + \beta^0 t^2) \\ \sum_{t=1}^{N} tw\left(\frac{t}{N}\right) X(t)(B^0\cos(\alpha^0 t + \beta^0 t^2) - A^0\sin(\alpha^0 t + \beta^0 t^2)) \\ \sum_{t=1}^{N} t^2 w\left(\frac{t}{N}\right) X(t)(B^0\cos(\alpha^0 t + \beta^0 t^2) - A^0\sin(\alpha^0 t + \beta^0 t^2)) \end{bmatrix}$$

$$Q^{''}(\theta^0) = \begin{bmatrix} \frac{\partial^2 Q(\theta)}{\partial A^2} & \frac{\partial^2 Q(\theta)}{\partial A \partial B} & \frac{\partial^2 Q(\theta)}{\partial A \partial \alpha} & \frac{\partial^2 Q(\theta)}{\partial A \partial \beta} \\ \frac{\partial^2 Q(\theta)}{\partial B \partial A} & \frac{\partial^2 Q(\theta)}{\partial B^2} & \frac{\partial^2 Q(\theta)}{\partial B \partial \alpha} & \frac{\partial^2 Q(\theta)}{\partial B \partial \beta} \\ \frac{\partial^2 Q(\theta)}{\partial \alpha \partial A} & \frac{\partial^2 Q(\theta)}{\partial \alpha \partial B} & \frac{\partial^2 Q(\theta)}{\partial \alpha^2} & \frac{\partial^2 Q(\theta)}{\partial \alpha \partial \beta} \\ \frac{\partial^2 Q(\theta)}{\partial \beta \partial A} & \frac{\partial^2 Q(\theta)}{\partial \beta \partial B} & \frac{\partial^2 Q(\theta)}{\partial \beta \partial \alpha} & \frac{\partial^2 Q(\theta)}{\partial \beta^2} \end{bmatrix}_{\theta=\theta^0}.$$

The elements of $Q^{''}(\theta^0)$ are given at the end of this appendix. Let us denote

$$\boldsymbol{D} = \operatorname{diag}(N^{-1/2}, N^{-1/2}, N^{-3/2}, N^{-5/2}). \tag{21}$$

Then using Result 1, it follows that

$$\boldsymbol{D}Q'(\theta^0) \xrightarrow{d} N_4(\mathbf{0}, 2\sigma^2\, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma}$ is same as defined in (8). Now expanding $Q'(\widehat{\theta})$ around θ^0 using Taylor series we obtain

$$Q'(\widehat{\theta}) = Q'(\theta^0) + Q''(\bar{\theta})(\widehat{\theta} - \theta^0),$$

where $\bar{\theta}$ lies on the line joining $\widehat{\theta}$ and θ^0. Since $Q'(\widehat{\theta}) = \mathbf{0}$, therefore

$$\boldsymbol{D}Q'(\theta^0) = -\boldsymbol{D}Q''(\bar{\theta})\boldsymbol{D}\boldsymbol{D}^{-1}(\widehat{\theta} - \theta^0).$$

Since $\widehat{\theta} \xrightarrow{a.s} \theta^0$, then using explicit expressions of the elements of $Q''(\theta^0)$, and repeated use of

Result 1, we obtain

$$\lim_{N\to\infty} \boldsymbol{D}Q^{''}(\bar{\theta})\boldsymbol{D} = \lim_{N\to\infty} \boldsymbol{D}Q^{''}(\theta^0)\boldsymbol{D} = \boldsymbol{G},$$

where $\boldsymbol{G}$ is same as in (9). Hence the results follow. ∎.

In the following we provide the second order derivatives of $Q(\theta)$ with respect to elements of θ at θ^0.

$$\begin{aligned}
\frac{\partial^2 Q(\theta^0)}{\partial A^2} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left\{\cos^2(\alpha^0 t + \beta^0 t^2)\right\}, \\
\frac{\partial^2 Q(\theta^0)}{\partial B^2} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left\{\sin^2(\alpha^0 t + \beta^0 t^2)\right\}, \\
\frac{\partial^2 Q(\theta^0)}{\partial \alpha^2} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left\{(B^0 t\cos(\alpha^0 t + \beta^0 t^2) - A^0 t\sin(\alpha^0 t + \beta^0 t^2))^2\right\} \\
&\quad -2\sum_{t=1}^{N} t^2 w\left(\frac{t}{N}\right) X(t)\left\{A^0 \sin(\alpha^0 t + \beta^0 t^2) + B^0\cos(\alpha^0 t + \beta^0 t^2)\right\}, \\
\frac{\partial^2 Q(\theta^0)}{\partial \beta^2} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left\{(B^0 t^2\cos(\alpha^0 t + \beta^0 t^2) - A^0 t^2\sin(\alpha^0 t + \beta^0 t^2))^2\right\} \\
&\quad -2\sum_{t=1}^{N} t^4 w\left(\frac{t}{N}\right) X(t)\left\{A^0 \sin(\alpha^0 t + \beta^0 t^2) + B^0\cos(\alpha^0 t + \beta^0 t^2)\right\}, \\
\frac{\partial^2 Q(\theta^0)}{\partial A\partial B} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left\{\cos(\alpha^0 t + \beta^0 t^2)\sin(\alpha^0 t + \beta^0 t^2)\right\}, \\
\frac{\partial^2 Q(\theta^0)}{\partial A\partial \alpha} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)(B^0 t\cos(\alpha^0 t + \beta^0 t^2) - A^0 t\sin(\alpha^0 t + \beta^0 t^2))\cos(\alpha^0 t + \beta^0 t^2) \\
&\quad +2\sum_{t=1}^{N} t w\left(\frac{t}{N}\right) X(t)\sin(\alpha^0 t + \beta^0 t^2), \\
\frac{\partial^2 Q(\theta^0)}{\partial A\partial \beta} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)(B^0 t^2\cos(\alpha^0 t + \beta^0 t^2) - A^0 t^2\sin(\alpha^0 t + \beta^0 t^2))\cos(\alpha^0 t + \beta^0 t^2) \\
&\quad +2\sum_{t=1}^{N} t^2 w\left(\frac{t}{N}\right) X(t)\sin(\alpha^0 t + \beta^0 t^2),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 Q(\theta^0)}{\partial B \partial \alpha} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left(B^0 t\cos(\alpha^0 t+\beta^0 t^2) - A^0 t\sin(\alpha^0 t+\beta^0 t^2)\right)\sin(\alpha^0 t+\beta^0 t^2) \\
&\quad -2\sum_{t=1}^{N} t w\left(\frac{t}{N}\right) X(t)\cos(\alpha^0 t+\beta^0 t^2), \\
\frac{\partial^2 Q(\theta^0)}{\partial B \partial \beta} &= 2\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left(B^0 t^2\cos(\alpha^0 t+\beta^0 t^2) - A^0 t^2\sin(\alpha^0 t+\beta^0 t^2)\right)\sin(\alpha^0 t+\beta^0 t^2) \\
&\quad -2\sum_{t=1}^{N} t^2 w\left(\frac{t}{N}\right) X(t)\cos(\alpha^0 t+\beta^0 t^2), \\
\frac{\partial^2 Q(\theta^0)}{\partial \alpha \partial \beta} &= 2\sum_{t=1}^{N} t^3 w\left(\frac{t}{N}\right)\left(B^0\cos(\alpha^0 t+\beta^0 t^2) - A^0\sin(\alpha^0 t+\beta^0 t^2)\right)^2 \\
&\quad -2\sum_{t=1}^{N} t^3 w\left(\frac{t}{N}\right) X(t)\left(A^0\sin(\alpha^0 t+\beta^0 t^2) + B^0\cos(\alpha^0 t+\beta^0 t^2)\right).
\end{aligned}$$

Appendix C:

We need the following lemma to prove Theorem 3.

Lemma 3: Let us denote

$$S_{1c} = \{\theta : \theta = (A, B, \alpha, \beta)^\top, |\theta - \theta_1^0| \geq 4c\}.$$

If there exists a $c > 0$,

$$\varliminf \inf_{\theta \in S_{1c}} \frac{1}{N}[Q_1(\theta) - Q_1(\theta_1^0)] > 0 \quad a.s. \tag{22}$$

then $\widetilde{\theta}_1$ that minimises $Q_1(\theta)$, is a strongly consistent estimator of θ_1^0.

Proof: It follows by contradiction using simple arguments. ∎

Proof of Theorem 5: Consider

$$\begin{aligned}
\frac{1}{N}[Q_1(\theta) - Q_1(\theta^0)] &= \frac{1}{N}\left[\sum_{t=1}^{N} w\left(\frac{t}{N}\right)(y(t) - \mu(t;\theta))^2 - \sum_{t=1}^{N} w\left(\frac{t}{N}\right)\left(X(t) + \sum_{k=2}^{p}\mu(t;\theta_k^0)\right)^2\right] \\
&= f_{11}(\theta) + f_{21}(\theta)
\end{aligned}$$

where

$$\begin{aligned}
f_{11}(\theta) &= \frac{1}{N}\left[\sum_{t=1}^{N} w\left(\frac{t}{N}\right)(\mu(t;\theta_1^0) - \mu(t;\theta))^2\right] + \frac{2}{N}\left[\sum_{t=1}^{N} w\left(\frac{t}{N}\right)(\mu(t;\theta_1^0) - \mu(t;\theta))\sum_{k=2}^{p}\mu(t;\theta_k^0)\right] \\
f_{21}(\theta) &= \frac{2}{N}\left[\sum_{t=1}^{N} w\left(\frac{t}{N}\right)X(t)(\mu(t;\theta_1^0) - \mu(t;\theta))\right].
\end{aligned}$$

Using Lemma 1, it follows that

$$\sup_{\theta \in S_{1c}} |f_{21}(\theta)| \stackrel{a.s.}{\to} 0,$$

and using lengthy but straight forward calculations and splitting the set S_{1c} as in Theorem 1, it follows that

$$\underline{\lim} \inf_{\theta \in S_{1c}} f_{11}(\theta) > 0 \quad a.s.$$

Hence,

$$\underline{\lim} \inf_{\theta \in S_{1c}} \frac{1}{N}[Q_1(\theta) - Q_1(\theta_1^0)] > 0 \quad a.s.$$

and the result follows. ∎

Appendix D:

In this Appendix we provide the proof of Statement 3 based on Conjecture A.

Let us denote $Q_1'(\theta)$ as the 4×1 derivative vector and $Q_1''(\theta)$ as the 4×4 second derivative matrix of $Q_1(\theta)$. Therefore, using multivariate Taylor series expansion of $Q_1'(\widetilde{\theta}_1)$ around θ_1^0, we can get

$$Q_1'(\widetilde{\theta}_1) - Q_1'(\theta_1^0) = Q_1^{''}(\bar{\theta}_1)(\widetilde{\theta}_1 - \theta_1^0), \tag{23}$$

where $\bar{\theta}_1$ is a point on the line joining $\widetilde{\theta}_1$ and θ_1^0. Now following exactly the same procedure as Theorem 2, we can obtain

$$\boldsymbol{D}Q_1'(\theta_1^0) \xrightarrow{d} N_4(\mathbf{0}, 2\sigma^2\boldsymbol{\Sigma}_1)$$

and

$$\lim_{N\to\infty} \boldsymbol{D}Q^{''}(\bar{\theta}_1)\boldsymbol{D} = \lim_{N\to\infty} \boldsymbol{D}Q^{''}(\theta_1^0)\boldsymbol{D} = \boldsymbol{G}_1,$$

hence the result follows. ∎

To prove Theorem 4, we need the following Lemma.

Lemma 6: $N(\widetilde{\alpha}_1 - \alpha_1^0) \xrightarrow{a.s} 0$ and $N^2(\widetilde{\beta}_1 - \beta_1^0) \xrightarrow{a.s} 0$.

Proof of Lemma 6: Let us denote the 4×4 diagonal matrix $\boldsymbol{D}_1 = \mathrm{diag}(1, 1, N^{-1}, N^{-2})$. Since $Q_1'(\widetilde{\theta}_1) = \mathbf{0}$, therefore, from (23), we can write

$$-\frac{1}{N}\boldsymbol{D}_1 Q_1'(\theta_1^0) = \left[\frac{1}{N}\boldsymbol{D}_1 Q_1^{''}(\bar{\theta}_1)\boldsymbol{D}_1\right] \boldsymbol{D}_1^{-1}(\widetilde{\theta}_1 - \theta_1^0).$$

Using Lemma 1, it can be shown that $\frac{1}{N}\boldsymbol{D}_1 Q_1'(\theta_1^0) \xrightarrow{a.s} \mathbf{0}$ and

$$\lim_{N\to\infty} \frac{1}{N}\boldsymbol{D}_1 Q_1^{''}(\bar{\theta}_1)\boldsymbol{D}_1 = \lim_{N\to\infty} \boldsymbol{D}Q_1^{''}(\bar{\theta}_1)\boldsymbol{D} = \boldsymbol{G}_1.$$

Since $\boldsymbol{G}_1$ is a positive definite matrix, the result follows. ∎

Now to prove Theorem 7, note that using Lemma 6, we obtain

$$\widetilde{A}_1 \stackrel{a.s.}{=} A_1^0 + o(1), \quad \widetilde{B}_1 \stackrel{a.s.}{=} B_1^0 + o(1), \quad \widetilde{\alpha}_1 \stackrel{a.s.}{=} \alpha_1^0 + o\left(\frac{1}{N}\right), \quad \widetilde{\beta}_1 \stackrel{a.s.}{=} \beta_1^0 + o\left(\frac{1}{N^2}\right).$$

Here a random variable $U = o(1)$ means $U \xrightarrow{a.s.} 0$, $U = o\left(\frac{1}{N}\right)$ means $NU \xrightarrow{a.s.} 0$ and $U = o\left(\frac{1}{N^2}\right)$

means $N^2 U \stackrel{a.s.}{\to} 0$. Therefore

$$\mu(t;\widetilde{\theta}_1) \stackrel{a.s}{=} \mu(t;\theta_1^0) + o(1).$$

Hence, the result follows. ∎

Proof of Theorem 5:

Note that it is enough to prove the following. If $X(t)$ is same as defined before and $\widehat{A}$, $\widehat{B}$, $\widehat{\alpha}$ and $\widehat{\beta}$ minimize

$$\frac{1}{N}\sum_{t=1}^{N} w\left(\frac{t}{N}\right)(X(t)-\mu(t;\theta))^2,$$

then $\widehat{A} \stackrel{a.s}{\to} 0$ and $\widehat{B} \stackrel{a.s}{\to} 0$.

To prove the above statement, we denote $\widehat{A}$, $\widehat{B}$, $\widehat{\alpha}$ and $\widehat{\beta}$ by $\widehat{A}_N$, $\widehat{B}_N$, $\widehat{\alpha}_N$ and $\widehat{\beta}_N$, respectively, to emphasis that they depend on N. Suppose $\widehat{A}_N$ does not converge to zero a.s., and the same for $\widehat{B}_N$. Since $\widehat{A}_N$, $\widehat{B}_N$, $\widehat{\alpha}_N$ and $\widehat{\beta}_N$ are all bounded, therefore, there exists a subsequence $\{N_k\}$ of $\{N\}$ such that $\widehat{A}_{N_k} \stackrel{a.s}{\to} \bar{A} > 0$, $\widehat{B}_{N_k} \stackrel{a.s}{\to} \bar{B} > 0$, $\widehat{\alpha}_{N_k} \stackrel{a.s}{\to} \bar{\alpha}$ and $\widehat{\beta}_{N_k} \stackrel{a.s}{\to} \bar{\beta}$. Therefore, if $\widehat{\theta}_{N_k} = (\widehat{A}_{N_k}, \widehat{B}_{N_k}, \widehat{\alpha}_{N_k}, \widehat{\beta}_{N_k})^\top$, then

$$\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{t=1}^{N_k} w\left(\frac{t}{N_k}\right)(X(t)-\mu(t;\widehat{\theta}_{N_k}))^2 \stackrel{a.s.}{\to} \sigma^2 c_1 + \frac{1}{2}(\bar{A}^2+\bar{B}^2).$$

Consider a point $\theta' = (\frac{1}{2}\bar{A}, \frac{1}{2}\bar{B}, \bar{\alpha}, \bar{\beta})^\top$, then

$$\begin{aligned}\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{t=1}^{N_k} w\left(\frac{t}{N_k}\right)(X(t)-\mu(t;\theta_{N_k}))^2 &\leq \lim_{N_k\to\infty}\frac{1}{N_k}\sum_{t=1}^{N_k} w\left(\frac{t}{N_k}\right)(X(t)-\mu(t;\theta'))^2 \\ &\stackrel{a.s.}{\to} \sigma^2 c_1 + \frac{1}{4}(\bar{A}^2+\bar{B}^2),\end{aligned}$$

which is a contradiction. Hence the result follows. ∎

Appendix E

We will show

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\cos^2(\theta_1 t+\theta_2 t^2)=\frac{1}{2}\int_0^1 w(t)dt.$$

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

$$\int_0^1 w(x)dx-\epsilon\le\int_0^1 p_\epsilon(x)dx\le\int_0^1 w(x)dx+\epsilon.$$

Further

$$\frac{1}{N}\sum_{t=1}^{N} p_\epsilon\left(\frac{t}{N}\right)\cos^2(\theta_1 t+\theta_2 t^2)-\frac{\epsilon}{N}\sum_{t=1}^{N}\cos^2(\theta_1 t+\theta_2 t^2)\le\frac{1}{N}\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\cos^2(\theta_1 t+\theta_2 t^2)\le$$
$$\frac{1}{N}\sum_{t=1}^{N} p_\epsilon\left(\frac{t}{N}\right)\cos^2(\theta_1 t+\theta_2 t^2)+\frac{\epsilon}{N}\sum_{t=1}^{N}\cos^2(\theta_1 t+\theta_2 t^2).$$

Suppose

$$p_\epsilon(x)=a_0+a_1x+\ldots+a_kx^k \qquad \Rightarrow \qquad \int_0^1 p_\epsilon(x)dx=a_0+\frac{a_1}{2}+\ldots+\frac{a_k}{k+1}.$$

Now due to Result 1,

$$\begin{aligned}\frac{1}{N}\sum_{t=1}^{N} p_\epsilon\left(\frac{t}{N}\right)\cos^2(\theta_1 t+\theta_2 t^2) &= \frac{1}{N}\sum_{t=1}^{N}\left\{a_0+\frac{a_1 t}{N}+\ldots+\frac{a_k t^k}{N^k}\right\}\cos^2(\theta_1 t+\theta_2 t^2)\\ &\longrightarrow \frac{1}{2}\left[a_0+\frac{a_1}{2}+\ldots+\frac{a_k}{k+1}\right]=\frac{1}{2}\int_0^1 p_\epsilon(x)dx.\end{aligned}$$

Therefore,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} p_\epsilon\left(\frac{t}{N}\right)\cos^2(\theta_1 t+\theta_2 t^2)-\frac{\epsilon}{2}\le\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\cos^2(\theta_1 t+\theta_2 t^2)\le$$
$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} p_\epsilon\left(\frac{t}{N}\right)\cos^2(\theta_1 t+\theta_2 t^2)+\frac{\epsilon}{2}.$$

Hence

$$\frac{1}{2}\int_0^1 w(t)dt - 2\epsilon \le \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\cos^2(\theta_1 t + \theta_2 t^2) \le \frac{1}{2}\int_0^1 w(t)dt + 2\epsilon.$$

Since ϵ is arbitrary, the result follows. Exactly, the same proof will go through for

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} w\left(\frac{t}{N}\right)\sin^2(\theta_1 t + \theta_2 t^2) = \frac{1}{2}\int_0^1 w(t)dt.$$

∎

References

[1] Abatzoglou, T. (1986), "Fast maximum likelihood joint estimation of frequency and frequency rate", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, 708 – 715.

[2] Djurić, P.M. and Kay, S.M. (1990), "Parameter estimation of chirp signals", *IEEE Trans. on Acoustics, Speech and Signal Process.*, Vol. 38, No. 12, 2118-2126.

[3] Farquharson, M., O'Shea, P. and Ledwich, G. (2005), "A computationally efficient technique for estimating the parameters phase signals from noisy observations", *IEEE Transactions on Signal Processing*, vol. 53, 3337 – 3342.

[4] Gini, F., Montanari, M. and Verrazzani, L. (2000), "Estimation of chirp signals in compound Gaussian clutter: a cyclostationary approach", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 48, 1029 – 1039.

[5] Grover, R., Kundu, D. and Mitra, A. (2018), "On approximate least squares estimators of parameters on one dimensional chirp signal", *Statistics*, vol. 52, 1060 – 1085.

[6] Grover, R., Kundu, D. and Mitra, A. (2021), “Asymptotic properties of least squares estimators and sequential least squares estimators of a chirp-like signal model parameters”, *Circuits, Systems and Signal Processing*, vol. 40, no. 11, 5421 - 5467.

[7] Jennrich, R.I. (1969), “Asymptotic properties of the nonlinear least squares estimators”, *Annals of Mathematical Statistics*, vol. 40, 633 – 643.

[8] Lahiri, A. (2012), *Estimators of parameters of chirp signals and their properties*, Ph.D. thesis, Indian Institute of Technology Kanpur, India.

[9] Lahiri, A., Kundu, D. and Mitra, A. (2014), “On least absolute deviation estimator of one dimensional chirp model”, *Statistics*, **48**, 405 - 420.

[10] Lahiri, A., Kundu, D. and Mitra, A. (2015), “Estimating the parameters of multiple chirp signals”, *Journal of Multivariate Analysis*, **139**, 189 - 205.

[11] Montgomery H.L. (1990), *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, American Mathematical Society, 196.

[12] Nandi, S. and Kundu, D. (2004),“Asymptotic properties of the least squares estimators of the parameters of the chirp signals”, *Annals of the Institute of Statistical Mathematics*, **56**, 529–544.

[13] Rousseeuw, P.J. and Leroy, A.M. (2003), *Robust Regression and Outlier Detection*, Wiley, New York.

[14] Vinogradov, I.M. (1954), “The method of trigonometrical sums in the theory of numbers”, *Interscience*, Translated from Russian. Revised and annotated by K.F. Roth and Anne Davenport. Reprint of the 1954 translation. Dover Publications,Inc., Mineola, NY, 2004.

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